ARL-TR-9035 • AUG 2020

Human–Autonomy Teaming: Interaction Metrics and Models for Next Generation Combat Vehicle Concepts

by Lixiao Huang, Nancy Cooke, Craig Johnson, Glenn Lematta, Shawaiz Bhatti, Michael Barnes, and Eric Holder

NOTICES

Disclaimers

ARL-TR-9035 ● AUG 2020

Human–Autonomy Teaming: Interaction Metrics and Models for Next Generation Combat Vehicle Concepts

Lixiao Huang, Nancy Cooke, Craig Johnson, Glenn Lematta, and Shawaiz Bhatti
Arizona State University, Mesa, AZ

Michael Barnes and Eric Holder
Human Research and Engineering Directorate, CCDC Army Research Laboratory

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* August 2020	**2. REPORT TYPE** Technical Report
3. DATES COVERED (From - To) 15 September 2018–14 September 2019	
4. TITLE AND SUBTITLE Human–Autonomy Teaming: Interaction Metrics and Models for Next Generation Combat Vehicle Concepts	**5a. CONTRACT NUMBER** W911NF1820271
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Lixiao Huang, Nancy Cooke, Craig Johnson, Glenn Lematta, Shawaiz Bhatti, Michael Barnes, and Eric Holder	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Arizona State University ATTN: Center for Human, Artificial Intelligence, and Robot Teaming Mesa, AZ 85212	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) CCDC Army Research Laboratory ATTN: FCDD-RLH-BD Fort Huachuca, AZ 85613-7002	**10. SPONSOR/MONITOR'S ACRONYM(S)** CCDC ARL
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) ARL-TR-9035
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	
13. SUPPLEMENTARY NOTES ORCID ID(s): Lixiao Huang, https://orcid.org/0000-0002-5543-8742	

14. ABSTRACT

Optimizing teamwork effectiveness within human–autonomy teams requires understanding the communications between humans and intelligent agents at their key tasks. This report provides a background to Next Generation Combat Vehicle concepts, a taxonomy of human–autonomy teaming interactions, and metrics and models for human–autonomy team effectiveness.

15. SUBJECT TERMS

human–autonomy teaming, HAT, metrics, modeling, team effectiveness, Next Generation Combat Vehicle, NGCV, trust

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Eric Holder
Unclassified	Unclassified	Unclassified	UU	134	**19b. TELEPHONE NUMBER (Include area code)** (520) 538-4701

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

Task 1 was conducted by Lixiao Huang, Craig Johnson, Nancy Cooke, and Eric Holder. We thank Eric Holder and Michael Barnes from US Army Combat Capabilities Development Command (CCDC) Army Research Laboratory (ARL), who arranged for access to Soldiers at Fort Huachuca, Arizona. We thank our subject-matter experts, Major Trenton Wilhite and Major John Hogan, for participating in our interviews.

Task 2 on interaction taxonomy and strategies was completed by Craig Johnson, Glenn Lematta, Lixiao Huang, Eric Holder, Nancy Cooke, and Shawaiz Bhatti.

Task 3 on human–autonomy teaming measures was completed by Lixiao Huang, Craig Johnson, Nancy Cooke, Glenn Lematta, Eric Holder, and Shawaiz Bhatti.

We thank CCDC Army Research Laboratory for funding our trip to AHFE2019, as well as Eric Holder and Michael Barnes for providing feedback to our slides and approving our slides through the ARL Form 1 process.

We thank our sponsor, ARL, under the Cooperative Agreement No. W911-NF-18-2-0271. Thank you to ARL's Dr Andrea Krausman and Dr Kristin Schaefer-Lay for extensive review and commentary to improve this report.

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARL or the US government.

Summary

Optimizing teamwork effectiveness in a heterogeneous team that involves humans with manned and unmanned combat vehicles requires understanding the communications and coordination required for key tasks. This knowledge can be used to guide the design of ergonomic interfaces for these interactions, as well as methods for evaluating teamwork effectiveness. This project systematically examines the tasks involved in Next Generation Combat Vehicle (NCGV) concepts to develop a taxonomy of human–autonomy team interactions and identify initial measures applicable to that context.

The ultimate goal is to develop metrics and models that can measure the human–autonomy teaming effectiveness in NGCV contexts. Without access to an actual NGCV with robotic combat vehicles (RCVs) in operation or existing experienced RCV crews, our approach breaks down into three tasks on the statement of work and embedded deliverables:

- Task 1 (Section 2) provides the background of NGCVs and related interactions during movement-to-contact operations.
- Task 2 (Section 3) develops the interaction taxonomy for the core baseline tasks of an armored platoon and potential interaction strategies, especially for the RCV operators' tasks based on relevant literature and subject-matter interviews regarding previous experiences of combat vehicles.
- Task 3 (Section 4) builds on Tasks 1 and 2 to propose a comprehensive teamwork effectiveness model and suggested potential team-level measures of teamwork effectiveness focusing on the team states with an emphasis on team situation awareness, team trust, team workload, and team resilience. These measures also shed light on the testbed requirements for later empirical testing. Many ideas are new and different from traditional static and dyad measures of teamwork effectiveness. Their strengths and weaknesses are included.

Moving forward, these identified interaction-based measures need empirical tests in team tasks that involve humans, autonomies, and interactions. Then the goal is to structure and define metrics and tease out the contextual variations and interrelations between measures that define team effectiveness in human–autonomous teaming contexts.

1. Introduction

With advances in artificial intelligence and machine learning, technology is increasingly able to serve as a full-fledged team member, rather than a device to supervise or control. Soldiers of the future will not only interact with fellow Soldiers, but with robots of many forms (including the Next Generation Combat Vehicle [NGCV]), intelligent decision aids, and wearable devices that sense their current physiological state. Manned vehicles will be required to interact with unmanned vehicles. The challenge is to get these heterogeneous and distributed agents to interact as an effective team, while managing workload and preserving team situation awareness (SA), resilience, and trust. To address this challenge, the research presented here focuses on identifying novel human–agent interaction models, measures, and metrics of human–agent teaming effectiveness.

1.1 Human–Autonomy Teaming (HAT)

Recently, teams, including military teams, have expanded to include intelligent artificial agents (Burke et al. 2004; Salas et al. 2008). An intelligent agent is defined as “anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors” (Russell and Norvig 2016, p. 34). This also includes nonhuman (i.e., artificial) entities such as robots, vehicles, and other automated systems in the vehicles. A *human–autonomy team* refers to a team comprising both humans and intelligent agents that act interdependently to achieve team-level goals. These human–autonomy teams may be preferred to traditional all-human teams when coordinated action leads to safer and more efficient task completion, or higher performance levels previously unachievable by all-human teams. For instance, maneuvering through a collapsed structure may be dangerous, difficult, or impossible for humans, but is critical for urban search and rescue operations (Burke et al. 2004).

Future autonomous combat vehicles may utilize several control structures to coordinate with other agents. Some existing models of control in human–automation interaction include teleoperation, supervisory control (Sheridan 2002), and various models of shared control (Allen et al. 1999; Chen and Barnes 2014; Johnson et al. 2014). Each model involves underlying strategies to achieve effective coordination of multiple human and artificial agents with varying capabilities. On the battlefield, it is impossible for any individual to have a complete understanding of the situation. Instead, the active integration of heterogeneous perspectives is required to coordinate efforts and achieve collective goals. Team-level cognitive processes such as planning, reasoning, decision making, and acting (i.e., team cognition) require team interactions (Cooke et al. 2013). Research on team interactions and coordination measurement is needed when developing NGCVs to support effective HAT design.

1.2 The NGCV Context

NGCVs are a family of future military vehicles, aiming to develop mobile protected firepower using modern technology and protect the lives of Soldiers. In NGCVs, the team composition may vary, such as the number of crew members and combat vehicles. The concept version of NGCVs in this research includes seven crew members, one manned combat vehicle (MCV*), and two unmanned robotic combat vehicles (RCVs) as wingmen to improve crew survivability and increase lethality. However, operating an unmanned vehicle while separated and sitting in a moving manned vehicle changes the nature of many current tasks, which requires appropriately reallocating functions and providing effective interfaces to support human decision making and team performance. This is especially challenging when uncertainty and possibilities abound in the problem space.

For one envisioned employment variation, the NGCV's platoon consists of two sections (Fig. 1), and each section comprises one MCV and two unmanned RCVs. Each vehicle is operated primarily by two people, and all operators in one section are seated inside the MCV in that section, with the seventh person possibly serving as the vehicle commander and section leader. One section leader (a.k.a., vehicle commander) might serve as a platoon leader, supervising both sections in a platoon; the other vehicle commander may serve as a platoon sergeant. The platoon leader may also interact with entities outside the platoon, including the company commander, area commander, and other infantry units. The NGCV platoon should be capable of conducting offensive, defensive, and stability tasks in support of unified land operations as a part of future armored combat. We choose one NGCV platoon section during movement-to-contact (MTC) scenarios to be a manageable piece to examine in detail at this first stage.

* A MCV is sometimes called a manned fighting vehicle (MFV) and is used interchangeably here.

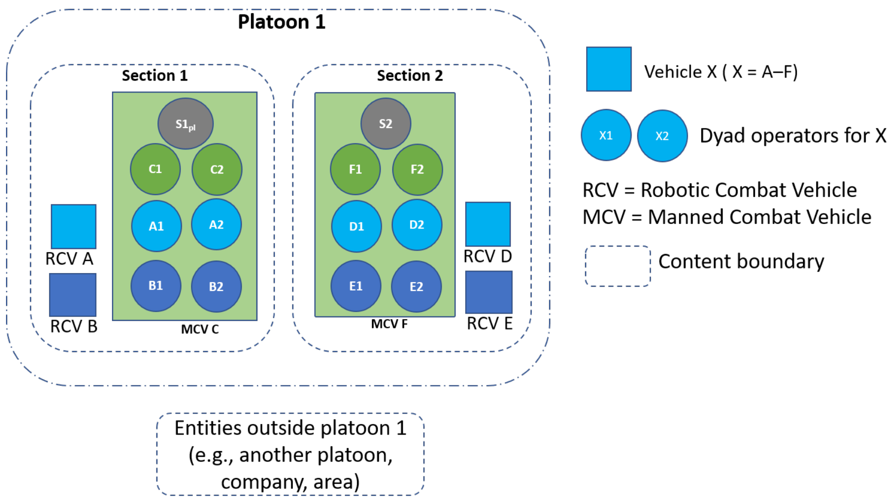

Fig. 1 Possible sitting structure of an NGCV's platoon. A–F is the label for vehicles, and each vehicle is controlled by two people within the same MCV and designated in the same color as the vehicle being controlled (e.g., A1 and A2 control RCV A and are all light blue). Section leaders are colored gray (S1pl = platoon leader, S2 = platoon sergeant) and can fill the role of vehicle commanders as well as supervising sections 1 and 2, respectively, while S1pl also oversees the whole platoon.

1.3 Current Work

The goal of this report is to develop metrics and models that can measure the HAT effectiveness in NGCV contexts. Without access to an actual NGCV with RCVs in operation or existing experienced RCV crews, our approach breaks down into three tasks:

- Task 1 (Section 2): Background of NGCVs and related interactions during MTC operations
- Task 2 (Section 3): Interaction taxonomy for the core baseline tasks of an armored platoon and potential interaction strategies
- Task 3 (Section 4): A comprehensive teamwork effectiveness model and suggested potential team-level measures of teamwork effectiveness

2. Interaction Analysis for the Development of NGCVs (Task 1)

HAT includes direct and indirect communications among humans, as well as technologies used in the tasks that help crew members to get information from the environment or execute some tasks. This section describes Task 1 of this project: conducting subject-matter expert (SME) interviews and interaction analysis. The goal of Task 1 was to understand the tasks, information requirements, and context in current operations that would inform the analysis of potential HAT in the NGCV context. Through these interviews, we also tried to understand the type of interactions typical of each, especially those envisioned to include autonomy or nonhuman agents. This section summarized the interviews, and describes the problem context and envisioned scenarios.

2.1 Interaction Analysis

Recent advances in autonomous technology suggest that near-future military warfighting systems will see more machine intelligence in the decision–making loop (Cummings 2017). These systems will be capable of executing complex and safety-critical tasks, such as maneuvering combat vehicles on a battlefield. As these individual systems are typically organized into teams and multiteam systems, the need for embedded autonomous agents to effectively coordinate and "team up" with humans grows.

Armored vehicles such as the M1 Abrams Tank and M2 Bradley Fighting Vehicle have been demonstrated to be an effective means of providing mobile, protected firepower on the battlefield. However, recent initiatives such as the *DOD Unmanned Systems Integrated Roadmap FY2017–2042* (2017) lay the groundwork for integrating unmanned vehicles within the future of the US military. Unmanned variants of NGCVs show promise for fulfilling roles that have traditionally been filled by manned vehicles. These vehicles will have several autonomous capabilities and will team up with Soldiers to accomplish mission tasks. System interdependence means that these new systems will likely change the nature of traditional tasks and may introduce unintended consequences.

For instance, interface changes as a consequence of teleoperation are well documented (Sheridan 2002). These include differences in feedback availability and format that can potentially decrease SA. Yet, the interactions among teleoperated systems, autonomous capabilities in these systems, and vehicles crews are poorly understood. For example, an operator could be responsible for controlling an RCV, switching between teleoperation and semi-autonomous mode. If their commander observes the RCV deviating from a planned route or formation, what communication should follow to remedy the situation? The commander might have several other concurrent tasks that preclude immediate communication to the RCV operator. In this instance, will the commander be able to distinguish autonomous RCV actions from human actions in the absence of explicit communication? Should the commander change how they rely on either agent in the moment? These questions become more relevant in battlefield contexts as assumptions of risk and tradeoffs are necessary for tactical decision making.

By conducting an interaction analysis of HAT related to NGCVs, we can identify the core tasks, information requirements, and interactions required, which provides critical insights into the communication issues that are related to germane tasks and procedures. From this, we can develop appropriate measures for design criteria that respect individual differences and constraints, and identify effective measurement strategies. This Task 1 summary describes the problem scope and presents findings from a literature review and SME interviews.

2.1.1 Movement to Contact

One specific task, MTC, has been selected for examination in this study due to the high degree of coordination and flexibility required to execute it effectively. MTC is described as "an offensive task designed to develop the situation and to establish or regain contact. It also creates favorable conditions for subsequent tactical actions. The platoon conducts MTC when the enemy situation is vague or not specific enough to conduct an attack. Movement to contact may result in a meeting engagement. Once contact is made with an enemy force, the platoon leader has five options: attack, defend, bypass, delay, withdraw" (US Army 2019). Preliminary examination of the interactions required to conduct MTC demonstrates a wide variety of potential interactions among both human and vehicle agents.

2.1.2 Military Literature Review

Researchers used the available literature to gain an initial understanding of the task and information requirements for armored vehicle operations overall and MTC in detail. The literature included materials such as the Manned–Unmanned Teaming (MUMT) RCV storyboard produced by the US Army Tank Automotive Research Development and Engineering Center (TARDEC[†]) and the US Army Combat Capabilities Development Command (CCDC) Army Research Laboratory (ARL); Bradley, Tank, Armor, as well as Platoon and Squad Training and Gunnery Manuals (US Army 1992, 2008, 2019), combined with team internal brainstorming and one researcher's tank platoon leader experience to understand the main tasks. The tasks of NGCVs are roughly categorized as mobility (regarding movement), reconnaissance (regarding information search only), and gunnery (regarding intentional target searching and shooting). Field Manual (FM) 7-8 does an especially good job of explaining the communications needed and principles (US Army 1992).

[†] TARDEC is now called the US Army CCDC Ground Vehicle Systems Center (GVSC).

2.2 SME Interviews

The literature review provided a broad knowledge about the platoon setting and tasks. However, some NGCV-related questions require an answer from people who are familiar interacting and operating within the platoon setting and battlefield operations. The SME interviews had three purposes for our project: 1) provide increased understanding of battlefield operations, 2) understand the Soldiers' needs in the existing situations, and 3) obtain Soldiers' feedback and input on conceptual NGCV operations and MTC in particular. The direct conversations helped address many questions and provide additional insights to the tasks and measurement requirements.

2.2.1 Participants

SME 1 was an Army Major with experience as an Abrams tank platoon leader, Bradley Cavalry Fighting Vehicle (CFV) scout platoon leader, and Armor officer.

SME 2 had 10 years of experience as an infantry officer, including time as a mechanized infantry platoon leader and an infantry school instructor.

2.2.2 Interview Procedure

We visited one of the CCDC Army Research Laboratory sites and interviewed one SME for 2 h and the other for 1 h, followed by 30-min team debrief. Our interview questions focused on SME experiences concerning MTC and other tactical tasks performed as a platoon leader. Special focus was paid to the challenges in communication and information management they had encountered on the battlefield.

2.3 Key Findings from the SMEs Interviews

Through the SMEs interviews, we identified two global characteristics of the MTC environment that helped to guide our analysis. The first is that risk perception and risk management will be an ever-present topic and goal. The second is that communication and interaction are essential elements that drive teamwork, and these processes can evolve with the team and their experiences. These are described more in the following and tie into the goal of the parent project, ARL's HAT Essential Research Program Project 5, looking at trust and team cohesion metrics in HAT environments. To best understand communication, we further investigated the communication tools, interactions, and information exchange in MTC. This section focuses on our feedback from the SMEs regarding these topics.

2.3.1 Two Essential Elements of the MTC Environment

2.3.1.1 Risks Perception and Risk Management

Any battlefield bears a lot of risks, which are dynamic over time. The definition of risk has three aspects: 1) the source of the hazards, 2) the severity of the consequence, and 3) the probability of the event happening (USNRC 1983). Soldiers can be trained to prepare and cope with some risks more effectively, but some risks are unpredictable and not acceptable. The perception of risk influences team strategies. Developing good SA of the battlefield environment is a key element to managing risks (Endsley 1995b). A recurrent theme across both SMEs was the concept of "acceptable risk" as a factor that impacted most of the coordination and maneuver decisions. Different risk types have varying requirements for information richness and accuracy. The line between acceptable and unacceptable risks was not explicit in the interviews. There may be objective criteria for mission success and damage evaluation, as well subjective individual differences, regarding what is acceptable and unacceptable. There are many ways to display and represent risks (Spiegelhalter 2017), and it is also critical to explore how the autonomy could help humans better understand risks (Ono et al. 2015; Huang et al. in review) and thus make better decisions.

2.3.1.2 Communication

Communication plays a significant role in developing SA, managing risks, and accomplishing tasks. Communication aids SA by enabling the exchange of information between agents. This exchange can improve the perception of the immediate situation, facilitate knowledge integration, enable projection of potential future states, and improve decision making. The ways in which information is communicated among agents within a system can directly impact the success of a mission. Good communication and good coordination walk side by side in task execution. Poorly timed, excessive, and irrelevant or incorrect communication can interfere with mental processing and task performance, including perception and decision making.

Interactions are the core of teaming and teamwork. It is therefore essential to understand both the communications that are taking place as well as how the environment provides information and structures those communications. We examined communication tools, environmental information and interaction pathways, constraints, and content in the current environment to set the baseline for interactions in NGCV environments.

2.3.2 Communication Tools

2.3.2.1 Intercom and Radio

Within many armored vehicles, an intercom system is used to communicate among the vehicle crew. The intercom system directly links the members of the crew and allows them to communicate verbally using a microphone and noise-canceling headset in noisy environments. Meanwhile, the radio is used to broadcast information between units across moderate distances. The intercom and the radio are integrated into the helmet (Fig. 2). The radio system can transmit and receive over a variety of frequencies, allowing for customization according to mission needs. The integration of the intercom and radio into compatible hardware allows for multiple different communication networks to be formed between different units (i.e., crew, section, and platoon). Controls are used by individual crew members to select which networks they are monitoring and communicating on. According to SMEs, the intercom and radio system is the most basic and expedient way to communicate with different units.

Fig. 2 Military combat vehicle crew tank helmet headset with radio intercom embedded (photo credit: US Army, Sgt. Mason Cutrer https://www.defense.gov/observe/photo-gallery/igphoto/2002137607/)

2.3.2.2 GPS

Crew members may have a GPS device to check real-time grid coordinates, azimuth, altitude, and time (Fig. 3) in conjunction with a printed map. Sometimes, several members in a squad may share one.

Fig. 3 Example of a GPS device (photo credit: https://www.defense.gov/observe/photo-gallery/igphoto/2002018244/)

2.3.2.3 Computer Screens or User Interfaces

There are typically computer screens within the vehicle where some crewmembers will have access to information. These systems are also prone to change with time as well as be updated. The SMEs provided examples from their experience. The SMEs discussed two types of computer screens in the combat vehicles they experienced. First was the Force XXI Battle Command Brigade and Below (FBCB2), a satellite network-based multipurpose communication platform (Fig. 4). The FBCB2 provides a common operating picture (COP) and a topographical map (Fig. 5). The COP displays the locations of friendly units, as well as graphic control measures such as mission routes, waypoints, and objectives, and fire control measures such as terrain reference points and engagement areas. Soldiers can use a touchscreen to add friendly force icons to the COP and use a keyboard to exchange text messages that are not subject to the constraints of transient verbal (radio) messages and send preformatted brief reports to describe an enemy activity or update supply status.

Fig. 4 **Example of monitors in a combat vehicle (Center=CITV; Right=FBCB2; photo credit: US Army: https://www.army.mil/article/63135/army_links_tactical_radios_chat_services_with_commercial_communications)**

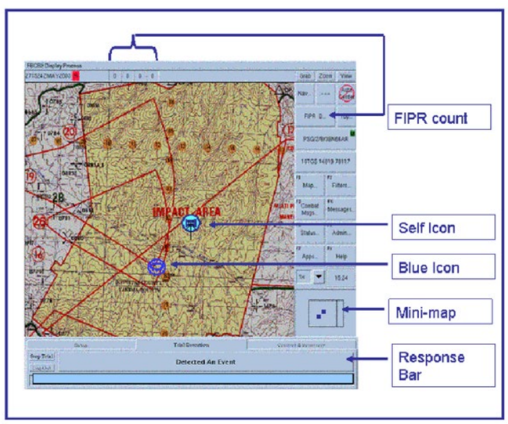

Fig. 5 **Example of an FBCB2 screen (photo credit: Durlach [2004])**

Second is the commander's independent thermal viewer (CITV) (Fig. 6). The CITV can be manually zoomed to find and designate targets. In a high-intensity conflict (i.e., people are shooting at you), the commander would use it to identify targets for their gunner. The gunner would often get “tunnel vision” and so the vehicle commander would continue to scan the area while the gunner engaged targets. In one SME’s experience, only the two leader vehicles had the monitors.

Fig. 6 Example of the CITV from the outside (in white circles)

SMEs expressed that they generally avoided looking at the FBCB2 screens during operations on the battlefield. Instead, they used a GPS device, and a printed map, overlays, and intercom/radio communication for coordination and decision making. As platoon leaders, they only used the FBCB2 during low-tempo periods. One reason for this disuse is that the icons on the screen are hard to read, along with slow update rates, and some icons that do not get updated. Another reason is the difficulty in precise input motion control in a moving vehicle, often on difficult terrain, under high workload and time pressure.

2.3.3 Environmental View

In addition to the CITV, traditional direct observation of the environment is accomplished through the use of the driver' periscopes, the gunner's aperture (Figs. 7–8), the vehicle commander's turret hatch (Fig. 9), and vision blocks (Fig. 10). Other crew members may have access to a hatch as well. In addition, sometimes crew members dismount from the vehicle and walk around to see the environment. These features may or may not be available in NGCVs, and their absence would introduce a greater reliance on displays and imagery to provide that input.

Fig. 7 Example of the M1A2 gunner's station (photo credit: US Army images by Spc J Hester-Heard, https://www.youtube.com/watch?v=KW2YMnD3HYk)

Fig. 8 Example of the gunner aperture view in a training simulation (photo credit: US Army photo by Sgt R Hale, 1st Inf. Div. Public Affairs, https://www.flickr.com/photos/soldiersmediacenter/6198162188/)

Fig. 9 Example of the M2 Bradley and M1A1 tank with multiple hatches being used by crew members highlighted with white circles (photo credit: https://en.wikipedia.org/wiki/M2_Bradley#/media/File:M2a3-bradley07.jpg)

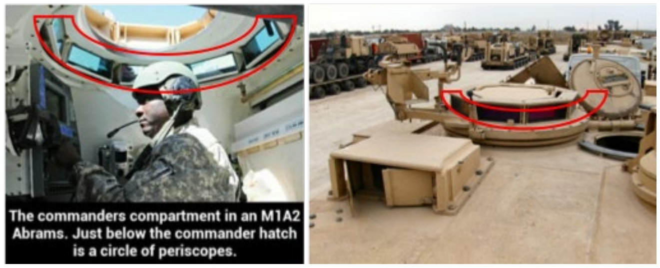

Fig. 10 Vision blocks highlighted in red (left=from the inside; right=from the outside)

The objective of any visual feedback is to provide necessary information about the environment. The traditional method of vehicle operation is to view through apertures and the open hatch to provide a 360° view when needed and safe to do so. It could be challenging for RCV operators to teleoperate from a moving MCV for a variety of reasons. For example, conflicting vestibular sensations from the movement of the MCV (where the operators are located) relative to the RCV may cause navigation difficulties and motion sickness, especially on rough, undulating terrain. Visual input from a screen can cause a decoupling in the perception–action cycle due to a lack of, or misalignment of, perceptual cues that are normally present in the natural environment such as depth cues and binocular stereopsis, vestibular acceleration sensations, a lag between input and feedback, and users experiencing a keyhole effect (Woods et al. 2004). Additionally, misalignment in fields of view of the various crew members with their own and each other's vehicles can add confusion to the coordination and actions.

Each crew member is expected to have up to three monitors in NGCV concepts, with potentially different information for their designated positions. There are at least three different roles: MCV operators, RCV operators, and platoon leader or platoon sergeant. If each pair of RCV operators needs different information for their specific task (e.g., mobility vs. gunnery), it will be helpful for them to have customizable screens so that they can find the information they need quickly based on their preference. However, having consistent display organization and settings may facilitate position rotation and efficient reporting and information sharing.

All these issues in environmental views may influence operators' SA and should be evaluated to determine the effectiveness of human–autonomy interfaces and interactions. With visual input as one primary factor, additional environmental issues need to be considered as well. For instance those associated with Soldiers' perceptions in a moving vehicle called "vehicle motion effects"—the effects of vehicle motion on Soldier performance, including motion sickness (Hill et al. 2004). The literature in this field is relatively old and the research needs to be updated using current technology.

2.3.4 Interaction Pathways, Constraints, and Content

2.3.4.1 Radio and Intercom Networks

Current radio network management strategies include volume control on one or both ears, call signs, and context/role-specific terms to differentiate the communication channels and conversations when multiple agents are online. However, this may raise an issue of attentional overload when managing multiple networks simultaneously. It is important for crew members to discern the networks correctly with an acceptable workload in a noisy environment. Strategies like call signs and channel management can be used, and crew members' workload and efficiency in information exchange are evaluated.

2.3.4.2 Span of Command

One platoon leader mentioned the span of command is four, as a common rule of thumb for interaction channels. In other words, a commander should try to avoid directly managing more than four people (with one person in charge of each of the following elements: supervisors, peers, and crew members by vehicles). Therefore, instead of interacting with all of the individual agents (i.e., operators of a vehicle, payloads and the vehicle) (Fig. 11), the platoon leader would choose to talk with one person in charge of a vehicle (typically, the vehicle commander; Fig. 11, right). However, it is often beneficial for other people in the vehicle to hear the commands and exchanges to maintain SA. In the case of the RCV in the context of NGCVs, the platoon leader is likely to communicate with A1 in the unit of RCV A, but A2 may still be aware of the information that the platoon leader communicates.

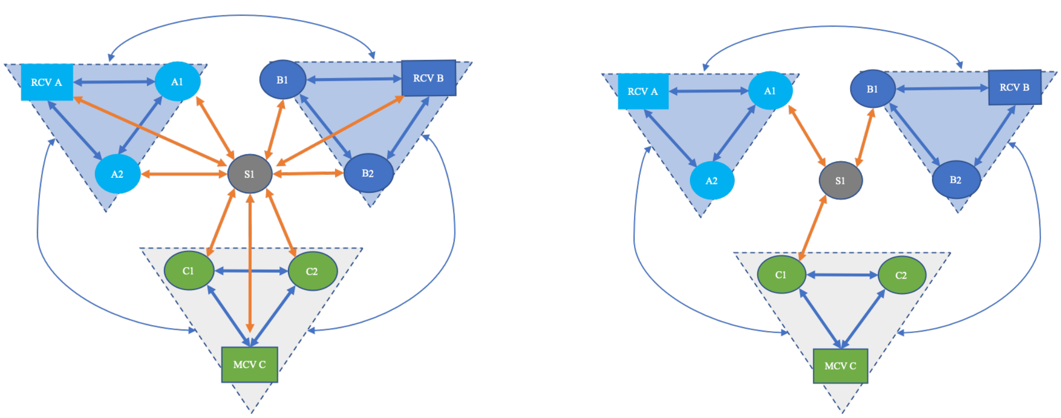

Fig. 11 Communication channel versions. Version A on the left: A1 = operator and vehicle commander for RCV A; B1 = operator and vehicle commander for RCV B; C1 = operator and vehicle commander for MCV C. Version B on the right, which is much simplified.

The platoon leader would be responsible for overseeing both sections, including a total of 14 crew members and 6 vehicles (Fig. 12). Likewise, the platoon leader would not interact with everyone in the other section but primarily only the section leader. Coordination across sections would be required for activities such as coordinating sectors of fire for longer-lasting or synchronized firepower and better shooting coverage and coordinating movements.

In addition, the platoon leader reports to the company commander (line 1T17 in Fig. 12), and may occasionally communicate directly with other units such as adjacent platoons, dismounted units, and when conducting a call for artillery support. Figure 13 and Tables 1–4 show examples of the layers of interaction, example content, and key questions to be answered, which were created based on Army literature and vetted by SMEs in separate sessions from those mentioned previously. There can be exceptions to this strict separation of coordination paths. For instance, the SMEs reported one example where direct communication between platoons may take place is if one platoon crosses into another's area of operations. However, coordination between platoons is generally conducted on the company commander level. The crew members do not report out except through the platoon leader.

When crew members dismounted, crew members would reorganize to different sections or elements: mounted and dismounted. For nets, the platoon leader and platoon sergeant had the platoon frequency, the company frequency, and the internal vehicle frequency. When the platoon leader dismounted, they only took the platoon frequency and the company frequency with them.

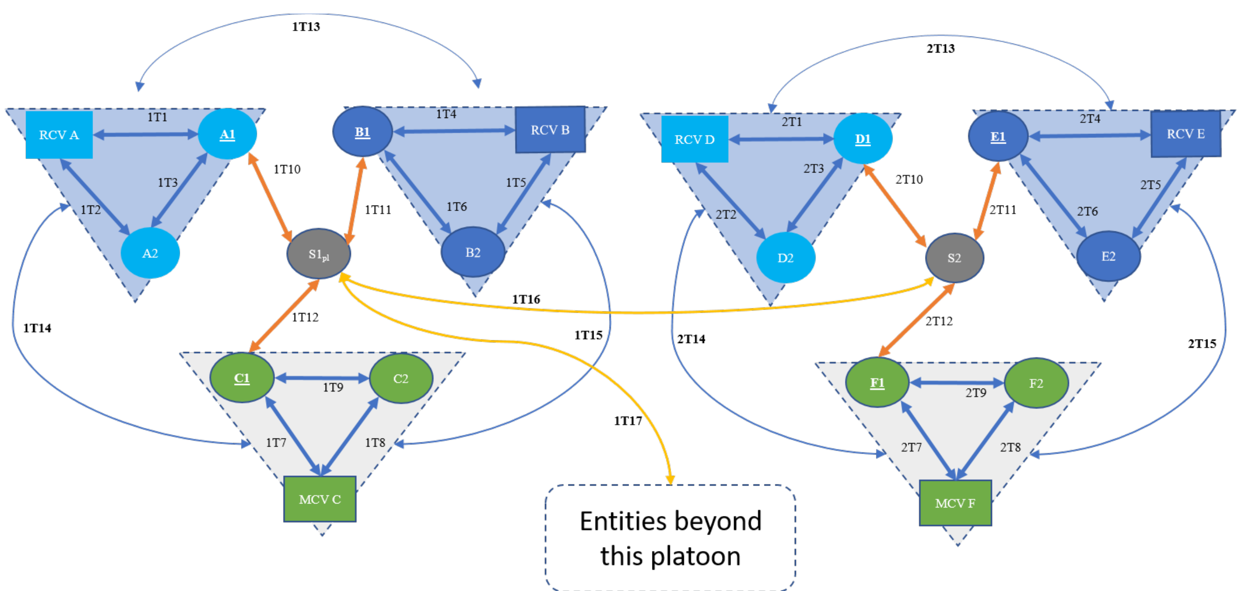

Fig. 12 Communication channels in a platoon. To track each interaction, we used interaction labels, for example, "1T1": the digit before T = section 1, T = a team of two/dyad, the digit(s) after T = the ID code of the interaction; and entities beyond this platoon = primarily the company commander.

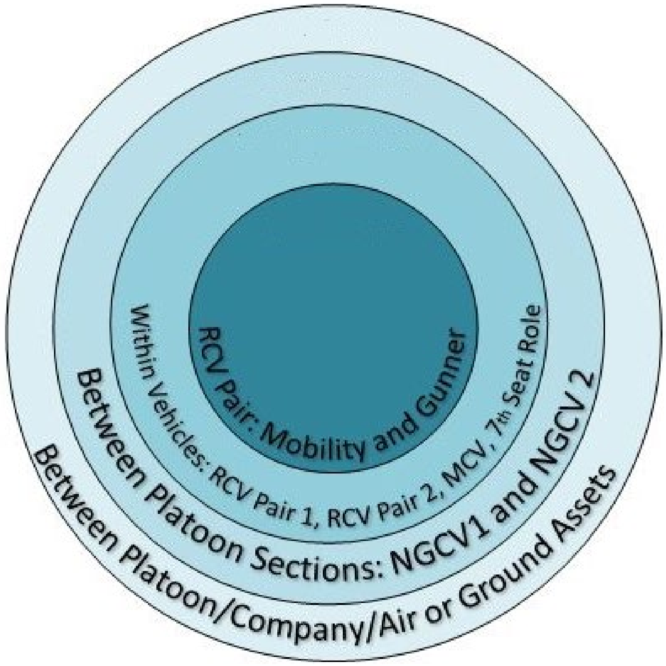

Fig. 13 System view of coordination complexity based on an NGCV example. GOTWA = 5 point contingency plan (G) where I'm Going, (O) Others I'm taking, (T) Time of my return, (W) What to do if I don't return, (A) Actions to take if I'm hit or Actions to take if you're hit; BDA = battle damage assessment; PoC = point of contact; and ACE = ammunition, casualty, and equipment report.

Table 1 **Layer 1 interactions between RCV pair: mobility and gunner**

Tasks	Considerations
Navigating/controlling RCV in support of NGCV tasking and platoon mission	o Understanding position in relationship to key references (e.g., alignment with NGCV/direction of travel/enemy, target reference points, sectors of fire and shift lines, fire control measures, objects on the ground that are references to these, location of friendly, enemy and neutral entities, obstacles, terrain) o Move in response to orders and situation (support by fire, assaulting force, recon pattern, movement type planned, e.g., bounding) o Inform #2 and #3 if making larger or unexpected movements, confirm when at assigned locations o Support targeting (e.g., positioning vehicle, movements) with partner o Avoiding accidents, injuries and damages (exact control and positioning)
Recon and information collection	o Identification/confirmation of enemy locations, strength movements, etc. o Accurately reporting through the chain (e.g., SALUTE reports on contact) intercom internal, radio external at minimum platoon leader and platoon sergeant informed
Enemy engagements (assume NGCV armed or can call for fires)	o Exercise actions as planned in Operations Orders (OPORD), GOTWA, rules of engagement (ROEs), battle drills, or standard operating procedures (SOPs) o Adjust or adapt as needed (inform #2 and #3 as required) o Understand sectors of fire assigned/allowed o Identify targets o Get any required approvals (positive target identification, to fire, to adjust/continue firing) o Confirm weapon status o Engage targets, adjust as allowed o Report status of target (BDA)

Note: Once in contact, traditionally Soldiers do not look down at their maps or displays much, and radio is the primary means of coordination. This may change with more unmanned vehicles and a requirement to be looking at a display but keep it in mind if assuming heavy interface interaction to complete tasks.

Table 2 Layer 2 interactions within vehicle: RCV pair 1, RCV pair 2, MCV pair, and 7th seat role

Tasks	Considerations
Who has primary role: platoon leader or platoon sergeant in vehicle, 7th seat?	o Consider concurrent roles for this person (driver, navigator, or gunner of MCV, PoC for #3, etc.) o Will there be a coordinator for actions in the back that communicates to those in the front and outside the vehicle?
Maintain SA of outside environment and battle situation (location and status of enemy and friendly entities and relevant neutral entities)	o Monitor section performance
Leader responsible for decision making on the vehicle	o Some pre-mission planning: routes, objectives, actions: SOPs o Section tasking, who, when and where it will move and the formations for the section o (option) Provide confirmations and approvals (positive target identification, to fire, to adjust/continue firing) as required by ROE and SOP o Designating references for the section (e.g., point on ground will align with) o May provide sectors of fire for Section vehicles
Accurately reporting through the chain (#3 and #4)	o BDAs, ACE reports, mission status, causalities, etc.

Note: The platoon leader and platoon sergeant could both be the leader for a vehicle and are usually split each leading a section in a platoon. In this case, there are only two manned vehicles in the platoon so each will likely have either the platoon leader or platoon sergeant. This does not rule out the addition of another Soldier, perhaps with the experience and qualification to take some of these tasks.

Table 3 Layer 3 interactions among platoon sections

Tasks	Considerations
Who has primary role: typically platoon leader but if they are driver or gunner could be 7th seat?	○ Consider concurrent roles for this person (driver, navigator, or gunner of MCV, vehicle coordinator #2.)
Maintain SA of outside environment and battle situation (location and status of enemy and friendly entities and relevant neutral entities)	○ Monitor platoon performance ○ Make sure all platoon members know their piece of the larger picture.
Leader responsible for decision making for the platoon	○ Pre-mission planning: routes, objectives, points-meeting/rally, OPORD, GOTWA (who does what for specific events), actions-SOPs, battle drills, communications and passwords ○ Locations and manning of Observation Posts (as directed by Company: #4) ○ Determining Platoon formations and changes in maneuver or planned actions ○ (option) Provide confirmations and approvals (positive target identification, to fire, to adjust/continue firing) as required by ROE and SOP ○ Designating references for the platoon (e.g., point on ground will align with, target reference points, sectors of fire, fire control measures, engagement areas), and alert all relevant friendly forces to any changes ○ Commands when to shift fires ○ Coordinates assault movements
Accurately reporting through the chain (#4) and back down to #1 and #2	

Note: The platoon leader and platoon sergeant could both be the leader for a vehicle and are usually split each leading a Section in a Platoon. In this case there are only two manned vehicles in the platoon so each will likely have either the platoon leader or platoon sergeant. This does not rule out the addition of another Soldier, perhaps with the experience and qualification to take some of these tasks.

Table 4 Layer 4 interactions among platoons, company, and air or ground forces

Tasks	Considerations
Who has primary role to coordinate: typically platoon leader but if they are driver or gunner could be 7th seat?	○ Consider concurrent roles for this person (driver, navigator, or gunner of MCV, vehicle coordinator #2.)
Provide reports (situation reports [SITREPS], SALUTE, contact, ACE)	
Inform others outside platoon on planned actions, key locations, signals and initiation of actions	
Inform others outside platoon on target information, range cards, sectors of fire, or engagement or disengagement criteria as needed	
Incorporate guidance from higher headquarters (HHQ) (company or battalion) or others outside platoon	○ Orders (OPORD, warning order [WARNO], fragmentary order [FRAGO], etc.) ○ Geographic points/lines/areas ○ Location of other friendly and enemy units
Incorporate intelligence and information received into plans and actions and disseminate new information	○ Terrain information ○ Changes in friendly situation or updates from other units that HHQ cannot monitor
Coordinate actions with others outside platoon as required	○ Indirect fire support plan ○ Calls for indirect fires (Forward Observer) ○ MEDEVAC ○ Coordination with dismounted forces ○ Any nuclear, biological or chemical concerns ○ Changes in platoon action from plan

2.3.4.3 Radio Network Management

A current platoon leader can maintain 3–4 radio networks, including the direct higher-level personnel (e.g., headquarters and/or company commander), the operators of other vehicles in the same platoon, and one's own crew. As the SMEs reported, there are a lot of individual preferences involved in how a leader sets up and monitors multiple channels. The two SMEs used different strategies to manage these radio networks. SME 1 used different volumes for each network, and each ear for a different net, with both ears open to the third network—enabling them to reference who was communicating with them spatially. Differently, SME 2 used the same volume for all three networks, though both SMEs use call signs and context/role-specific terms, so that command receivers have two ways to recognize the sender and commands. In general, more information is passed down than sent up. If radio networks get oversaturated and a piece of important information needs to be pushed through, an operator could say, "Break, break, break," which signals for everyone to stop talking and "clear the net".

2.3.4.4 Information Content and Medium

To fulfill the role of a platoon leader, the individual needs to know information about the enemy, any abnormal conditions that prevent crew members from continuing their tasks, whether the status of a vehicle is set for the other vehicle to move, as well as terrain and locations of friendly forces or neutrals, and so on. However, a platoon sergeant, an RCV driver, an RCV gunner, and the MCV operators may need different information for them to continue their work. For each of these roles, the types of information, the frequency of communication, and the medium of communication need to be investigated to evaluate whether this information is efficiently and effectively communicated.

2.3.4.5 Push versus Pull Communication

When developing human–machine interfaces, it is important to make the information available for users to pull information from the monitors as needed rather than proactively pushing additional alerts when the crew members are busy handling other things. When alerts become annoying, people might turn them off or silence them. Artificial intelligence may be developed to recognize when people are occupied with tasks and wait (Huang et al. 2015). However, proactively pushing information to the crew members may be better for germane and critical information in some cases.

The criteria for effective information exchange are to build good SA by providing the right information to the right person at the right time through the appropriate medium. Any specific design should consider the human's limit of attention and workload. Multiple Resource Theory (Wickens 2002) recommends against overloading one modality during time sharing (simultaneous multitasking) and the signaling principle (Mayer 2001), which suggests highlighting important information.

2.3.5 Operator Interaction Preferences

During the interviews SMEs provided their personal experiences and the preferences that came out of those. These are informative to document and should be validated across other Soldiers and, if determined to apply to NGCVs contexts, factored into both design and evaluation decisions. This section provides discussion of some of those experiences and preferences reported.

2.3.5.1 Common Operating Picture (COP)

The COP in this context refers to the digital map interface displayed on the FBCB2. The COP shows the vehicle-level fidelity, but not the dismounted infantry. Much of the information on the FBCB2 is shared among all the users over a distributed satellite network. The intent of the COP is to enhance shared awareness of the battlefield situation. However, the SMEs expressed that they generally disliked the COP functionalities in the FBCB2 for a few reasons, including excess irrelevant information leading to display clutter and slow refresh rates relative to mounted vehicle travel speed. This generally resulted in increased workload during activity peaks and disuse. They expressed that in future systems they would prefer to have information available, but not forced upon them. That is, the ability to easily select and filter information as needed. They also expressed that non-FBCB2 system alarms such as the maintenance warning were often very distracting during peaks of activity as a vehicle commander and led to counterproductive interruptions.

2.3.5.2 Less Communication is Preferred

One SME stated that less communication is preferred and that less communication is often indicative of a well-functioning team, but this highly depends on the familiarity between the team members, and the expertise level of the crew members. Crew members who have worked together for a long time and know each other well develop a rich understanding of each other's patterns of interactions and intentions on the battlefield. This includes a deeper understanding of the meaning inherent in seemingly simple communications (e.g., "I am going to flank to the right"). In essence, as teams get better at working with one another, they develop shared schemas and scripts for responding to situations. This allows them to coordinate with less need for communication. In contrast, working with new or inexperienced team members requires more communication and clarification.

2.3.5.3 Principle-based Operation

Current doctrine might be more flexible (open to interpretation) than it was in the past. Military teams develop their own tactics, techniques, and procedures (TTPs) and SOPs (a.k.a. way of doing things) over time through training according to their specific mission set and team makeup. Under the guidance of TTPs and SOPs, there is tolerance and room for the squad crew in a vehicle to work out the most effective way to communicate among themselves.

2.3.5.4 Team Autonomy is Highly Favored

In order to contend with the complexity on the battlefield, the SMEs expressed that they would convey the c*ommander's intent* regarding the current mission or situation to subordinates and would push an initiative for local decision making to subordinates whenever feasible and appropriate. This allows subordinates to take quick and decisive actions without appealing to the platoon leader for every decision. Additionally, control measures, contingency plans, SOPs, TTPs, and battle drills are used to enhance team autonomy. If at some point radio communication is lost, a good team that has interoperability would still know what to do next to complete the mission. To this end, team members are expected to be very familiar with contingency plans developed in the planning process and able to react instinctively to some situations due to extensive practice of battle drills. Battle drills are used to limit the number of things one needs to plan for deliberately but rather give one a set of actions to follow in specific situations (e.g., closing eyes while putting on a gas mask or following manual operational procedure when autonomy malfunctions and cannot be fixed).

2.3.5.5 Automation Level of the Robotic Combat Vehicle

The technology level of the RCV also influences communication as does the human interface to those automated functions. For example, the RCV's target recognition functionality and its accuracy, capabilities, and transparency can influence the gunner's interaction with the RCV; self-driving functionality influences the driver's interaction with the RCV. In addition, whether the RCV can execute contingency plans also influences the communication and coordination intensity in contingency situations. Though high autonomy of the RCV may be favored in some situations so that less communication is needed to function, target engagement is still likely to require a human in the loop for the foreseeable future to avoid costly consequences before the decision automation is perfectly reliable.

2.3.5.6 Significant Content Items to Report to the Commander

With the expectation of team autonomy and reduced communication, the SMEs listed a few principles regarding significant content to report to the vehicle commander. First, contact reports refer to enemy identification and typically include a brief description, the distance, and direction of the enemy contact (e.g., "contact, two enemy tanks, 1500 m to the west!"). Second, SALUTE is an acronym that stands for size, activity, location, unit identification, time, and equipment. A SALUTE report is typically sent a few minutes after the initial contact report. Third, any situation that prevents the person from continuing a task. These principles are helpful, yet a detailed list of examples that are frequently reported, such as confirmation of reaching a point and set, will need to be developed to flesh out the communication model and develop metrics.

2.3.5.7 Trust in the Information

One SME mentioned that when part of the information from the computer system turns out to be confusing or incorrect, they would turn off or ignore the source entirely to ensure that what they work on is the correct information. False information and uncertainty cause extreme discomfort and workload during decision making. These are factors that will be critical to examine in future contexts, especially when the amount and sources of information are expected to grow exponentially and cross-checking every piece of information is not always feasible. This may force users toward working more at the source level, rather than the information item or feature level, when dealing with trust.

2.3.5.8 Seating

With uncertainty in the NGCVs seating arrangement for the MCV, one SME expressed the preference to sit in the front if there is an advantage by being able to see outside. This was critical for a platoon leader to develop good SA. But when the front is sealed up with no outside view, it was suggested placing the platoon leader at the back of the MCV because the back is the safest place to protect the highest-ranking person on the team. Even sitting in the back, there should be ways to provide access to see the outside, such as hatches, especially when cameras fail or have a limited field of view. Crew members with highly interdependent tasks should sit next to each other because physical proximity fits human mental model of relatedness of team members' tasks and provides a contingency communication channel (e.g., facial expression and body gestures, talking loudly) in case of technology failure.

2.3.6 Innovative Communication Approaches

NGCVs may carry many novel features to improve human–autonomy communication via user-friendly interfaces and features such as automation transparency. For example, when two or more communications use the same modality (e.g., perceiving two types of information through audio messages simultaneously), they tend to fight for mental resources more than having two different modalities (e.g., one in audio and the other in visual). To mitigate the mental workload of processing similar information, novel communication methods may be helpful, such as adding tactile force and vibration for the road surface (Corbett et al. 2013).

For another example, the RCV operator will need to operate and interpret the RCV input and actions efficiently and accurately and report RCV information correctly to the team. In case of technology failure, such as some camera system(s) malfunction or are destroyed or disabled during the mission, there might be a backup viewing system to resume the operation of the RCV to prevent losing the whole RCV simply because of the loss of the camera system. As part of team resilience strategies, it might be worth exploring whether crew members can get in a disabled RCV and drive it back to a safe area manually.

2.3.7 Limitation

Battlefield communication is complex. The SMEs provided rich experiences and feedback from the platoon leader's perspective, but less information was available from the driver or gunner's viewpoint regarding the communication issues for RCV. This is partially because existing combat vehicles are operated from within the vehicle rather than from outside, and it is unknown how exactly the teleoperation of the RCV works and what automation capabilities it has at this stage.

2.4 Future Directions

The background research and SME input on task, information, and interactions provided the team a better understanding concerning the actors, tasks, information, and interactions. This analysis also raised a set of topics for discussion moving forward when considering the evaluation of HAT in NGCVs in terms of human–autonomy interfaces and communication approaches. These topics are introduced in the following sections along with the literature that informs them.

As technology advances, future HAT interactions may be fundamentally different from existing ones. Therefore, the primary teaming tasks are more important than the specific tools and interfaces in use. This project utilized an iterative and interactive process where multiteams adapt and evolve by learning from each other. The steps that followed were to 1) develop an interaction taxonomy for possible HAT interactions (Section 3), 2) propose interaction measures appropriate for critical tasks and HAT (Section 4), and 3) use empirical data (Brewer et al. 2019; Schaefer et al. 2019) to examine the communications taxonomy and implement the measures to create metrics.

3. HAT: Interaction Taxonomy and Strategies for NGCVs (Task 2)

Building on the results of SME interviews, Task 2 was designed to further explore the interactions, discuss the factors that define the interaction context, identify essential interactions, and propose example interaction strategies required by changes in the anticipated NGCV environment. The outputs are an interaction taxonomy and NGCV HAT interaction strategies.

3.1 Interaction Taxonomies

An interaction taxonomy is a thorough analysis of interactions involved in HAT, and here we specify the taxonomy in the NGCV context. The interactions are less about the literal physical interface between a human and an autonomy and more about team configuration and role assignment for effective interactions and coordination. Thus, this work included some low-level interactions but aims to develop an interaction taxonomy based on a more macro and high-level view, such as each team member's interdependent tasks, information exchange approaches, and entities involved in the information exchange. An interaction taxonomy is essential in developing metrics to measure the frequency of information exchange, the communication contents, and the communication flow (who is talking to whom). Examples of relevant taxonomies and uses are provided in the following.

First, based on a review of taxonomies in the past decades, Beer et al. (2014) argued that a taxonomy could provide a framework for examining levels of robot autonomy: from manual control to decision support to full automation. They further presented a high–level taxonomy of possible levels of automation where interactions between a human and an autonomous agent are described. For example, in decision support, the autonomy would communicate to the human what decision options are available and then implement the decision made by the human.

Second, a taxonomy can also help improve understanding of how robots, or sets of intelligent agents, interact with one another (Dudek et al. 2002). According to Dudek et al. (2002), important factors to include in this type of taxonomy include communication, learning abilities, and path planning. Critical to our work on HAT, their communication taxonomy divided robots into three groups: those that cannot communicate, those that can communicate to only robots within a certain radius, and those that can communicate with any other robot. They further proposed ways in which robots may communicate with one another regarding hierarchical, partial, and complete communication networks.

Third, another taxonomy of human–robot interactions considered the composition, capacities, and interactions of the human–robot group (Yanco and Drury 2004). Notably, they included an examination of shared interactions in robot control. This included combinations of single and multiple robot systems with either one or multiple operators. However, their taxonomy did not closely explore the team interactions between individual operators, or larger human–autonomy teams.

Fourth, more directed at HAT, Save et al. (2012) describe an extensive taxonomy that summarized four generic cognitive functions: information acquisition, information analysis, decision and action selection, and action implementation, each striated by increasing automation levels (Save et al. 2012). Furthermore, they proposed three HAT principles: 1) a system cannot just have one level of automation; 2) an automated system can support more than one function thus, having more than one level of automation; and 3) automation being analyzed in support of human performance is not just a technical improvement but impacts how a person is supported in a task and how that human performs that task. They stated two purposes for taxonomies used in human–automation design: 1) support design choices early on regarding the optimal level of automation to best support team performance in a joint human–machine task and 2) put forward specific human factors related recommendations by classifying examples of automation.

These taxonomies provide an initial foundation for understanding human–autonomy team interactions that may be critical for military operations. Our taxonomy expands on this foundation to provide a more detailed, domain-specific interaction taxonomy for human–autonomy teams. From this, we are able to recommend team interaction strategies supporting NGCV crews.

The future RCV is anticipated to have a range of capabilities to maneuver in various environments and perform with degraded sensing capabilities. Many of these capabilities may be automated, including driving, decision support, and target detection or recognition assistance. Insight from academic literature on the limitations of autonomous systems (Woods et al. 2004) suggests that the RCV may be less capable in more context-sensitive activities, such as detecting inconsistencies in the environment and anticipating threats.

However, there are also areas where the RCV may be more capable than humans. For instance, the capacity to rapidly sense the environment and update this information on shared digital displays (Brewer et al. 2019) may be considerably faster for artificial agents. Additionally, the RCV may communicate with other artificial agents with a level of efficiency that humans cannot parallel. Critical to successful use is understanding the competence boundaries of the RCV—where the RCV will perform reliably on its own and when human intervention is necessary. These capabilities and limitations are important to consider in how they will structure the HAT interactions, as well as system design and the TTPs developed and employed.

One concern with these future systems is the need to track additional vehicles as a result of anticipated increases in the number of vehicles per section, including the addition of unmanned aerial vehicles (UAVs), and the potential complexity of coordinating both air and ground vehicles. This is particularly challenging for the platoon leader, who is responsible for maintaining SA of both sections. Multiple radio nets may all be simultaneously active during a fast-tempo situation (e.g., engagement or replanning), during which the platoon leader must integrate disparate information and plan. Therefore, there is a need to manage workload introduced by a stretched span of control and increased available information volume, particularly during high-tempo situations.

3.1.1.1 The Battlefield Environment

The battlefield environment is complex. Mission, enemy, terrain and weather, troops and support available, time available, and civil considerations (METT–TC) are a set of doctrinal mission variables that are used to focus analysis during mission planning and execution, and also help to provide context (US Army 2019):

- *Mission*. The mission is generally dictated by a higher-level commander to gain or maintain a desired effect on the battlefield. Key elements of the mission include the specific task that needs to be accomplished by the unit, the purpose, or intent of the mission, mission constraints such as maneuver boundaries and externally dependent execution times, and the mission of other units.
- *Enemy*. Key elements of the enemy include the enemy's composition, disposition, strengths, recent activities, ability to reinforce, and possible courses of action.
- *Terrain and weather*. Key elements of terrain and weather include observation and fields of fire, avenues of approach, key terrain, obstacles, cover, and concealment.
- *Troops and support available.* This variable represents combat power and the ability to sustain it over time. It includes the status of Soldiers and equipment, logistical support such as fuel, capacities of individual Soldiers, and equipment (including vehicles).
- *Time available.* This describes the time available due to internal and external dependencies and approximate time required to execute anticipated actions.
- *Civil considerations*. Civil considerations are the influence of manmade infrastructure, civilian institutions, and attitudes and activities of the civilian leaders, populations, and organizations within an area of operation on the conduct of military operations (US Army 2019).

3.1.1.2 Commander's Intent

In the battlefield environment, it is often necessary for actors to make decisions independently of their remote supervisors. The concept of *commander's intent* describes how collective goals are pursued despite the decoupling of the commander's view and pre-specified plans with local actors' access to immediate environmental information (Shattuck and Woods 2000). Commander's intent is generally conveyed as a clear and concise expression of the purpose and desired end state of an operation in order to provide focus and allow subordinates to act in the absence of further orders (JP 3–0), though the concept has also been used to refer to a convergence and focus of efforts in the absence of an explicit central authority (Alberts 2007). Novel interaction strategies are necessary to balance the initiative provided to agents on the battlefield, and their ability to fulfill the commander's intent in a predictable and appropriate fashion (Holder 2017).

Team interactions take many forms in sociotechnical environments. In a vehicle crew, information could be conveyed on a shared display, radio channel, or in physical space. These interactions can have different properties, such as meanings (e.g., status update, request), mechanisms (e.g., verbal behavior, touchscreen displays, auditory alerts), exchange structures (Chiou et al. 2019), and communication flows (Cooke and Gorman 2009). Communication dimensions have inherent capacity requirements and constraints. The design of the vehicle and context (e.g., availability of outside view) can also drive or constrain the strategies for providing information (e.g., no view implies reliance on a display for visual information).

3.1.1.3 Existing Strategies

A synthesis of existing interaction strategies in combat vehicle crews indicated three basic themes: efficiency, predictability, and adaptability. Examples of efficient and predictable coordination strategy implementations include closed-loop communication (STP 21–1–SMCT), standardization of tactics, phraseology (standard terminology, pro–words, call signs, radio etiquette), standardized reports and orders process, clear hierarchy, rank structure, and standardized roles and responsibilities. Examples of adaptive coordination strategy implementation include commander's intent, progressive elaboration in the planning process, and unit–level standard operating procedures (FM 3–20.15) (US Army 2019, p. 15).

3.2 Taxonomy Development

3.2.1 Taxonomy of Interactions

The HAT interaction taxonomy presented in this work is a combination of low-level interactions within the human–autonomy team (human–human, human–agent, and agent–agent) as well as team-level interactions and strategies. The taxonomy is split into three broad categories including *task, team composition,* and *communication*. Each category contains dimensions that can be used to characterize interactions. The taxonomy can be thought of as a "toolbox" with which to characterize team interactions and implement interactions strategies.

3.2.2 Task Dimensions

To develop the task category, a task analysis was performed on the tactical task *MTC*. We identified and categorized the tasks into the following levels of abstraction: *task classes, tasks,* and *subtasks*. This resulted in the identification of four task classes: *mobility, gunnery, actions on contact, and crew management*. Twenty-three tasks were identified. Abbreviated examples are provided here, with the full task analysis available in Appendix A. Table 5 shows each task class, its definition, and example tasks. The task analysis structure formed the base that the interaction analysis and taxonomy was built upon.

Table 5 Task class, definition, and tasks

Task classes	Definition	Tasks
Gunnery	The operation of sensors and weapon systems for detecting and identifying potential targets, making engagement decisions, executing, and assessing engagements (FM 3–20.21).	• Prepare and maintain weapon systems • Detect surroundings for potential targets • Identify targets • Determine target engagement method • Engage target • Call for indirect fire • Assess targets
Mobility	Tasks related to movement and navigation	• Prepare and maintain vehicle system • Drive • Navigate
Actions on contact	A series of actions, often conducted simultaneously in reaction to contact with the enemy (US Army 2019)	• Identify and report contact • React to direct or indirect fire contact • Develop situation and choose a course of action • Execute selected course of action
Crew management	The management of crew and nonhuman agent requirements for short term and sustained operations.	• Prepare/ maintain communications equipment • Monitor communications equipment • Monitor crew status • Provide or request medical support • Share mission • Mobility mode selection • Gunnery mode selection

Subtasks are lower-level tasks that are typically required to complete the higher-level tasks under nominal conditions. For example, the *detect surroundings for potential targets* task include subtasks *operate weapons*, *designate sectors of responsibility*, and *communicate to operate weapon*, and so. A total of 169 subtasks were identified in the task analysis. Table 6 presents example subtasks.

Table 6 Subtask examples

Task	Definition	Example subtasks
Detect surroundings for potential targets	The acquisition and location of an object in the operational environment (FM 3–20.21)	• Operate weapons, sights, and/or turret to sense surroundings • Designate sectors of responsibility to observe targets • Communicate to operate weapons, sights, or turret to sense surroundings • Communicate sectors of responsibility

Twenty-eight essential interactions spanning the entire task analysis were identified. Essential interactions were defined as interactions that 1) require interaction among two or more agents and 2) are essential to accomplish the higher-level task under nominal conditions. Failure to complete an essential interaction would be likely to result in negative consequences for the crew, such as missed targets, damage or mission failure. A sample of the essential interactions is included in Table 7.

Table 7 Essential interaction examples

Task class	Task	Essential interactions
Gunnery	Detect surroundings for potential targets/conduct surveillance	• Communicate sectors of responsibility • Communicate detection of targets • Communicate location of targets
	Identify targets	• Communicate target classification • Communicate target identity • Communicate target alignment
	Determine target engagement	• Communicate range to target • Communicate target engagement method (weapon, ammo, technique) • Communicate about the decision to engage the target
	Engage target	• Give command to fire • Confirm firing execution
	Assess targets	• Give cease fire command
Mobility	Drive	• Communicate local route
	Navigate	• Communicate vehicle location

3.2.3 Team Composition Dimensions

The team composition dimensions characterize the agents, their roles, and their relationships within the team—the team makeup and member relationships:

- *Agent types* identifies the agents involved in order to differentiate interactions among human and artificial agents. Agent types in this taxonomy include artificial and human.
- *Role fluidity* describes the level of rigidity regarding roles and responsibilities. Classifications of role fluidity are differentiated among high, moderate, low, or none. Traditional armor crews are relatively rigid, though some shared roles are exhibited (e.g., both gunner and tank commander can control the main gun). The NGCV systems are anticipated to have higher role fluidity.
- *Human–vehicle ratio* describes the number of operators assigned control or coordination responsibility for an unmanned vehicle. For example, if two operators are the primary responsibility holders (monitors) for one RCV the ratio is 2:1 (adapted from Murphy and Burke [2016]).
- *Agent formal role.* Formal roles describe sets of permissions and responsibilities officially sanctioned within the organizational structure, typically identified in relation to a formal title (e.g., vehicle commander, driver, gunner, operator, and so on). Examples are included since the specific formal roles are not yet in existence for NGCV systems (Table 8).

Table 8 Example formal roles

Agent formal role	Responsibilities
Vehicle commander	Serves as supervisor of the section by coordinating all vehicles and crew members to achieve tactical objects
RCV mobility operator	Navigates and maneuvers the robotic combat vehicle while avoiding hazards
RCV gunner	Responsible for operating the primary weapon system on the RCV in order to search for, properly identify, and destroy enemy targets.

- *Agent task role* describes the team-level roles taken on by an agent during a specific task. Roles are separated into the simplified cognitive roles of sensor (sensing and interpreting), decision maker (judgment and decision making), and effector (acts on the environment) (adapted from Russell and Norvig [2016]). This is differentiated, though related to, the agent's formal role and may take place over much shorter time scales. An example is shown in Fig. 14. Agents can take on more than one task role simultaneously.

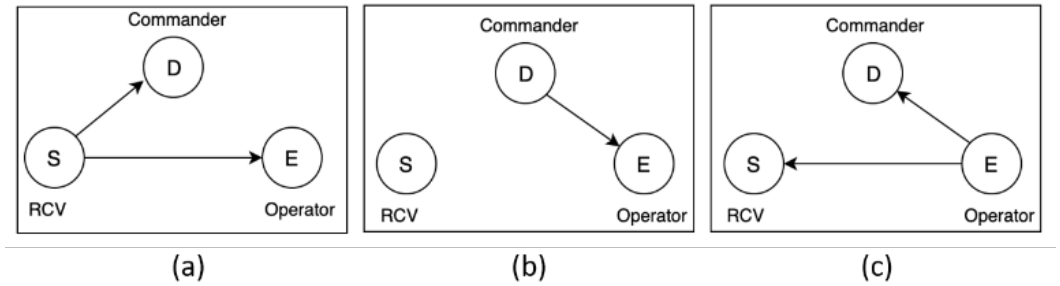

Fig. 14 Agent task roles example, where S = sensor, D = decision-maker, E = effector, arrows indicate interactions between agents. a) RCV operating in semi-autonomous mobility mode senses an obstacle in the environment and notifies the commander and operator. b) The commander confirms that the object is an improvised explosive device (IED) and instructs the operator to initiate evasive action (decision). c) The operator acknowledges the commander and manually operates the RCV in the environment to avoid the IED (effects).

3.2.4 Communication Dimensions

Communication dimensions characterize the elements contained in communication between team members. *Communication type* differentiates between explicit communication (clearly communicated) and implicit communication (indirect or implied):

- *Media* describes the device, technology, or artifact that is utilized to interact. An example would be a vehicle intercom system (Hollan and Stornetta 1992). Media considerations are important as they describe the actual tools that are implemented to support agent interactions.
- *Mechanism d*escribes the interactions taken to meet communication needs that are enabled by a medium such as text chat messages (a mechanism) displayed on a computer screen (media). Mechanisms are tied to media, but a single medium may be able to support a variety of different interaction mechanisms such as graphical map features and text messages on a single computer screen (Hollan and Stornetta 1992). Conversely, similar mechanisms (e.g., voice commands) may be involved in different media. We did not use modes of communication here because modes seem to include both the media and the mechanism. A shortlist of example communication mechanisms corresponding to media is shown in Table 9.

Table 9 Example mechanisms and media

Modes	Media	Example interaction
Voice communications	Intercom	Vehicle commander gives driver driving instructions.
Graphics	Graphical user interface	RCV sensor places potential obstacle on a digital map.
Text message	Computer display	RCV displays text-based information about statements, warnings, cautions, and alert related to the communications with the vehicle and state of the vehicle (e.g., autonomy mode selected, and so on).
Gestures	Body parts	Fingers showing an ok sign to confirm.

- *Modality* concerns the cognitive resource requirements for an interaction such as "auditory" modality for spoken word over a radio (Wickens 2008). Further granularity can be achieved by examining the additional dimensions of *code (*spatial or verbal) and processing *stage* (perception, cognition, responding). Modalities in this taxonomy include visual, auditory, vestibular, and tactile, while the encoding of the information can take the form of spatial or verbal information (Table 10).

Table 10 Example interaction modalities and codes

Modality and code	Example
Visual–spatial	Identifying unit icon on the digital map
Visual–verbal	Receiving text communications
Auditory–spatial	Hearing a directional alarm
Auditory–verbal	Hearing commands from vehicle commander
Vestibular–verbal	Feedback that MCV is turning in response to a command to the driver
Tactile–spatial	Tactile feedback on teleoperation system indicates loss of RCV traction

- *Communication flow* describes who communicates to whom (Cooke and Gorman 2009), defined by the sender and receiver. An example would be an RCV mobility operator talking to the vehicle commander to share some information. Flow may include one-to-one communication or other combinations involving multiple agents. An example would be dispersal (e.g., broadcast) of the location and direction of a possible enemy by one member of the crew to the rest of the crew, including digital formats available to intelligent agents (Fig. 15). This taxonomy includes the flow categories of one-way, exchange, transfer, dispersal, and consolidation to facilitate examining the impacts of flow on performance and the HAT environment. Research is needed to find out which type of communication flow is better than the other and in what conditions. *Communication contents* describe what people communicate about. Other researchers have also examined communication patterns (Baker et al. 2019).

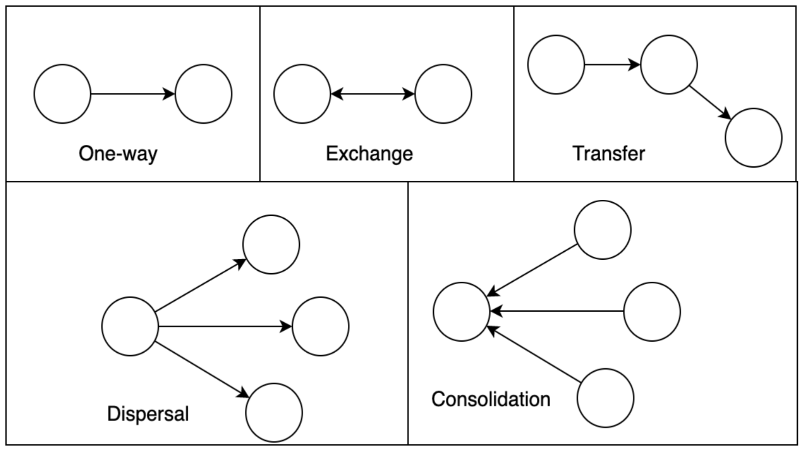

Fig. 15 Examples of communication flow

3.2.5 Interaction Strategies

Although there are numerous instances of ways in which research has incorporated specific intentions to structure team interactions (Fussell et al. 1998; Johnson et al. 2014; Chen et al. 2018), the concept of a team interaction strategy remains relatively undefined. In this work, we define a *team interaction strategy* as the specification of some properties of team interactions (e.g., how, when, with whom, at what times) to achieve one or more objectives. A benefit of focusing on team interaction strategies is that they exploit specific qualities of team interdependence, allowing them to generalize better across similar teaming contexts. For example, Patterson et al. (2004) identified 21 role handoff strategies across various safety-critical domains to compare to processes in observational healthcare data. They were able to provide suggestions for improvement in the handoff process within the healthcare context, including training and interface design. This approach is particularly useful for systems that have not been fully defined and developed, such as NGCVs.

An interaction strategy specifies a general framework for interactions that achieve one or more objectives under particular constraints. Interaction strategies can be executed using a variety of different means and can be broad or specific depending on the objective. Team interaction strategies in this study include assumptions about the anticipated interaction requirements for team effectiveness in NGCV crews. The future environment may impact teams at smaller (e.g., one operator and one vehicle) and larger (e.g., multiple crews) scales. Thus, team interaction strategies should address coordination needs at low levels of organization as well as from a macro level. These strategies are also grounded in team states and outcomes, including appropriate trust, SA, manageable workload, and resilience. An ineffective strategy may increase awareness of status information in ways that reduce trust or increase workload. Alternatively, a strategy may only be viable when a team is already working well together or lose viability in the context of competing priorities. Thus, consideration for the tradeoffs and dynamic context of these strategies should also be given.

Finally, as strategies provide structure to interactions, some insight into the possible interactions is needed for these strategies to be developed. They may be derived from an analysis of essential interactions involved within the defined scope of application as described in Sections 2 and 3. Team interaction categories and dimensions discussed in the taxonomy helped characterize the possible interactions and are determined by the composition of the system/team and constraints imposed by the task environment. Although we focus on interaction strategies for NGCVs with unmanned ground vehicles, the strategies in this work are presented at a level of abstraction intended to be appropriate for implementing in a variety of human–autonomy vehicle crew systems.

The team interaction strategies that we propose are based on identified objectives for effective HAT in the context of changes anticipated in NGCVs. The NGCV section concept introduces notable shifts from existing systems. Increasing interactions with artificial agents suggest vehicle operation may adopt different forms of control (e.g., supervisory control). Humans have extensive experiences interacting with humans but not necessarily with the various intelligent agents of NGCVs. Humans must have accurate expectations for effective team interactions. Therefore, one of our objectives for crew interaction strategies is *increasing awareness of teammate behaviors, roles, and responsibilities.* In addition, vehicle operation tasks may be conducted by artificial agents, freeing vehicle operators to perform other tasks. As operators become more available, the nature of tasking may change from more predefined responsibilities to task switching and on-demand tasking. In order to allocate tasks appropriately, agents assigning tasks should be able to observe the current task allocation of the team.

Invoking on-demand tasking alludes to the objective of *managing crew flexibility for changing conditions*. Although mission planning may be extensive, flexibility is needed to adapt plans in surprising situations and can allow wider optimization of resources at the potentially higher risk of confusion. These adaptations may include revisions to task allocation, courses of action, and reorganization of military assets. Such adaptations must be coordinated across agents in a vehicle crew, including artificial agents. For on-demand tasking to benefit flexibility and adaptability, coordinators must consider each agent's capacity to perform tasks when replanning under surprise. For instance, although artificial agents may offer more efficient sensing, they may fail to interpret anomalies or complex situations accurately. Thus, a human may be required to compensate for this gap or even override the system when necessary. For these reasons, vehicle crew members are anticipated to have many ways of coordinating the same tasks, which must be explored over time.

Finally, the constraints of the new working environment are anticipated to affect vehicle crew interactions. For example, the shift from manned control to remote operations has several implications. First, human operators will be physically distant from activity and direct sensation, relying more heavily on inputs from intermediary sources (e.g., agents, displays) whose location and perspective will be different from that of the operators. Feedback previously accessible passively (e.g., motion cues) and actively (e.g., using a hatch) via physical presence will need to be provided through other means. Second, improved sensing may afford increased capacity to observe teammates and receive situation information but may also introduce needs for information filtering to ensure that each agent obtains information relevant for them in a timely manner. A third constraint is the need to control additional vehicles per section, which may affect the span of control for each agent. Overall, these changes lead our final objective: *understanding and working within the constraints of the new environment.*

Based on the anticipated system changes and human system requirements, we proposed 22 example team interaction strategies for achieving these requirements and meeting the three objectives (Table 11):

- *Objective I: Increasing awareness of teammate behaviors, roles, and responsibilities.* The first objective driving strategy development was to improve awareness and the need to scope the relevant teammate behaviors, roles, and responsibilities pertinent for different agents and to increase the team's understanding of each other's strengths and performance limitations. We also identified the need for information about current task allocation, workload, and underlying causes of behaviors. Then, we determined a need for understanding which agents need what status information (e.g., operators, artificial agents, commanders). Finally, we identified a need for possible rules or pathways for exchanging the information, such as when information should be pushed or pulled or how information may get from an artificial agent to a specific team member. Overall, 13 strategies address this objective.

- *Objective II: Managing crew flexibility for changing conditions.* We identified the need for information concerning the conditions that demand performance variability, including environmental changes, changes in team states (e.g., workload), and unexpected events. Then, we considered how to define interactions that could proactively benefit flexibility by either increasing the range of possible coordination solutions, increasing awareness of when adaptation may be required, or facilitating the transitions between teammates. Finally, we considered how to define interactions that may occur in response to changing conditions, including on-demand tasking, replanning, or adapting communication to reduce demand. Eleven strategies were for improving flexibility.
- *Objective III: Understanding and working within the constraints of the new environment.* Finally, addressing the constraints of the new environment on teaming first involved identifying the need for information concerning changes in feedback and control. These include a need to coordinate more entities (e.g., vehicles, humans, and artificial agents), the shift from manual to supervisory control across multiple tasks, and potential limitations to control (e.g., remote control boundaries). Then, we brainstormed potential compensations for these changes, including increasing operators' initiative when making vehicle-level decisions and reproducing the feedback previously obtained through physical presence (e.g., motion cues). In total, 12 strategies address this objective.

Table 11 Interaction strategies

Objective	Requirement	Strategy
I	Each agent that depends on other teammate's input will need transparency regarding the underlying reasons or supporting information for behaviors of other teammates (Chen et al. 2018).	[1] Agents utilizing another teammate's input are provided supporting information as needed to understand and use the information.
I	Teammates need to understand which teammates need information updates and which do not.	[2] Each agent pushes relevant contextual information when that information is not easily observable by relevant recipients.
I	Teammates need to understand which teammates are likely to have information that they might require.	[3] Each agent obtains information about who will likely have required information before it is needed.
I	Humans must be able to distinguish actions and input as provided by humans vs. agents.	[4] The inputs of human and artificial agents are represented in a manner that is easy to distinguish from one another.
I, II	For crews with fluid roles, current task allocation status will need to be available to all crew members.	[5] The team's ongoing and upcoming tasks and overall workload demands are made observable when coordinating task allocation.
I, II	Exchanges in responsibility during role handoffs needs to be efficient and understood by each agent (Patterson et al., 2004).	[6] Each stage of a role handoff follows a clear and consistent phraseology with closed-loop take-over communicated to all relevant agents.
I, III	The human in the loop of RCV control will need to translate RCV inputs and perspective to his/her own and other teammates with different perspectives and orientations to the battlefield.	[7] Translating artificial agent inputs between humans follows a clear and consistent phraseology and common reference systems.
I, II. III	The performance limits for human and agents need to be clearly understood and observable for each agent.	[8] Agents receive feedback when their requests of other teammates are likely to exceed that teammate's performance limits.
I, II, III	Teams will need to calibrate trust over time as agents gain knowledge, skills, or change in condition.	[9] Agents are provided supporting information when their expectations of other teammates deviate from actual knowledge, skills, or states of those teammates.

Table 11 Interaction strategies (continued)

Objective	Requirement	Strategy
I, II, III	Acceptable conditions for task reallocation need be established in advance (workload limits, proficiency sets, etc.)	[10] Agents are provided information about their own or other performance limits (e.g., during mission planning, predictive analytics, in real-time). [11] Organizational efforts are taken to identify situational trigger conditions and response procedures
I, II, III	Operators will need to coordinate with RCVs in multiple ways to respond effectively to unexpected events (Gorman et al. 2010).	[12] Humans proactively explore multiple ways of coordinating with artificial agents. [13] The levels of autonomous support are negotiated between the RCV and operators over time.
I, II, III	Crews with remote vehicles will have to coordinate within the control boundaries of all vehicles.	[14] Remote control boundaries of robotic combat vehicles are observable to agents coordinating with those vehicles. [15] Agents involved in planning maneuvers obtain information regarding remote control boundaries and terrain considerations.
II	The human needs to be able to override the system when required and have the information required to get up to speed.	[16] Humans may override automated systems when required.
II, III	Responsibility for control and tasking of agents must be available at multiple levels of the human chain of command.	[17] RCV Operators may be designated to make decisions regarding tasks of RCV automated agents as situations demand. [18] Who is in control of an automated agent functionality should be clearly assigned and acknowledged
III	Remote or distributed systems will need to provide supplementary cues to operators in order to compensate for lost or reduced feedback.	[19] Information normally obtained passively via physical presence (e.g., motion cues) is also provided in remote operation. [20] Compare information and decision information requirements between legacy systems and current systems to identify needs are met.
III	RCV Operators will need to be able to operate vehicles remotely without access to direct environmental feedback.	[21] Information normally obtained via direct environmental feedback is provided during remote operations.

Note: Interaction requirements and strategies based on our three scoped objectives: I) increasing awareness of team members' behaviors, roles, and responsibilities, II) managing flexibility for changing situations, and III) understanding and working within the constraints of the new environment.

3.3 Discussion of Interaction Taxonomy and Strategies

We examined the team tasks in the mission of MTC to conceptualize NGCV crew communication in dynamic and complex battlefield operations. Rather than predicting and specifying the exact interfaces to be used, the interaction taxonomy and team interaction strategies proposed were abstracted to fit a variety of technological capabilities and crew structures.

When designing a system to support an interaction strategy, the interdependence of the system should be considered throughout the design process. Our strategies attempted to consider the broader context of military operations by specifying particular elements and remaining broad on others. Yet, interaction strategies taken out of context may be ineffective or counterproductive if implemented without care for system interdependence. For instance, several proposed strategies could be implemented over a particular communication medium, even though a considerable risk may be overloading sensory modalities associated with that medium. The ideal implementation of these team interaction strategies would consider such constraints in addition to other factors such as organizational context and other interaction strategies for how they might synergize or conflict with one another.

Operating unmanned vehicles also introduces some sensory difficulties, which may limit the feedback RCV operators have access to on the RCV and its local problem space. Interaction strategies are needed between the RCV and RCV operators to maintain adequate shared and team SA. Additionally, connectivity becomes a challenge as RCVs require connectivity to the MCV for both control inputs and information transfer and can have constraints based on operational range. Furthermore, as technology improves, the level of automation in human–autonomy teams may similarly increase. This may have implications for team interactions and require modifications to this taxonomy.

Implications. The task analysis and interaction taxonomy presented in this work will play a significant role in guiding and evaluating NGCVs to enhance effective HAT. Designers may use the identified essential tasks to design technologies to achieve team goals. Researchers may be able to adapt these tasks and essential interactions to measure mission success and teaming effectiveness.

Future work should use experiments to evaluate the effectiveness of applying the proposed interaction strategies to guide the understanding of how to implement and measure effective teaming. Furthermore, measures and metrics are needed to empirically test the effectiveness of the proposed interactions. Instead of aggregating individual performance measures, HAT should be evaluated at the team level (Cooke et al. 2013) and the system level (Gorman et al. 2019) to capture the team dynamics, such as team SA (Gorman et al. 2005) and system reorganization (Gorman et al. 2019). Future inquiries on these systems should also expand to include interactions between teams (e.g., crews, sections, platoons, domain, and so on).

4. HAT Measures for NGCV Concepts (Task 3)

Based on past work and interviews, Task 3 translates the interaction-based metrics of HAT effectiveness to the identified Army contexts. We proposed a framework for capturing HAT effectiveness in NGCV concepts, which consist of inputs, team interaction processes, team states, and multi-criteria outcomes. Because we are examining a dynamic and complex HAT context, we proposed interaction-based measures to capture team states according to the interactive cognition theory. The measures will be refined and tested to create performance metrics. These metrics can be used to guide the design of robot autonomy and the training of Soldiers to achieve effective teamwork. This section includes potential measures of HAT effectiveness that are appropriate for the Army context and envisioned scenarios.

4.1 Introduction

As technology advances the capabilities of artificial agents (e.g., autonomous vehicles, robots, and software agents), the agents will have great potential in enhancing Soldiers' safety and mission success in the future battlefield. However, one key to exploiting their potential and what is also most challenging is effective HAT in heterogeneous teams involving humans and agents. Defining and measuring team effectiveness in HAT will help guide the development of the capabilities and the design of interfaces of artificial agents, as well as the training of Soldiers. HAT is even harder when the context is fast paced and high stakes, and with high uncertainty. To this end, Section 4 aims to identify and develop measures to assess teamwork in such combat settings than can lead to defined performance metrics. This process views teaming effectiveness in terms of the inputs, team interaction processes, team states (SA, trust, workload, resilience) and multi-criteria outcomes.

4.2 Teamwork Effectiveness and Input-Mediator-Output (IMO) Model

Teamwork effectiveness should not be only reflected in one single outcome, but also in the impacts of, and interactions among, the inputs, team interaction processes, team states, and multi-criteria outcomes (Mathieu et al. 2008; Baker et al. 2019). Take cardiopulmonary resuscitation, for example, even if the rescue team works effectively, it is possible that the patient still dies due to the severity of the patient's medical situation. In other words, life or death should not be the only criterion to judge the effectiveness of the medical teamwork. The same conditions apply to combat contexts where any single outcome can be determined by factors that are not always controllable by the crew despite effective teaming. Further, wars are won in the longer timeframe and consistently using effective team processes is likely to show an overall impact on success. More importantly, the teamwork processes and team states may influence the relationship between inputs and outcomes (Fig. 16). In this IMO model, the mediators include processes and emergent states, and they are expected to mediate the relationship between the team's initial setting and their final outcomes. The input–process–output (IPO) model conveys a similar idea, but failed to capture the difference between the process and emergent states (Marks et al. 2001; Ilgen et al. 2005). Processes focus on the interactions, while emergent states focus on a variety of cognitive and affective team traits. This IMO model covers the essential components and the iterative nature of the Input–Mediator–Output–Input (IMOI) model (Ilgen et al. 2005), so it is simpler to have just three letters. In addition, it is easier to identify the sequential inputs, mediators, and the outcomes in the IMO model than the vaguely defined components of teamwork described in the Big Five model: team leadership, mutual performance monitoring, backup behavior, adaptability, and team orientation (Salas et al. 2005).

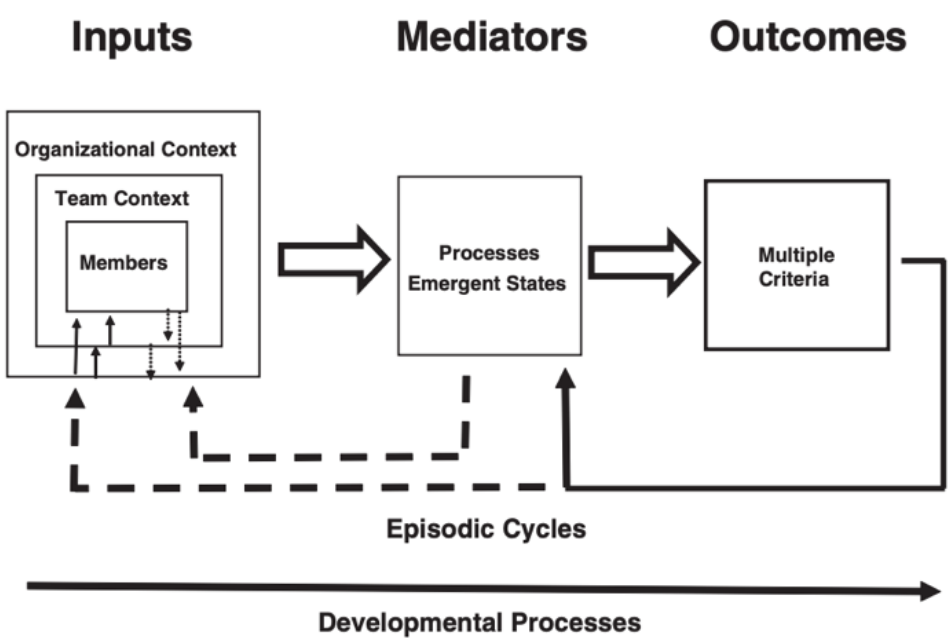

Fig. 16 IMO team effectiveness framework (Mathieu et al. 2008)

The IMO model is an operational way to examine HAT in the NGCV context. It can breakdown the vague concepts and operationalize them in the temporal and operational phases. For this NGCV project, we focus on developing metrics for the processes and emergent states in this team effectiveness framework, and it requires two instrumental pieces: 1) identifying critical subtasks and the interaction taxonomy and 2) measuring teamwork effectiveness through teamwork processes and team-level interactions. In Section 3, we have described the interaction taxonomy, involving seven crew members, two unmanned robotic combat vehicles, and a manned vehicle, with each pair of crew members controlling one vehicle (Johnson et al. in press). The following sections provide more details about our approach.

4.3 Interactive Team Cognition and IPSO Model

HAT is operationally defined as at least one human and one agent interdependently working together toward a common goal. Interactive team cognition (ITC) theory (Cooke et al. 2013) suggests 1) team cognition is an activity, not a property or a product; 2) team cognition should be measured and studied at the team level; and 3) team cognition is inextricably tied to context. Team interaction is team cognition. Therefore, team cognition and team states are indicated through interactions, such as communication and coordination (Cooke and Gorman 2009).

One example of using ITC to study HAT is through the uninhabited aerial vehicle–synthetic task environment (UAV-STE), a testbed that mimics the US Air Force predator ground control station (Cooke and Shope 2004). It allows researchers to analyze the interactions within a team, team SA, and team performance. Findings showed that team performance decreased when team members changed, but these new teams continued to form flexible and stable interactions. Thus, the newly formed teams continued to improve interactions after the retention period compared to teams that stayed intact after the retention period (Gorman and Cooke 2011). Intact teams were rigid with their interactions in that their interaction techniques did not change after the retention period even when there were novel events. Further, when it comes to disruptions or perturbations, new teams showed more flexibility regarding coordination patterns due to flexible interaction techniques thus, displaying better team performance in conditions in which disruptions and increased workload were present (Gorman et al. 2010). Team cognition, as measured by interactions (determined by task analysis), can capture and characterize the factors that lead to differences in team states and outcomes, when teams face novel situations and disruptions. These are all inherent characteristics of the environment of Soldiers and semi-autonomous vehicles interacting in a battlefield scenario (e.g., MTC).

A related but different team concept is shared mental models, which is a static measure of team cognition (Cooke et al. 2013). A mental model is an individual's stored representation of the environment, which allows the individual to describe, explain, and predict the environment (Rouse and Morris 1986). Shared mental models can only capture a snapshot in time, whereas interactive team cognition is constantly adapting and more accurately captures the team dynamics. Further, shared mental model approaches assume that the mental model of a team is equal to the sum of the individuals' mental models that make up the team. However, team members have heterogeneous mental models as required by their task roles and background. It is more important to understand how these heterogeneous team members effectively interact with each other across different positions than the teammates sharing a large portion of the same information.

Therefore, the team interaction processes and team states do not necessarily only mediate the relationship between the inputs and output, but they themselves may also influence each other and have a different relationship with the outputs and the iterative inputs in the next cycle. To capture both the interaction processes and their emergent team states, we modified the IMO model (Mathieu et al. 2008) to propose this Inputs–Team Interaction Processes–Team States–Outcomes (IPSO) framework of HAT effectiveness (Fig. 17). In this new framework, the teamwork processes may mediate the relationship between the inputs and the outputs, the team states may mediate or moderate the relationship between the interaction processes and the outcomes, the interaction processes and team states may influence each other, and team states may or may not have direct impact on the outcomes. Each of these relationships in the framework requires empirical testing, and thus appropriate measures of each of these constructs are important to help with the verification.

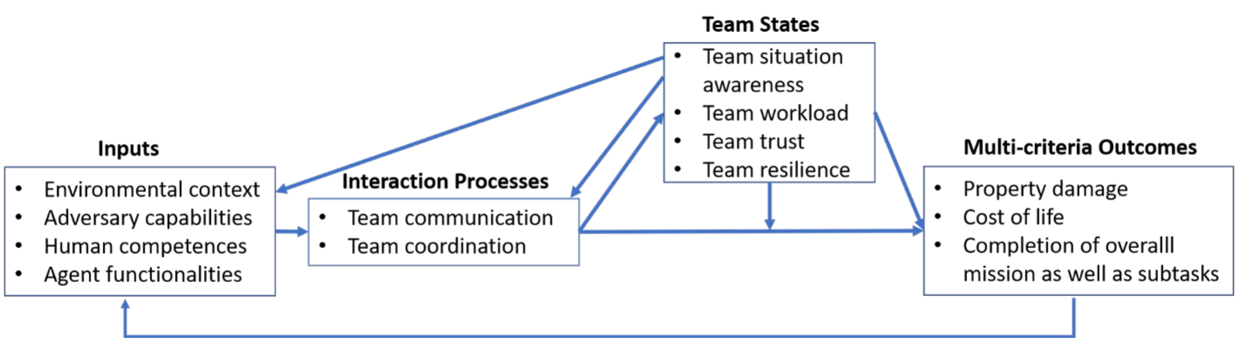

Fig. 17 IPSO human–autonomy team effectiveness framework

Inputs include environmental context, adversary capabilities, human Soldiers' competences (e.g., knowledge, skills, and attitudes), and agent functionalities (e.g., automation level). Inputs do not result in the outcomes directly but through other variables.

Team states are variables that characterize the team's mental and physical conditions in performing tasks. Depending on the effectiveness of teamwork interaction processes (e.g., team communication, team coordination), the inputs will result in different team states (e.g., team trust, team workload, team SA, and team resilience) and likewise these team states may impact the teamwork interaction processes. For example, teammates may only interact with trusted teammates or they might change interaction patterns based on workload. Effective teamwork interactions may result in good team states, poor teamwork interactions may result in bad team states. Team states can also impact the inputs (context) and either mediate or moderate the relationships with outcomes. Therefore, these teamwork interactions are interconnected with the relationships between the inputs and team states and outcomes. At the team level, existing research has mainly focused on humans. However, as the intelligent functionality of the autonomy increases along with interdependent tasks, it is vital to consider the states of the autonomy and its interactions part of the team states and interaction processes. Thus the concept of team states will also need to expand to include operationalized state indicators of the autonomy. Humans and autonomies could have common states and unique states, and this work focuses on the states that both humans and autonomy could have. This work also seeks to better clarify the relationships and constraints by looking at teamwork interaction processes and team states as two different categories and as they relate to each other as well as inputs and multi-criteria outputs. The following section introduces the components of this framework in further detail to identify appropriate metrics.

4.3.1 Inputs for HAT Effectiveness Framework Context

In a combat context, team inputs refer to team composition (i.e., who is on the team, number of humans and autonomies), the initial condition and capabilities of the team members (e.g., Soldiers' knowledge, skill, and attitude), and autonomy' functionalities (e.g., automation level and functions; Table 12), initial external conditions such as assigned mission and environmental factors (e.g., terrain, locations of landmarks), and even the enemy capabilities (Spiker et al. 2007). The team inputs set the stage for team interaction processes, team states, and team outcomes. In other words, team inputs may be used to predict the team interaction processes and team states and their team outcomes in some way. Therefore, we could evaluate each part of the inputs to determine and predict the team effectiveness. The key is to further breakdown the key components of inputs.

Table 12 Example types of artificial agents

Type of agents	Example studies
Decision aid	• Path-planning assistant (Huang et al. in review) • Planning decision aid agent (Chen et al. 2018)
Autonomous vehicle	• Brewer et al. 2019
Search and rescue robots	• Bartlett and Cooke 2015
Virtual intelligent tutor	• Lester et al. 1997

An autonomy (or an intelligent agent) is often referred to as an intact entity, such as the whole RCV as one agent, rather than treating a RCV as multiple agents based on their different autonomy-enabled systems. However, an intelligent agent is different from a human agent due to a lack of central neural control and thus should be modularized according to its functions. Each human is one agent because all the body parts and corresponding voluntary functions have one and only one central neural control for decision making. In contrast, an artificial agent consists of different parts designed by different people, and each module may function independently under more than one central control system. Therefore, unrelated models should be treated individually according to their functionality, though in some cases, some modules can be connected (e.g., designed by the same group of people, or dependent on each other's functions), with one extreme example being HAL9000 with one central decision control in the movie *2001: A Space Odyssey*.

This principle of treating functioning models separately aligns with the concept of high specificity in trust calibration, which suggests that calibrated trust in an agent should match the expectation of the functions with the actual functions of an autonomy component (Lee and See 2004). If an autonomy entity has multiple functioning modules like a mosaic picture of different elements, calibrated trust requires operators to treat each module with matching expectations (e.g., I will use the automated driving component on marked roads but not the object recognition in low light without strict supervision). However, the inferences between the modules should also be considered depending on their connectedness. When asked about the whole vehicle as an agent, the component with the most salient impression may be used as a heuristic to represent the whole (Gigerenzer and Gaissmaier 2011).

4.3.2 Teamwork Interaction Processes

HAT interactions roughly fall under the following categories according to the experimental manipulation: verbal communication (i.e., text or auditory dialog), visual interactions (e.g., gazing at images and pictures on the screen), tactile, and physical operations (e.g., pushing buttons and turning wheels in teleoperation, and gesture controls). Brain–computer interactions are possible (Jeon et al. 2011), but their development speed is expected to take more than decades for an operator to function at the novice level, so we currently do not consider it in our HAT project. Table 13 provides examples of these categories. In this work, we propose to use the interaction processes to analyze the team states. How these interaction modalities and interaction processes might be used as indicators of team states is explained in the next section on team states.

Table 13 Categories of interactions

Type of interactions	Variables studied
Chat text	Communication, coordination, coordinated SA (Gorman et al. 2005; Cooke and Gorman 2009)
Radio/intercom	Workload, SA, time to identify targets, number of communications (Hutchins et al. 2010)
Pressing buttons	Task completion time; preference (Guo and Sharlin 2008)
Body gesture	Task completion time; preference (Guo and Sharlin 2008)
System log (e.g., mouse click)	Workload, operation strategies (Gao and Cummings 2012)
Eye gazing	Trust (Meyer et al. 2014)

4.3.3 Team States

Team states are defined as the physical and mental conditions of teams at any moment during a mission. Some commonly studied team states include team SA (Gorman et al. 2005; Salmon et al. 2017; Schaefer et al. 2017; Stanton et al. 2017) and team resilience (Cooke and Gorman 2009; Gorman et al. 2010, 2005; Bowers et al. 2017; Hoffman and Hancock 2017). Using interaction-based metrics (see Table 13), we also propose to examine distributed dynamic team trust (Huang et al. in press) and team workload (Bowers et al. 1997; Funke et al. 2012). These team states are expected to influence the inputs, interactions, and various team outcomes, such as mission success. The following section explains each team state in more details.

4.3.3.1 Team Situation Awareness

SA is defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley 1988, p. 97). It is hypothesized that having good SA mutually influences effective team interaction processes (e.g., communication and coordination).

Given the prevalence of teams in modern sociotechnical systems, efforts have been made to understand SA on the team level (i.e., *team* SA) and system level (e.g., distributed SA; Stanton et al. [2017]). Team SA is defined as the extent to which each team member possesses the SA that is required for their responsibilities (Endsley 1995b). Practically, good SA should not require crew members to know all the elements in the environment but only the critical information required for the crew members to accomplish their tasks.

- Shared SA is often mentioned in the literature of team SA, but these two are different because shared SA focuses on the same SA between team members (see overlapping parts in Fig. 18), whereas team SA represents team members' distributed SA covering content that may or may not be the same for all team members (Endsley and Jones 1997). Figure 18a illustrates highly overlapping SA, while the team members in Fig. 18b only have shared SA that is required for their responsibilities. Comparing these two figures, the team in Fig. 18b covers a larger task area than that of Fig. 18a. Therefore, Fig. 18b is expected to be more effective for a heterogeneous team to share a high level of team SA but less shared SA, because each team member has different task requirements, thus minimizing the necessary overlap.

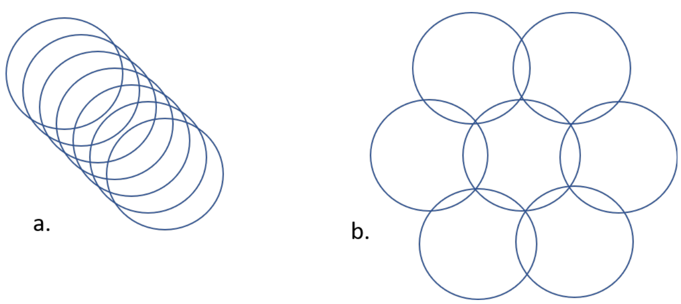

Fig. 18 Team SA. Each circle represents an individual's SA. The seven circles in a) represent highly overlapped SA and b) illustrates only necessary shared SA for the individuals' tasks to support others' responsibilities.

Measurement of team SA should reflect the two key elements of team SA in Endsley and Jones' (1997) definition: 1) each team member's SA should be included and 2) each team member's SA is required to fulfill their responsibilities, especially the tasks that require the team to work interdependency. Situation Awareness Global Assessment Technique (SAGAT) involves freezing a simulation or scenario of interest at selected times, blanking any displays or information sources, and asking participants SA probes about the situation (both system and external) at the three levels of SA. The answers can be compared to the actual state of the world to provide an objective assessment of accuracy (Endsley 2017). SAGAT was shown to have multiple benefits in capturing situation knowledge beyond the physiological techniques, performance measures, subjective techniques, and questionnaires (Endsley 1995a). However, SAGAT and other existing knowledge-based measures of team SA face challenges regarding the criteria of effective team SA (Endsley 1995a).

There are three additional issues that the SAGAT approach misses. First, not all knowledge is needed to complete tasks. Thus, it is important to examine each team member's tasks and relevant SA requirements. Second, team members do not need to know the knowledge during the entire mission to accomplish their responsibilities. In other words, team members may need the knowledge at different points of time when the tasks come up or some critical information needs to be pushed to all related parties immediately such as identifying an enemy. Therefore, using shared SA like a screenshot would miss the importance of timing and interdependence of tasks, also called the dynamic process of task flow. Further, for some tasks, it is not critical to have some knowledge in your head but rather to quickly know where it is located when needed or recognize it when it appears in the process of interactions. Based on Norman's seven stages of action cycle, people can use information feedback in the world (i.e., from the interface and environment) during their operations to complete their tasks without remembering all the detailed information (Norman 1988). The blank screen SAGAT approach does not account for these types of situations. Third, some SA may be too subtle to be recalled in words but plays a critical role in accomplishing the tasks. For example, we do not need to recall everything on a coin to know the value of a coin, but some cues at a glance (e.g., the shape and color) are enough to help us recognize the value of the coin to successfully complete transactions. Therefore, we use other team SA measures to complement SAGAT.

Other SA methods commonly used in the literature include the Situation Present Assessment Method (SPAM) and Situation Awareness Rating Technique (SART), and can alleviate some of the issues of SAGAT but these methods are currently tailored more to capturing individual SA.

SPAM (Durso et al. 1998) is an online, real-time probe method that was developed for use in the assessment of air traffic controller SA. The idea behind real-time, online probe methods is that they retain the objectivity of online freeze probe approaches but reduce the level of intrusion on task performance by not using task freezes. SPAM focuses on operator ability to locate information in the environment as an indicator of SA, rather than the recall of specific information regarding the current situation. In addition, when using scripted applications, scenario prompts can be tailored via agents or confederates to pull or push SA prompts into the team environment with minimal intrusion if tailored to the mission context (e.g., reports/requests from HQ), although they can focus Soldiers' attention on those items rather than what they would naturally attend to (Endsley 2017).

SART (Taylor 1990) is a simplistic post-trial subjective rating technique that was originally developed for the assessment of pilot SA. SART uses the following 10 dimensions to measure operator SA: familiarity of the situation, focusing of attention, information quantity, information quality, instability of the situation, concentration of attention, complexity of the situation, variability of the situation, arousal, and spare mental capacity. SART is administered post trial and involves participants subjectively rating each dimension on a seven-point rating scale (1 = Low, 7 = High) based on their performance of the task under analysis. The ratings are then combined to calculate a measure of participant SA. A quicker version of the SART approach also exists, known as the 3D SART. The 3D SART uses the 10 dimensions described earlier grouped into 3 overarching dimensions of demands on attentional resources, supply of attentional resources, and understanding the situation. This method is relatively quick and easy to administer but is again tailored more to individual SA and is reliant on recall and subjective biases.

To capture true team SA, one promising method is the *coordinated awareness of situation by teams* (CAST) (Gorman et al. 2005). CAST defines the necessary actions in an order that allows the introduction of obstacles, called roadblocks, to test whether the relevant crew members all have the correct SA, which is determined by whether they communicate it to the correct crew member to fix the obstacles. However, to construct a CAST instrument, the first concept to be operationally defined is the roadblocks—disruptive events requiring adaptive and timely team-level solutions. This method can be effective for pulling out team SA but is also limited in its focus on disruptive events because people may also miss critical situations in normal events. Further, CAST not only examines team SA through crew members' information coordination but also aims to identify the most effective way of coordinated team SA when encountering roadblocks. However, real-world operations often tolerate less-efficient ways of achieving team SA as long as the team task is successful. This makes the method potentially more applicable to training rather than real-world paradigms. This goes along with difficulties applying this method outside of the scripted experimental or training context to the complexity of real-world operations.

To overcome the flaws of these measures of team SA, we propose another approach, called incorrect SA in Failed Team Tasks (iSAFT). Incorrect team SA indicates missing or inaccurate team SA. SAGAT and CAST focus on what operators know about the situation and accuracy of this knowledge at a few selected moments; in contrast, iSAFT focuses on what operators did not know that they should have known during failed team tasks. This method has several advantages. First, it is more targeted at improving team performance by focusing on what went wrong and whether it was caused by incorrect team SA, while SAGAT and CAST do not differentiate failed tasks and successful tasks. Second, this approach may significantly reduce the total data size to be analyzed.

To conduct iSAFT, step one is to conduct a task analysis for the team members based on the planned scenario or task goals. This task analysis will identify the tasks that individuals can accomplish, the tasks that will need inputs from other team members, and the criteria of success (Table 14). After the trials, at identified failed tasks related to incorrect SA, we can use retrospective interviews, screen recording analysis, and system log analysis to explore the incorrect SA and its corresponding communication channels that caused the failure. Building on literature on identifying the task procedure and interaction modalities (Endsley and Jones 1997; Huang et al. 2019), Table 15 shows potential modalities used for team SA for NGCV concepts. The missed information can be categorized and counted for crew members to compare the frequency and types of incorrect SA (Table 16 and Fig. 19 for a notional example of the measurement method).

Table 14 Example of task analysis for TSA purpose

Team member	Hierarchical task analysis	Information needed from teammates	Examples of failed tasks
Driver	1. Task 1 (e.g., target recognition) a. Subtask 1 b. Subtask 2 2. Task 2 … 3. Task N	For each task, we will identify inputs from teammates	• RCV gets stuck in mud • Driver misjudged the hardness of the terrain • Gunner saw the mud-like terrain but with uncertainty and did not warn the driver quickly enough

Table 15 Team SA modality

Modality	Devices	Explanation
Visual	Nonverbal information from teammates	Seeing the gestures of the teammates
	Printed maps	Showing the geography of the area
	Direct observation window of the shared environment	Seeing the outside environment through the vehicle window
	Navigation watch or compass	Showing altitude and directions
	Digital text chat or reports	Showing important information from other teammates
	Screen monitors	Camera views of the surrounding (indirect views)
Audial	Radio/Intercom	Hearing messages from higher level and same level crew members
Haptic	Wearable haptic devices	Maybe showing threat detection (e.g., vibration near a bomb)
Vestibular	N/A	Natural sensor of speedometer, orientation

Table 16 Counts of incorrect SA types for each mission

	Interaction types for incorrect SA in Mission 1		
Crew ID	Type 1	Type 2	Type 3
A1	1	0	3
A2	1	0	0
B1	0	3	2
B2	2	1	0
C1	1	0	0
C2	3	0	0
S1	2	0	1

Note. Types are the categorizations of the incorrect SA based on expected information contents and sources.

- For position A1: type 3 (e.g., screen) caused most miscommunication.
- For position B1: type 2 (e.g., intercom) caused most miscommunication.
- For position C2: type 1 (e.g., vibration) caused most miscommunication.

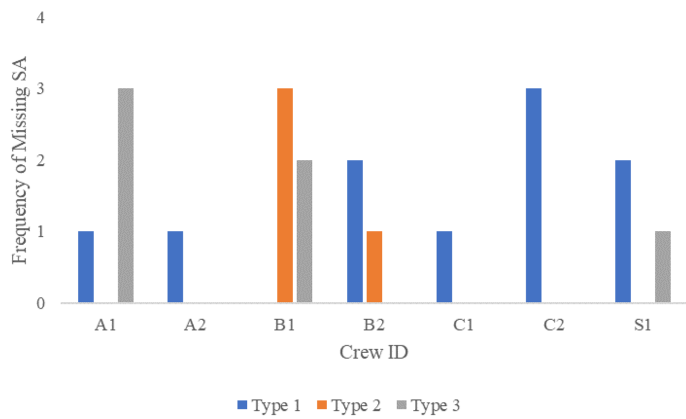

Fig. 19 Example of incorrect team SA during mission 1

iSAFT focuses on failures to get the right information to the right people. A limitation of iSAFT is its focus on poor team SA to the neglect of good team SA. To make up for that, a systematic analysis of good SA should be studied as a baseline for comparison.

Using interactive cognitive theory, the focus on interactions and communications does port easily across environments though to capture natural interactions in an unobtrusive manner that can be analyzed offline. The process of identifying the key tasks and information requirements also lends itself to a lengthier analysis of interactions for indicators of good team SA. These identified SA measures will be put through experimental testing for validation and refinement with the goal of identifying performance indicating patterns and metrics. These metrics will be developed in local testbeds and validated on NGCV scenarios during live or simulated NGCV events. Other criteria of team SA effectiveness may need to be examined as well (Table 17). These criteria are open for empirical examination, enrichment, and refinement.

Table 17 Team SA effectiveness dimensions and examples

Dimensions	Ineffective teams	Effective teams
Information checking	SA black hole: one member would lead others off	Self-checking: checked against others at each step
Information sharing	Did not share pertinent information: group norm	Coordinated to get information from each other
Prioritization	Failure to prioritize: members went in own directions; lost track of main goal	Prioritized: set up contingencies and rejoining
Expectations	Relied on expectations: unprepared to deal with false expectations	Be able to adapt to unexpected events

Note: Adapted from Endsley and Jones (1997).

4.3.3.2 Team Trust

Trust is defined as a person's belief of whether an agent will help accomplish the individual's goal in a situation that bears uncertainty and risks of failure (Lee and See 2004). Appropriate trust is critical in human–human teaming and HAT because it may impact human interaction strategies and teaming outcomes. This is especially true in the complex battlefield with both humans and agents involved in interdependent functional positions. Traditional trust measures have focused on individuals' global trust in the whole autonomy and employ a static view of trust, which should be updated to fit the complexity of HAT dynamics.

Most existing team trust measures focus on individuals' trust toward one entity (either human or artificial agent) and aggregate them. However, team trust is not an isolated phenomenon but a distributed and connected trust net at a team level. Each crew member may trust an artificial agent according to individuals' responsibilities and interaction experiences. Moreover, each crew member may trust other team members as well, and these attitudes may influence each other's trust in an artificial agent in a transitive way. For example, if Soldier A trusts Soldier B, and Soldier B trusts agent C, then Soldier A may be influenced by B (even by hearing comments by B concerning C or overhearing interactions between B and C and end up trusting C). From a team perspective, the distributed trust patterns of all related members on the net should be captured and analyzed to study team trust and how the patterns change over time. Assigning or removing a person from the team would be reflected in the team trust network. In an ideal situation, all involved stakeholders know the capabilities of one another and agents well, and their trust relationships would display as a fully trusted network (Fig. 20). If any of the stakeholders or autonomies malfunctions or a relationship shows a negative sign, it indicates the need to check for potential problems with the functionalities of the involved parties, or misaligned trust.

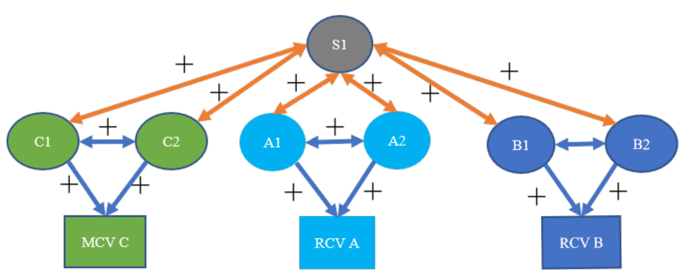

Fig. 20 One extreme example of a fully trusted network at one moment, as made possible through the trust network. S1= platoon leader; C1 & C2 = MCV operators; A1 & A2 = RCV A operators; B1 & B2 = RCV B operators. Arrow = trust direction, + indicates trust, and – indicates distrust; the gradient trust spectrum is not the focus here. In the normal situations, some of the arrows would be negative.

Further, team trust is not a static phenomenon but changes dynamically. The dynamic time scale could be seconds, minutes, hours, or days. However, trust scales are typically administrated once after the task or twice comparing before and after treatment (e.g., Jian et al. [2000]). Measuring team trust like a screenshot of team status oversimplifies the team state. The three-layer trust model (e.g., dispositional trust, situational trust, and learned trust; Hoff and Bashir [2015]) and their corresponding influencing factors for each layer of trust depict the dynamic nature of trust, especially situational trust. Each type of trust requires corresponding dynamic measures of trust. A multi-method approach to measuring trust is recommended, capturing each of these layers and in multiple ways as best fit the context and needs to provide a more complete trust picture (Schaefer et al. 2019; Baker et al. 2020; Milner et al. 2020).

Since human–human interactions and human–autonomy interactions do not occur all at the same time, in the same way, team trust may be demonstrated by phases (roughly divided by before, during, and after interaction) and the trust between different entities could use different measures depending on the interaction types. For example, before the team interactions, trust can be measured through questionnaires to build the trust net; controlled training environment may use obtrusive trust measures like pop-up questions or wearable devices (e.g., electroencephalography cap); fast-paced, real-life applications require nonobtrusive measures of trust through interactions (recorded communications, eye tracking, and interface inputs) that happen in their natural process. Together the data can roughly show the dynamics in one graph (Fig. 21).

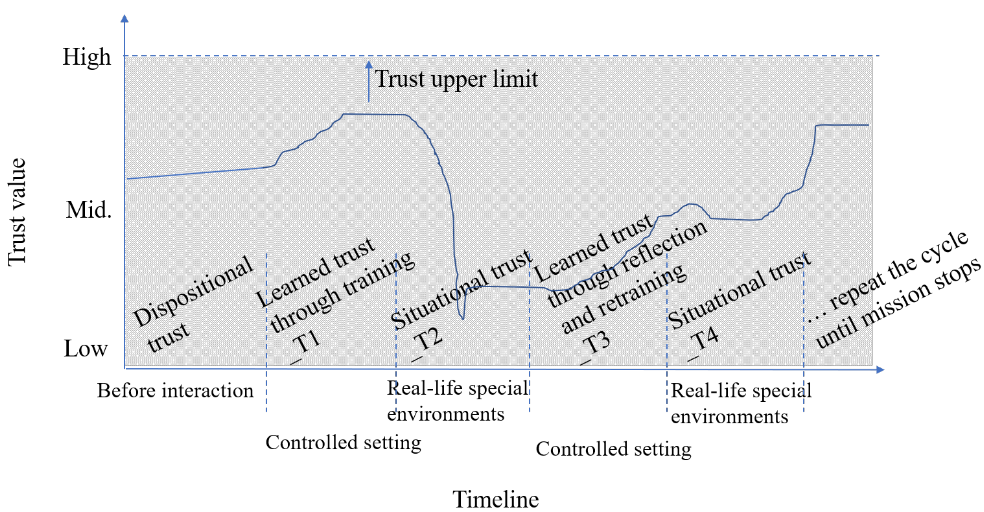

Fig. 21 A rough illustration of three-layer trust dynamics over time for each of the trust relationships. The vertical dotted lines divide phases of time that correspond to the three trust layers; T1–T4 are not time points but phases. The solid line in the shaded area represents one example of trust variation in different situations.

Dispositional trust is measured through questionnaires. The learned trust could be assessed through surveys and interaction-based measures. Situational trust is measured through interaction-based measures. Next are different types of interaction-based measures for team trust according to the ITC theory.

A verbal-communication-based analysis of trust may examine the communication content patterns, such as repeated checking whether the agent has done the assigned tasks. For example, "Did you check monitor A? Did you check monitor B? Remember when you hit that point, do this and that." These communication patterns can be compared to a control group interacting with a trusted entity (Table 18) and also explored in terms of what level of trust calibration leads to the optimal level of team effectiveness in various contexts. To determine the thresholds of the communication contents and frequency, we will also need other validation methods, such as their usage patterns of the autonomy and the performance of the team.

Table 18 Examples of trust analysis through communication content and flow analysis

Analysis type	Communication examples	
	with a trusted entity	**with a distrusted entity**
Communication content analysis	No repeated checking of the agreed tasks, but only ask for the information needed for their own tasks when situations arise.	Repeated checking on tasks: "Did you check monitor A? Did you check monitor B? Remember when you hit that point, do this and that."
Communication flow analysis	The percentage of text pulling and anticipatory pushing of information to or from the entity	The percentage of text pulling and anticipatory pushing of information to or from the entity

Further, the patterns or percentages of communications representing pulling information from an entity versus pushing information in anticipation of another entity's information needs may also show a direct relationship to team trust. Good teams comprise teammates that trust each other, and communicate and coordinate effectively (Reina et al. 2017). Effective communication in teams, especially, military teams is of the utmost importance as it could mean the difference between life and death. In military teams, Soldiers need to be able to effectively and efficiently communicate with each other so that potential threats do not result in loss of life. Further, team members must have the confidence that their teammates will ask for information when they need it, provide information required by other teammates when appropriate, and not distract teammates with unnecessary information at critical times.

Suppose, for example, an UAV drone spots an enemy tank moving in on a squadron from behind but does not push that information to the squadron. The enemy tank gains a tactical advantage and, therefore, is likely to destroy the squadron. In another example, suppose a Soldier is not aware of the next phase of a plan once the team reaches a decision point. In this case, not asking for information could be dangerous for that Soldier and also everyone else involved. For these reasons, an argument can be made that the appropriate push and pull of information can be used as proxy measures for trust. It is hypothesized that the more anticipatory pushing (the pushing of information from teammate A to teammate B based on teammate A's accurate expectation or anticipation that teammate B needs said information) occurs within a team, the more trust there will be in a team. It is hypothesized that the accurate calibration and timing of these actions may help to differentiate between effective and ineffective teams. The communication of each teammate would need to be coded as an anticipatory push, pull, or neither. Afterward, the amount of pushing and pulling would need to be calculated and would need to be correlated with other measures such as performance, known thresholds, or perhaps other measures of trust. For example, if it is found that higher anticipatory pushes and low pulling are correlated with higher scores on the trust questionnaire, it can be said that evidence has been provided to validate anticipatory pushing as a measure of dynamical trust. Patterns of push-pull related to scenario events can also be analyzed and compared between high- and low-performing teams. Note that this method would require a corpus of text-chat communications from multiple teams to be refined.

A visual interaction-based measure may examine the cross-checking ratio: monitoring how much time the operator looks at the status of the agent through eye-tracking data (Jenkins and Jiang 2010) compared against expert suggested monitoring time or data from baseline and control conditions (Table 19). This measurement approach should also be validated with other types of measures of trust, such as verbal explanations of their trust rationale, actual usage of the autonomy, and requests for more information or displays with higher levels of transparency.

Table 19 Example of trust measure through visual interactions

Interaction	Trust	Distrust
Eye-tracking data (a.k.a., cross-check ratio)	Only check the status when needed (this desired frequency should be established in the condition when experienced users interacting with a trusted entity)	Frequently looking at the status to make sure all the values are at the expected range.

A behavioral interaction-based analysis of trust may examine the actions that indicate trust. Reliance is a behavior to fulfill a function by using a tool, an agent, or a system, typically because the agent either trusts another system will fulfill that capacity or they cannot complete the task alone (Johnson et al. 2014). There could be three types of reliance by degree (Table 20). Implementing the level of acceptance approach may overcome the obstacle of using a dichotomy in cases when the autonomy is required to complete a task or ordered by a superior. Data collection could be accomplished by recording the changes the users enact to the default autonomy recommendation (Cummings et al. 2019). Table 20 shows examples of behavioral interaction-based measures. Even when users have the option and authority to choose whether to use the autonomy and choose so, it may not guarantee trust because of lack of other better options. However, these behavioral indicators provide an objective input into the multi-method approach. As we stated in the IPSO model, trust may or may not directly impact outcomes, and the factors that contribute to the positive outcomes are what really matters. Multi-method approaches are needed for studying the complex phenomenon of trust.

Table 20 Example of behavioral interaction-based analyses

Type	Definition	Example	Interpretation
Binary behavior	Whether the user chooses to use the autonomy for single indicator situations, and may appear as accepting the autonomy's recommendations, hand off a task to the autonomy, or disabling the autonomy.	When a task could be done through the aid of RCV's automated driving function or other ways, whether the user chooses to use the function indicates trust.	When the user has the option to use or disuse the autonomy, if they choose to use the autonomy, then it indicates their trust.
Categorical gradual behavior	The extent that a user takes the recommendation, inputs, and plans on a categorical scale of acceptance.	The user chooses to see the agent's recommended path plan, and then alters some aspects of the recommendations with a number of options.	When the user has the option to make modifications, less modification may indicate a higher level of acceptance and trust.
Continuous gradual behavior	The extent that a user takes the recommendation, inputs, and plans on a continuous scale of acceptance.	In an automatic target recognition (ATR) embedded UAV, the camera view automatically focuses on the target 70% of the time. The users can pan the camera view to locate the target if it was not found at the first view.	The high number of clicks panning the camera may indicate less trust.

The complexity of trust in teams may be measured by examining continuous interactions between humans and agents. These interactions may include verbal communications, visual interactions, and behavioral operations. These methods may complement or verify each other depending on the availability of the data. Once we have collected each individual's trust in various entities (i.e., other crew members and the robotic agents) they interact with, we can then show the trust dynamics on a trust network (as shown in Fig. 20). Using this trust network, we can also analyze the impact of other crew members' attitudes on one individual's trust in the robotic agent via an assortment of surveys, verbal probes, and behavioral indicators.

4.3.3.3 Team Workload

Workload has been defined as a "hypothetical construct that represents the cost incurred by a human operator to arrive at a particular level of performance" (Hart and Staveland 1988, p.140). Workload is often described using a resource model in which task performance is dependent on the availability of a fixed quantity of resources (e.g., Wickens and Hollands [2000]). The assumption is that as task demands increase, more resources are required for performance and workload increases. If mental or physical resource availability is exceeded, an individual is required to change their strategy to compensate or task performance may suffer. Workload is a construct that describes the relationship among task demands, operator skill, and task performance (Funke et al. 2012).

Team workload can be indicated by team workload capacity—the potential amount of taskwork and teamwork that teams can engage in assuming the optimal allocation of resources at the moment (Bedwell et al. 2014). The capacity framework includes the interdependencies among the work environment, characteristics of the task (e.g., complexity and difficulty), and team characteristics such as coordination. It has been hypothesized that effective team coordination can mitigate the influence of task complexity on team workload capacity (Bedwell et al. 2014). This hypothesis aligns with other research suggesting that teams alter their coordination strategies in response to high workload (Entin and Serfaty 1999; Stephens et al. 2011; Funke et al. 2012).

Team workload distribution can reflect the bottleneck of teamwork (see the narrow point in Fig. 22) and therefore is important for strategic reasons, such as function allocation, staffing, and coordination. Team workload is defined as a function of the team's interactions as well as their relationship to the environment, available tools, and the task (Funke et al. 2012). Whereas individual workload is directly tied to taskwork demands, members of a team also incur teamwork demands (e.g., communication, coordination, monitoring, and so on). For example, a team may be structured in a rigid, hierarchical fashion, which may be indicated by the commander's node exhibiting high sociometric status. When the team is pushed toward its workload capacity boundary, the central node (e.g., the commander), limited by a finite individual workload capacity, may *bottleneck* the coordination required for a team-level shift in strategy or task reallocation, and limit the compensatory potential of the team. This may result in performance decrements. Conversely, another team may be structured in a decentralized fashion, reducing the potential for coordination bottlenecks, and increasing the team's overall workload capacity. This decentralized organization can also result in extra teamwork demands for communication and coordination, challenges in monitoring and redundant individual efforts. Therefore, it is important to consider workload at a team level.

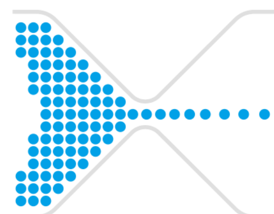

Fig. 22 An example illustration of teamwork bottleneck due to the critical person's high workload. (Note. The road traffic is determined by the narrowest point of the road.)

Well-rehearsed and habitual tasks require fewer mental resources (e.g., knitting, driving) so that people may engage in multitasking. Similarly, working with teammates one has worked with for 15 years requires much less communication than working with a new teammate. Familiarity of a teammate is related to shared mental models about how things should work and established patterns of coordination and task expectations, either in a human–human or human–autonomy relationship. Familiarity of a teammate belongs to the inputs section of the teamwork and influences teamwork interaction processes (Cannon-Bowers et al. 1993; Orasanu and Salas 1993; Endsley and Jones 1997).

The relationship between workload and performance at a team level may resemble the Yerkes–Dodson relationship for individuals (Singh 2009). When it applies to teams, a well-coordinated team may have increased workload capacity and continue to perform well under high workload (see the green line in Fig. 23b), whereas a poorly coordinated team may have decreased team workload capacity and only be able to handle a smaller workload while maintaining acceptable performance (see the red line in Fig. 23c). Teammate familiarity is related to the shared mental models, but they are not necessarily a linear relationship. Core tenants of functional team familiarity include knowing teammates' roles, information specialties, information requirements, and high and low workload periods related to the task and mission.

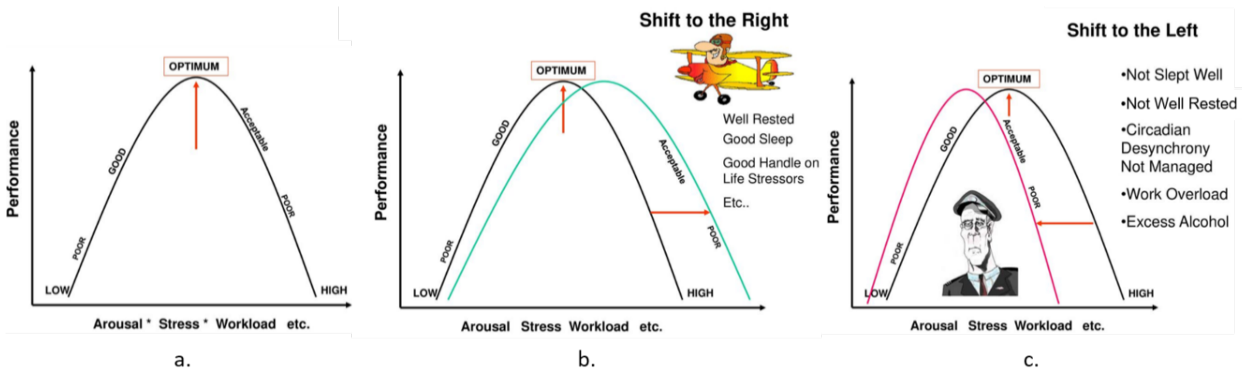

Fig. 23 Yerkes–Dodson relationship between workload and performance (adapted from Singh [2009])

Measurements of team workload have generally taken the form of extensions of individual workload measures, including three categories:

1) **Self-report,** team workload measures using *subjective self-report* include individual workload questionnaires like the NASA-Task Load Index (TLX) (Hart and Staveland 1988) or adaptations intended to differentiate between the dimension of team workload such as the Team Workload Questionnaire (TWLQ; Sellers et al. 2014). These measures are necessarily intrusive if administered during a task, or rely on subjective recall if administered afterward, and may not capture the dynamics of rapidly shifting workload distributions across a team in high-tempo, complex environments.

2) **Task performance** as an indicator of team workload is typically measured as a relationship between primary task performance, such as navigating a vehicle, and secondary task performance, such as monitoring and reporting certain information to the vehicle commander. Using this paradigm, performance decrements on the secondary task may indicate a higher workload (Lenné et al. 2014). When their performance of the secondary task decreases, that means their workload is high. However, having the secondary task adds workload to the performer and in experimental contexts it is often an artificial secondary task that is imposed. Table 21 shows an example of applying the method with non-artificial tasks in the NGCV concept, which has two robotics combat vehicles and one manned combat vehicle. One avenue to consider for future work is identifying appropriate team level secondary measures that tap teamwork demands (e.g., communication, coordination, monitoring, etc.) and could be effectively applied to NGCVs applications.

Table 21 Example of task performance as an indicator of workload

High workload	Low workload
Operator A1 is working on task 1. Meanwhile, they need to complete a secondary task locating positions on a map on demand on the side. If they ignore the map problem, or their performance continually declines, that indicates their workload is beyond his capacity and they are not functioning well.	If operator A1 is doing well on the primary task and the secondary task, then their workload is within their capacity.

3. **Physiological** approaches to measuring team workload have focused on measuring autonomic bodily processes such as heart rate variability and blood pressure. For example, higher variability in the amount of time between heartbeats has been shown to be correlated with higher workload states (Hughes et al. 2019). Many physiological methods have relied on individual measurement and aggregation across the team. However, recent research in this area has explored synchronization and entropy of heart-rate variability between team members over time (Dias et al. 2019). Unfortunately, these measures alone may fail to adequately capture how team workload is tied to team interactions.

However, the disadvantages of these methods above are that they are either static or individual oriented in a dyad relationship. Interaction-based measures of team workload take into consideration interactions among the teammates, and these interactions may change over time. Next are six potential interaction-based measures of team workload.

Social Network Analysis (SNA)

Adaptations of SNA methods have been used to model processes in systems, such as distributed SA, communication flow, and structural relationships between subunits (Stanton 2014; Salmon et al. 2017; Stanton et al. 2017, 2018). We propose for the first time to use SNA to model team workload. Workload through SNA is determined by these dimensions: 1) the number of agents to interact with, 2) the number of interactions through various media, 3) the extent to which the interaction demands mental resources (the difficulty level of a task; see the fifth measure), 4) the length of interaction, and 5) the pattern change of interaction over time. Metrics such as *centrality* or *sociometric* status (Stanton 2014) may correlate with team workload "bottlenecks" and reveal possible structural effects.

For this SNA, the data are typically put into the form of a matrix, which represents interactions over some dimension (verbal communications in this example), during a specified amount of time (i.e., over a full mission or during a single task). By convention, message *sender* labels are in the first column and *receiver* labels are in the first row (Table 22). This example is a bidirectional network, which means the direction of communication is recorded (i.e., A to B ≠ B to A).

Table 22 Example of SNA data format A

Sender/ receiver	RCV_A1	RCV_A2	RCV_A	S1
RCV_A1	…	10	8	12
RCV_A2	7	…	14	13
RCV_A	14	15	…	30
S1	13	21	16	…

The agents within the network are commonly referred to as *nodes*. Nodes are connected by *links*, which may be weighted according to some parameter, such as the number of occurrences of an interaction over a dimension. In this example of communications within RCV operators and the section leader (Fig. 24), the interactions are weighted according to the number of verbal communications: bigger numbers of interaction frequency indicate potentially high workload and small numbers potentially low workload. *Degree of centrality*, a measure of the number of connections that a node has, is an example of a node-level measure that has been studied extensively. *Density* is an example of a network-level SNA metric, characterizing the number of connections in a network compared to how many are possible (Borgatti et al. 2018). This SNA provides a view of the team workload distribution in terms of communication channels.

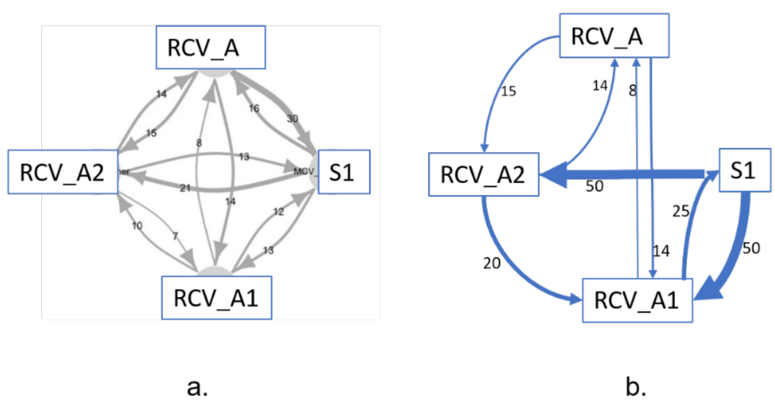

Fig. 24 SNA of workload output for the RCV related entities. SNA diagram for the taxonomy of interactions for controlling one RCV (RCV_A = all interfaces of RCV A; RCV_A1 and RCV_A2 control RCV A; S1 = section leader/platoon leader). a) assumes S1 interacts with both the vehicle RCV_A and operators RCV_A1 and RCV_A2 and b) S1 primarily interacts with RCV_A1. a) used R and b) used PPT to illustrate the concept.

Table 23 shows a different SNA data format, taking multiple dimensions into consideration.

Table 23 Example of SNA data format B

Subtask	From	To	Detection	Identification	Assessment	Engagement
Detection	S1	RCV_A1	1	0	0	0
Identification	S1	RCV_A1	0	1	0	0
Assessment	S1	RCV_A1	0	0	1	0
Engagement	S1	RCV_A1	0	0	0	1

Additionally, SNA may be able to retroactively capture the restructuring of teams under conditions of high workload to better understand their collective responses. This information could be used to inform system design and training. In related research, Barth et al. (2015) used SNA methods to characterize surgical team adaptation process during periods of high and low *complexity* (Fig. 25). They found that team communication increased overall during high-complexity periods but was marked by fluctuations between long periods of silence and peaks of activity. They also found that team communication patterns “flattened” during high-complexity periods (indicated by measures of network density), suggesting an adaptation to support higher information sharing without as strict regard for protocol, hierarchy, or organizational structure. Similar measures could examine how teams adapt their communications under periods of high workload.

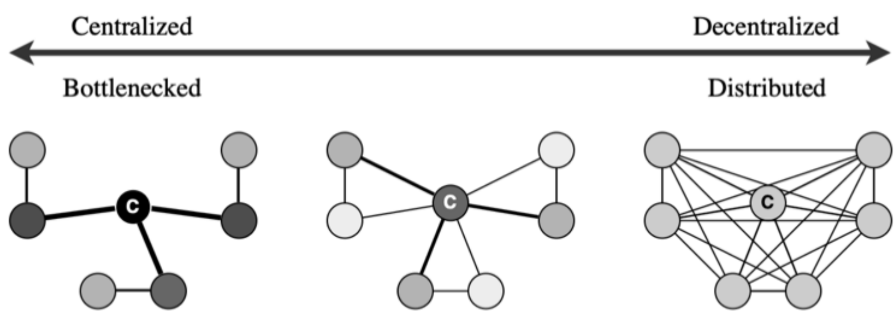

Fig. 25 Example centralized and decentralized communication networks under high task demand. Thicker links represent higher communication flow. Darker nodes represent higher individual workload. (Left) Bottlenecking takes place due to the limited ability of the center node to convey necessary information and direction to periphery nodes. (Right) The central node avoids overload due to the existence of additional links between periphery nodes, and the team exhibits better workload distribution overall. A hybrid structure is displayed in the middle. Hypothesized relative SNA measures displayed by arrows. Adapted from Barth et al. (2015).

Task Load

We may consider the task load across the team to predict team workload levels. The higher task load an operator has, the higher workload they may experience. For example, if the team needs to accomplish five tasks and each team member needs to finish a number of subtasks (Table 24). The task load distribution may help predict team members' workload distribution for each team member during these task categories (Fig. 26). The limitation of this method is that some tasks are more difficult than others and causes a higher workload. Therefore, the mental effort requirements for the tasks need to be defined.

Table 24 Example data for the number of subtasks to be finished

Team role	T1	T2	T3	T4	T5
S1	6	4	5	7	5
C1	2	3	1	5	3
C2	3	2	4	1	5
A1	3	5	4	2	1
A2	2	4	3	5	2
B1	4	3	5	2	1
B2	2	5	7	2	6

Note. S1 = section leader, C1 & C2 = vehicle C operators, A1 & A2 = vehicle A operators, B1 & B2 = vehicle B operators. The hypothetical data are for demonstration of the concept.

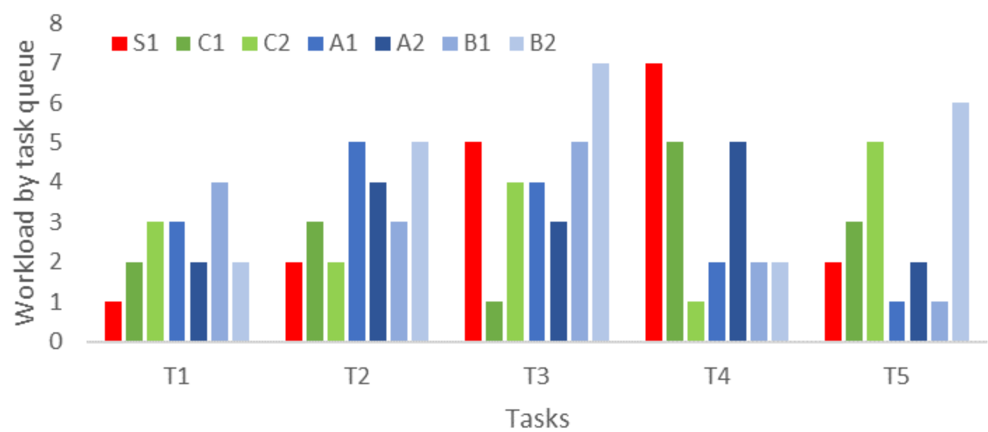

Fig. 26 Team workload by task load

Discrete Event Simulation

Discrete event simulation provides an even more detailed analysis of bottleneck identification to model team workload. In this method, each task with a start time and end time is treated as an event; workload is defined as the percentage of working time out of the total length of the shift (Huang et al. 2018). Each of the seven crew members distributes their time differently. A task analysis is conducted to identify the main tasks that each crew member works on (Fig. 27).

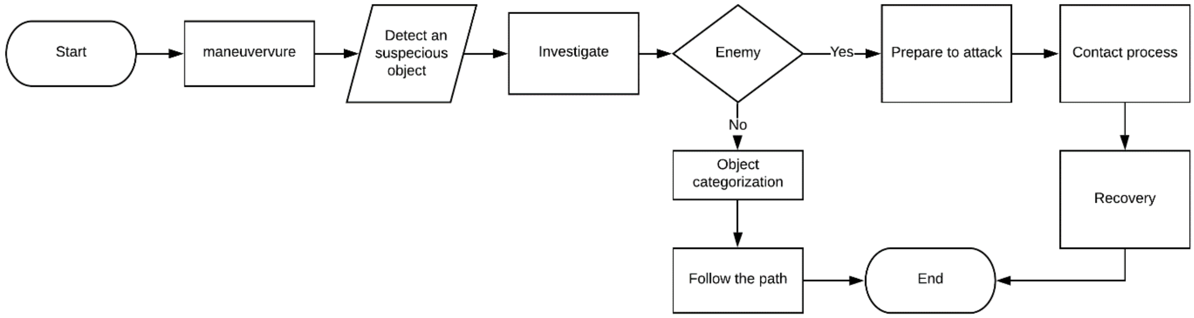

Fig. 27 Example of task analysis using flowcharts

Discrete event simulation can be used to develop a workload model to predict whether a crew member is undergoing a high or low workload when conducting certain tasks. It takes three steps to build the workload simulation (Huang et al. 2018). Step 1 is to identify the types of primary tasks involved in the procedure of interest. Step 2 is to collect data of the task duration by taking the start and end time stamps of the tasks. Based on the example of railroad dispatchers' log data, a hypothetical example of task duration in the NGCV context is shown in Table 25. Step 3 is to convert the single task duration to the average duration per task category per hour (Table 26). Step 3 is to select the best distribution fit in Matlab based on low square error from common models and build the model (Stimpson et al. 2016; Huang et al. 2018).

Table 25 Hypothetical sample manual log for one person in the NGCV context

Start	End	Notes	Task category
10:48:52	10:50:25	Moving forward and keeping a distance from the front vehicle	Maneuver
10:50:25	10:54:30	Searching for the next enemy vehicle	Target searching
10:54:30	10:56:15	Confirming the object category and marking on the map	Target identification
10:56:15	11:01:23	Encountering a block on the road and making a detour	Navigation

Table 26 Summary of frequencies of observed task categories per hour

Sub	Maneuver	Target search	Target identification	Navigation	Ave. time per task
S1	0.10	3.33	10.11	0.00	0:01:34
C1	46.00	2.00	16.23	60.20	0:03:11
C2	0.15	0.00	22.41	0.00	0:01:09
A1	20.25	0.00	33.91	45.10	0:01:26
A2	0.00	25.23	15.00	10.34	0:05:12
B1	30.00	0.00	0.00	40.15	0:01:12
B2	0.00	28.05	20.25	15.02	0:10:15

Note. The first column is the subject. Column 2–5 are the accumulated time for the each task category in one hour. Column 6 is the average time duration for each task.

Each person's task category and duration data are used to make the workload distribution plot (Fig. 28 shows an example from the railroad dispatchers' workload plot [Huang et al. 2018]). If we plot the workload for each crew member, we can build a workload network (Fig. 29). Considering the distribution of all seven crew members on a team may result in suggestions on workload adjustment and identify patterns that characterize either high- or low-performing teams. This method can also be used to build a model for special events, such as quiet time, when people face road obstacles, and when people are being attacked.

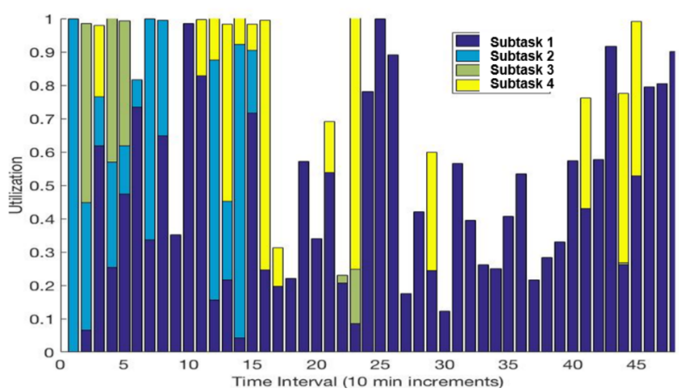

Fig. 28 Representative utilization of one person during 8-h shift using discrete event simulation (adapted from Huang et al. [2018])

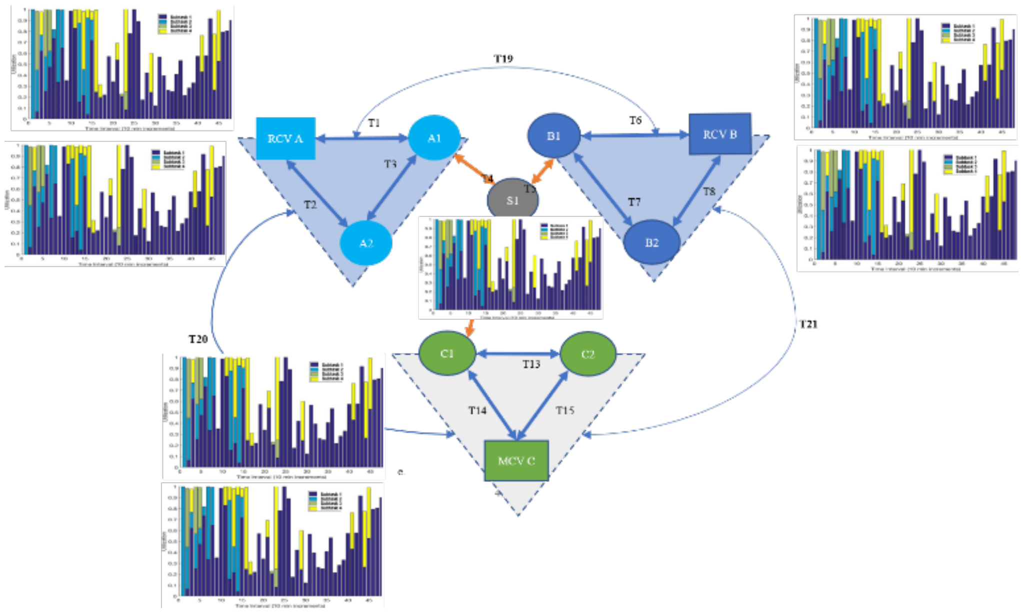

Fig. 29 Team workload distribution by time utilization (Note: Utilization refers to the percent of busy time spent on the tasks over the allotted time interval. In this network, each crew member has a workload distribution plot in a typical hour.)

To make the individual workload simulation to model the overlap or causation of workload from one team member to another, it requires expanding and customizing the model to account for the time spent on the interactions with the autonomy, or even which types of interactions with the autonomy. It will also need to account for the tasks associated with each crew member and see if the workload simulation reflects that person's task load. By understanding the percentage of time/effort spent on each type of interaction with the autonomy, we may identify which types of interactions are taking too much time away from the crew member and whether the autonomy needs to be optimized in that aspect. So this analysis may help us find out the impact of operating, reasoning with, understanding, and working with the autonomy.

To compare team workload across teams, we can also try aggregating the whole team's workload distribution and compare it with other teams. This team-level discrete event simulation for workload network needs to be tested with empirical data and scenarios.

Workload Simulation Tools

The Improved Performance Research Integration Tool (IMPRINT) is a related workload simulation tool that has multiple scenarios that can be applied to HAT (Mitchell 2009; Plott et al. 2017). Users can assign values to each task and other influencing factors to model the changes in individuals' workload based on the changes of the influencing factors. The parameters in this workload simulation tool may be refined based on actual data from similar applications or expert ratings and thus be used to predict workload in conditions that do not have available empirical data and variations in task organizations, assignments, and conditions can be tested out. IMPRINT is ARL-developed tool for workload simulation. It is more structured than the discrete event simulation method. This type of workload simulation tool has already set up the framework, and since it is built for ARL, many of the components may be borrowed to determine the parameters in the NGCV context. This tool requires both time, data to produce estimates and some IMPRINT expertise to build out a model.

Patterns of Team Communication

This method assesses team-level communication patterns (i.e., anticipation ratios—the number of transfers to individual X by the number of requests made by X, pushing information in anticipation of a teammate's needs), or changes in the amount of overall communication between specific team members, which can be representative of high and low team workload (Entin and Serfaty 1999). Language communication is one type of interaction-based data, and it can fill the gaps where other data are not typically available and make it possible to indicate crew members' workload by using interactions such as communication patterns. This can allow for interventions such as training, facilitation, or real-time alerts to deviations from optimal performance patterns.

Measuring Information Channels based on the Multiple Resource Theory

Multiple resource theory (Wickens 2008) can be applied to interdependent team-level activities to measure team workload. Understanding how the various resources are being utilized across the HAT teammates can identify conditions for overload and also underutilization. These measurements can lead to metrics defining these states and guide responses from teammates, especially autonomous ones that are prone to miss situational or implicit cues, to more effectively time or deliver input. It is important to consider how the NGCV environment uniquely impacts various resources as well (impacts of ambient noise and vibration on communications). For instance, if a vehicle commander is communicating over the radio to the entire crew, they are primarily using their verbal responding attentional resources, and the crew attending to the radio are using auditory perceptive and cognitive resources. These attentional resources could be quantified in real time to indicate when these resources are being used, and if occurring simultaneously with other events that demand the same resource, to indicate times where attention may be overloaded. Another example is exchange of information via graphic control measures on a shared map display common in combat vehicles. Depending on the interface, placing the graphic may involve different attentional modalities while teammates who eventually view the graphic use visual attentional resources. Generally, interactions involving attentional resources for one teammate that incur attentional demands on another may constitute a metric of team workload dynamically. However, it is noteworthy that the graphic control measure example involves an interdependent workload that may be asynchronous, while radio communications occur simultaneously. Table 27 illustrates the input data for analyzing radio communication over time. Figure 30 illustrates an example of team communication viewed in terms of the number of teammates engaged in verbally responding and auditory perception and cognition. This area needs further exploration.

Table 27 Example data structure for the attentional resource approach to measuring team workload

Mission time (s)	Senders	Recipients
1–5	1	6
6–10	0	0
11–15	1	2
16–20	1	6
21–25	0	0
26–30	1	1
31–35	2	6
36–40	1	1
41–45	1	13
46–50	1	1

Note. The purpose of the hypothetical data is to demonstrate the concept. Senders and recipients of natural-language communication via radio are recorded over time.

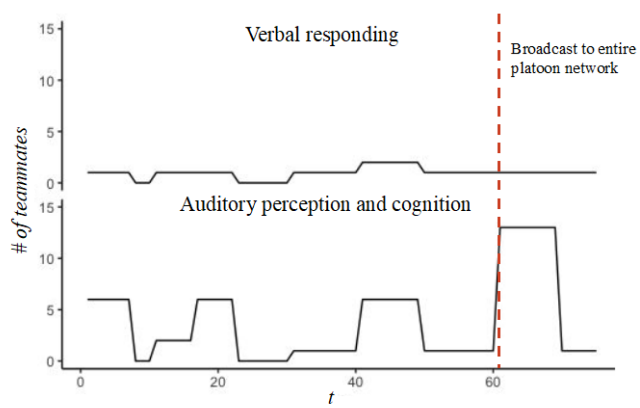

Fig. 30 Team communication over time decomposed into multiple resources of attention. Each line represents different attentional resources engaged as a result of team interactions. When a teammate communicates over the radio, they are using verbal responding resources, while recipients are using auditory-verbal perceptive and cognitive resources.

The section has defined a field of possible workload measures and options that could be brought to bear on the NGCV HAT situation. This will be used as a starting point to prioritize the best measures and then test and refine those in testbed experiments during the next phase of this project.

4.3.3.4 Team Resilience

Team resilience is a phenomenon characterized by positive adaptation at the team-level and maintenance of performance despite experiencing challenging or off-nominal events. A resilient team can effectively reorganize themselves when their experienced competence envelope is exceeded by demands. At a broader system-level, resilience is often contrasted with optimization, which are adaptations to refine capabilities to best fit the expected situations. By contrast, adaptations that expand capabilities increase extensibility, or resilience (Woods 2018). The opposite of resilience is brittleness, which is a breakdown of a system at its boundary of competence (Woods 2015). A common metaphor for resilience is the stress–strain curve of solid mechanics as displayed in Fig. 31 (Sheridan 2008), where a material may be stretched and return to its original state (analogous to mental workload) up to a limit in which it will "give" and no longer return to its original state. For a team, this would entail experiencing an extreme event that pushes them beyond their experienced limit for adaptation, leading to performance breakdowns.

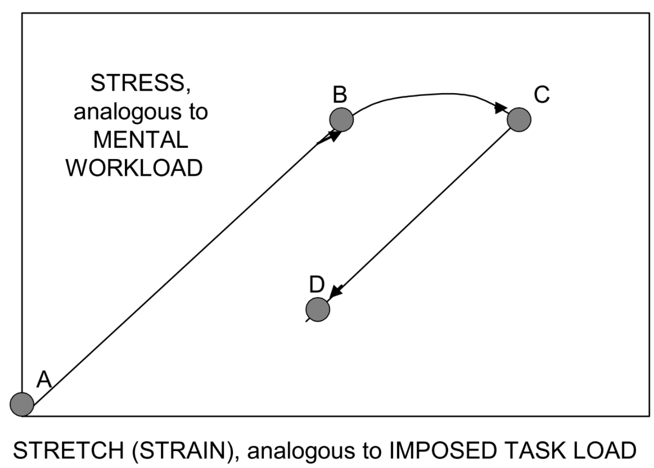

Fig. 31 Stress–strain curve metaphor for resilience (Sheridan 2008)

Importantly, a team of resilient individuals does not necessarily make a resilient team. A resilient individual is one who engages in coping strategies such as relaxation, mental simulation, or anticipation to deal with extreme of unexpected events (Bowers et al. 2017). This would be effective if each individual experienced challenges in their independent tasks, but these processes may not contribute to graceful adaptation at the team-level given a challenging or off-nominal task that is interdependent. Processes for anticipatory thinking, for instance, are characterized differently at the individual and team level (Klein et al. 2004). A resilient individual may even inhibit processes that enable team resilience, for instance, by taking responsibility for the unexpected demands of the entire team, thereby reducing the need for other teammates to cultivate adaptive capacity. Teams, on the other hand, may need to gain awareness and form understandings of novel events, make flexible plans and revise them, and reground the team as events bring into question their basic compact and interdependence (Hoffman and Hancock 2017).

There is evidence that team resilience can be instantiated proactively in teams. Gorman et al. (2010) examined the effects of perturbation training on team coordination and found that introducing perturbations to the team's normal coordination in training scenarios improved the overall flexibility of their coordination. Similarly, Gorman and Cooke (2011) found that mixing up the team composition after a retention interval allowed teams to form stronger team processes while also improving flexibility. In general, it appears that diversifying the constraints in teamwork (e.g., team composition and situations) may be a critical component to preparing teams for unexpected conditions to the extent that it affects how team coordination evolves. A recent effort has begun to evaluate how teams explore novel interactions and gain a broader understanding of the constraints that lead to novel variations in coordination that are adaptive (Lematta 2019).

A critical condition for measuring resilience is that the work system experiences events that exceed its experienced competence envelope (Hoffman and Hancock 2017). Adapting to situations a team expects (e.g., contingency planning) is more appropriately characterized as robustness (Hewitt and Woods 2015). Examining team-level processes, such as team cognition, that enable and maintain resilience is critical to establishing context-independent and leading indicators. With that said, measures of resilience are typically categorized in terms of leading and lagging indicators. A leading indicator is predictive for how a system may be resilient, while a lagging indicator concerns how resilient the system was (Hollnagel 2016).

One example of a lagging indicator examines the time for a system to recognize a need to change (r), plan this reorganization, and implement the change (i) compared to total event time (t) (Hoffman and Hancock 2017). This is represented quantitatively with the following equation:

$$R = [(i - r) / t\,] \times 100$$

For this equation, smaller numbers correspond with greater resilience. Other work has explored nonlinear dynamics of multiple system layers to examine team reorganization (Gorman et al. 2019). This method utilizes measures of information entropy (predictability of the communication pattern based on prior communication at a specified point in time) to quantify changes in system composition in real time. Figure 32 shows an example analysis with a sample US Army Wingman Joint Capabilities Technology Demonstration data set of a manned-unmanned team gunnery exercise as Fort Benning, Georgia (Brewer et al. 2019), where entropy was calculated as turn-taking behaviors consistent or inconsistent with prior behaviors, where high entropy represented less consistency and predictability. From these data, the predictability of a team's communication may be correlated with events and team activities such as perturbations and team-level adaptations. Because adaptations may include changes to the way teams interact, measures of communication entropy may be effective in capturing a team's response to perturbations.

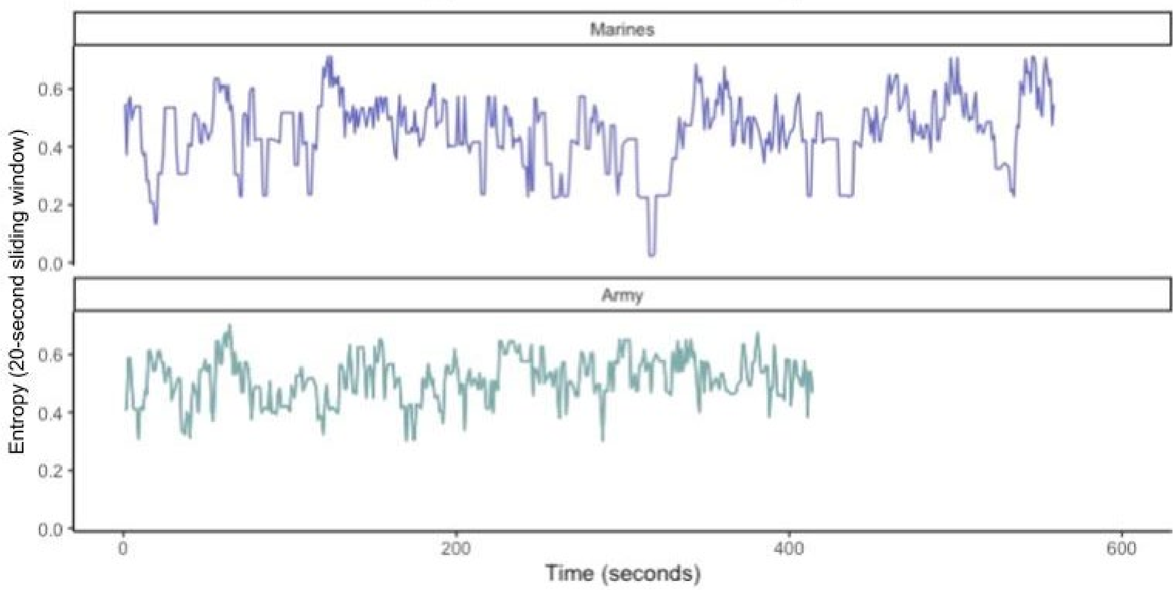

Fig. 32 Example of communication flow entropy in gunnery exercises

A leading indicator of team resilience may consider team-level activities that occur before an off-nominal event requires adaptation. Woods (2018) argues that agents *expressing initiative* and *engaging in reciprocity* are critical for networked agent systems to sustain resilience over time. In the NGCV context, a gunner may express initiative to engage a target with discretion because a commander has provided authority to engage it in advance. This exchange may be beneficial when the commander has trust in the gunner and anticipates their workload may increase during the target engagement process.

Additionally, recent work has suggested that *exploratory team interactions* may be associated with effectiveness in normal and degraded conditions in a three-agent aerial reconnaissance task (Lematta et al. 2019). Measures of exploratory communication represent an evolution in a team's process by quantifying or qualifying novel variation in team interactions over time as a metric of team resilience. This metric assumes that a team that explores new ways of interacting is more capable of identifying effective solutions to subsequent perturbations. One approach may decompose team interactions in natural language over radio into dimensions (e.g., content, communication flow) and a set of relevant states. Then, coding the time-series of team communication in terms of these states and applying nonlinear dynamical analyses may quantify the predictability of a team's overall coordination pattern, represented as a percentage of determinism (%DET – a summary plot of recurrences in identified signal states over time). For a detailed example of this method, see Gorman et al. (2019). The key difference between the example in Gorman et al. (2019) and the exploratory communication metric proposed here is considering exploration as predictive of effective adaptation and performance recovery rather than identifying perturbations and roadblocks.

Table 28 provides an example of how determinism and resilience may correspond. An example analysis may be a correlation between communication determinism (Gorman et al. 2019) before the perturbation onset and the resilience formula (Hoffman and Hancock 2017). Figure 33 provides a sample exploratory communication analysis with sample Wingman Data (Brewer et al. 2019). No conclusions are warranted from this analysis as key data parameters to make valid comparisons were missing and comparisons across non-matched trials is not warranted. The example was just to illustrate the concept with the type of data expected to be used.

Table 28 Example using mock data showing how proactive exploration might be associated with reorganization during a perturbation

Team	%DET of communication before perturbation	Resilience *(R = [(i – r) / t] x 100)*
Resilient Team	65%	80
Brittle Team	85%	50

Note: The resilience of the team is determined by the %DET of communication. The lower %DET the more resilient of the team, while the greater the R value, the more resilient the team is.

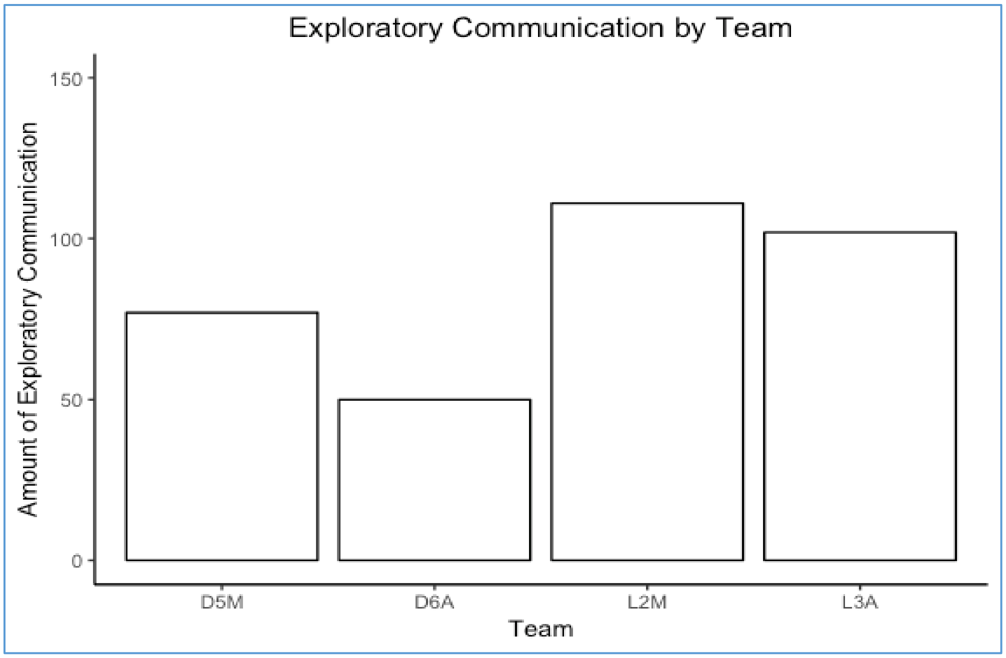

Fig. 33 Example of overall exploratory communication concept using teams in a gunnery exercises, where D = dry run, L = live fire, A = Army, and M = Marines

Along with other team-level states of interest in this paper, team resilience may benefit from SNA approaches. Some current approaches, such as Event Analysis of Systemic Teamwork (EAST; Walker et al. 2006), also show promise by focusing on "hidden links". EAST provides a high-level view of a social system by focusing on tasks, information, and the social system. The output of the analysis is in the form of graphics that represent these various aspects of the system that can be examined and compared qualitatively and quantitatively (Fig. 34 shows examples of graphical outputs of EAST from a cybersecurity study; Rajivan and Cooke 2017). To measure resilience this method can be applied similar to the EAST "Broken Links" approach (Stanton and Harvey 2017) by injecting nodes at the information, task, and agent level. By then breaking these nodes the fragile and resilient aspects of a network may be identified. This process requires recording who interacts with who during a team task (e.g., 30-min simulated mission), the frequency of their interactions, and codifying the tasks that correspond with team communication. The data are then represented as an EAST network. Nodes and edges are systematically introduced into the network at the information and task level, which represent unexpected dependencies. Then, the procedures in the "broken links" approach are followed for each iteration of the network that introduces or breaks a dependency.

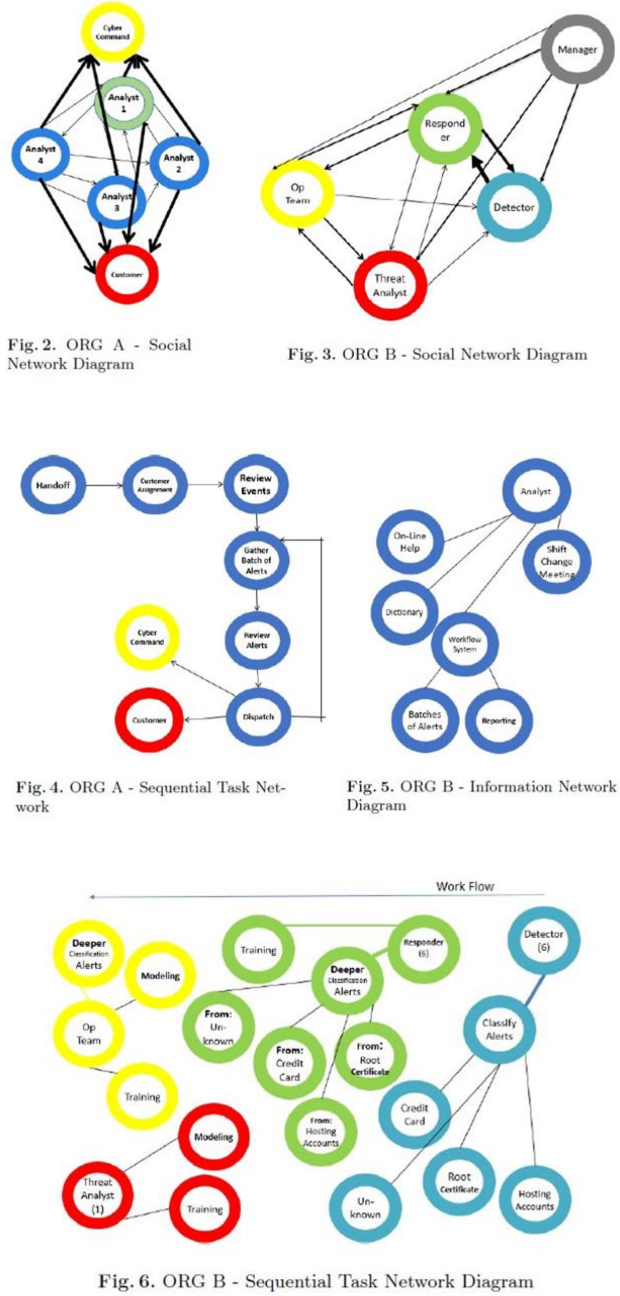

Fig. 34 Examples of graphical EAST outputs from Rajivan and Cooke (2017)

For example, in a target engagement task, the gunner utilizes aided target recognition (AiTR) functionality combined with visual determination and coordination with the mobility operator and commander to respond to targets. The gunner relies on a visual display and sensors to detect and identify targets but in the scenario loses that visual display due to damage, thereby losing their ability to see the outside environment. The equipment is new to this unit and they have not yet trained for this contingency. This situation will force the team alter their operations, for example, to rely on the mobility operator or commander to do the visual confirmation of AiTR detected targets. Effectively, this process asks, "Given that a particular task or information dependency exists outside of the team's experienced competence envelope, will the team be able to continue normal processes if that dependency disrupts a normal team communication and workflow process?"

A limitation of current EAST applications is that they do not account for events outside the competence envelope described in the team network. Future work will consider network modeling that can incorporate unexpected events to anticipate how a system may adapt, and thus show their patterns of broken links at different levels of team resilience.

4.3.4 Multi-Criteria Team Outcomes

When it comes to team outcomes, team effectiveness should not depend on one criterion. For example, the emergency room medical team effectiveness should not be determined by the patient's survival rate alone, but also consider given the condition of the patient's initial status, how many subtasks the team conduct successfully and efficiently. Likewise, in the NGCV context, team outcomes should not only include the ultimate mission success alone, but also the platoon crew members' successful completion of the subtasks and efficiency. In some cases, the mission may have several objectives or multiple aspects. Even if the ultimate mission fails, completion of sub-missions, subgoals, and subtasks may also have great value for the team's learning and improvement, increasing team effectiveness for future missions. For example, the safety of the Soldiers, and reservation of the weapons, enemies attacked, enemy properties damaged, resources depleted, and critical information acquired. Therefore, the desired outcomes of the NGCVs should be identified for each type of combat mission. Then any team with any mission can quickly select outcome criteria for appropriate evaluation.

In this work, we propose to identify the criteria that are critical for the combat context and evaluate the success of each materialized outcome, rather than one global assessment of whether the team achieved mission success on one combat mission. These measures may include the following: 1) successful completion of essential subtasks identified on the battlefield (Johnson et al. in press), 2) amount of resources available (e.g., fuel and ammo expended and associated with the ability to continue the mission, 3) causalities, including effective evacuation of injured Soldiers or civilians, and damage to property, and 4) efficient completion of tasks to make necessary time for other tasks, as well as meeting timelines and checkpoints. These measures will be reviewed by SMEs for validation.

4.3.5 Team Effectiveness Modeling

As explained at the beginning of this report, team effectiveness is a multi-dimensional concept that is not only reflected in the teamwork processes but also in multi-criteria team outcomes. To be more specific, this project is focused on effective team states—team SA, team trust, team workload, and team resilience. We also propose to use interaction-based measures to model these team states.

Hypothesis 1: Effective team interaction processes will positively correlate with desirable team states.

- 1a: Effective team interaction processes will have a positive correlation with trust.
- 1b: Effective team interaction processes will have a positive correlation with SA.
- 1c: Effective team interaction processes will a have an optimal range of workload.
- 1d: Effective team interaction processes will have a positive correlation with resilience.

Hypothesis 2: Desirable team states will positively correlate with positive multi-criteria outcomes.

Hypothesis 3: Effective team interaction processes will positively correlate with positive multi-criteria outcomes.

Each team state has a range of values and boundaries that differentiate between desired and undesired conditions of those states. When one variable violates the boundary, it is counted as one violation.

We propose three model approaches to determine the best model for team effectiveness.

Approach 1: Models of separate indicators of team effectiveness. This approach tracks each team state separately and establishes the desired range for each observed team state.

Approach 2: A regression model of team effectiveness. The second approach standardizes the values of the team states in approach 1 and combines their standardized values and weights through a regression model. Here, the aggregate model can be related to team effectiveness:

$$Team\ state\ effectiveness\ = a_0 + \beta_1(team\ trust) + \beta_2(team\ SA) + \beta_3(team\ workload) + \beta_4(team\ resilience)\ +\ \varepsilon \quad (1)$$

Approach 3: Discrete model. The third approach is to combine the indicators and, rather than focus on a regressed model, instead generates a discrete team effectiveness score that is readily interpretable. When an indicator violates its desired boundaries, it is counted as 1. If all variables violate their desired range of responses, the team effectiveness alert value will be 4 in this case. This provides a predictive metric between 0 (normal state) and 4 (extremely abnormal state) when low team effectiveness (ineffectiveness) is anticipated. A preliminary math model is described here:

$$Ineffectiveness\ alert = \sum_{i=1}^{n} w_i \, 1s_i\{x_i\} \tag{2}$$

$$1s\{x\} = \begin{cases} 1, & if\ x \in s \\ 0, & if\ x \notin s \end{cases} \tag{3}$$

$$S = [0, a) \cup (b, \propto] \tag{4}$$

Function 2 is a sum of ineffectiveness alerts, in which w_i is a relative weight for each team state; currently, it is assigned as equal weight, but weights may vary. Function 3 defines the team states' output value. Function 4 defines the set of desired values.

Applying the approaches to expert operators will help generate the desired range for each team effectiveness indicator. The model and online functions can be used to monitor operators' real-time state during simulated training and live actions. Each different formulation provides a different perspective on all of these novel measures. We will improve the ability of our measurements and model to predict team effectiveness and align with team interaction processes as the program progresses. A final model will be delivered to include test results and a determination of the ability to predict team effectiveness and the relationship to team interaction behaviors.

The choice of approach depends on the variable of interest and the availability of the interaction data. Take team states for example, we proposed several approaches for each of the team states, such as team trust. At our current stage, these approaches have not been validated yet. We will need to collect empirical data and try out the measures to better understand the validity of each and their costs. This process is necessary to develop metrics and modeling methods.

4.3.6 Metrics Requirements

Effective HAT metrics should use multiple methods for assessing team cognition, with a focus on interaction-based measures, such as teammates' verbal and behavioral interactions during their tasks. In some cases, different approaches tell us different types of information about the interactions, for instance, objective measures of interaction frequency, social network measures and self-reports of team experience may reveal different aspects of teamwork effectiveness. Interaction-based measures are relatively objective. Subjective assessment is often biased or inherently flawed but can also determine how teams and human team member behave despite objective performance and their perceptions that can influence future teaming. Combining data from multiple analysis methods should provide a stronger metric and better understanding of the interaction (e.g., trust calibration, workload, stress, and so on) and HAT dynamics.

We are currently at the stage of identifying measurements. It will require extensive testing, validation, and refinements to explore the performance standards for each measure, the adjustments by various contexts, and the interactions across measures required for metric development. The metrics developed should have a good reliability, construct validity, and also convergence validity with other data concerning performance and other overlapping measures. In addition, ideal interaction-based metrics should include the following characteristics (Table 29). The rankings provided were initial rankings among the project team to support prioritization discussions. The goal is to also confer with the larger team of ARL HAT collaborators working these themes and in reference to testbed progress to make prioritization judgements.

Table 29 Measurement options and comparisons

Team state	Measure	Definition	Data required	Data collection equipment and procedure	Data Availability (0–3) unlikely to already available	Team level constructs	Added value (0–3) none to high
Team SA	SAGAT (Endsley, 2017)	Each team member's knowledge of the situation at random selected moments	Voice recording; Pop-up survey responses;	freezing a simulation or scenario of interest at selected times, blanking any displays or information sources and asking participants SA probes about the situation (both system and external) at the three levels of SA.	1	Yes	1
	CAST (Gorman et al. 2006)	Whether the relevant crew members communicate SA to the correct crew member to fix the obstacles.	Text chat, system logs, voice recording	1. Identify roadblocks that requires adaptive and timely team-level solutions. 2. Document perceptions 3. Document coordinated perceptions 5. Document coordinated actions	1	Yes	2
	iSAFT	What operators did not know that they should have known at the failed tasks	Task list, task performance criteria, failed tasks, required SA for the team tasks, interaction channel that failed the SA development	Retrospective interview; Screen recording replay; System log analysis;	1	Yes	2

Table 29 Measurement options and comparisons (continued)

Team state	Measure	Definition	Data required	Data collection equipment and procedure	Data Availability (0–3) unlikely to already available	Team level constructs	Added value (0–3) none to high
	SPAM (Durso et al. 1998)	Operator's ability to locate information in the environment	Understanding of tasks and SA requirements to develop probes	1. Probe participants during performance about critical SA items. 2. An adaptation is to embed these probes as natural scenario events	1	Maybe depending on probes	1
	SART (Taylor 1990)	Subjective measure of operator's perceived SA on 10 dimensions	subjective response	Provide instrument post-trial	3	No	1
Team workload	Social network structure	The density centrality of communications among the team members over time.	Time, sender, receiver, and messages (Who interacts with who during a team task (e.g., 30-minute simulated mission), the frequency of their interactions, and how these change over time)	Voice chat system and audio recording, and then chat transcription and text data frequency coding	3	Yes	2
	Team coordination demand	The attentional resources used by a team.	Time, sender, recipients, modality	Voice chat system and audio recording or transcripts, interface inputs and screen capturing	2	Yes	2

Table 29 Measurement options and comparisons (continued)

Team state	Measure	Definition	Data required	Data collection equipment and procedure	Data Availability (0–3) unlikely to already available	Team level constructs	Added value (0–3) none to high
	Discrete event simulation (Huang et al. 2018)	The percentage of time used to do the task over a prescribed time period.	Type, duration, and frequency of tasks	Receive training on the tasks. Minimally use naturalistic observation with paper, pencil, and a clock. Video recording or voice recording, and system logs are preferred.	1	Yes	2
Team trust	Binary behavior	Behaviors allowing another agent to fulfill or assist with a specific capacity.	Time, Manual control activation, RCV control parameters, RCV automated functions	Identify action options, collect interactions	2	Maybe	1
	Continuous gradual behavior	The extent of behavior that a user takes the recommendation, inputs, and plans on a continuous scale of acceptance.	•Whether operators choose to use this AiTR function •The number of clicks operators use to relocate the target on the screen •The number of targets successfully identified •The number of misidentified targets	Export computer system logs to .CSV	2	Maybe	1
	Eye-tracking analysis	A visual interaction-based measure	Frequency and duration of fixing ones' eyes on the agent while doing secondary tasks	Define areas of interest, use Eye-tracking devices and data synchronization software (e.g., iMotion)	1	Maybe	2

Table 29 Measurement options and comparisons (continued)

Team state	Measure	Definition	Data required	Data collection equipment and procedure	Data Availability (0–3) unlikely to already available	Team level constructs	Added value (0–3) none to high
	Communication analysis	A verbal interaction-based measure: Push and pull percentages in team communication	Transcribed communications (transcripts or text messages) between team members and the robotic agents	A chat software that records human agent communication messages; then the data will be manually coded or use programs like Python to sort the data.	3	Maybe	2
Team resilience	Team interaction exploration	Novel variations in team communication	Time, sender, recipients, content	Voice chat system and audio recording or transcripts	3	Yes	2
	Response to off-nominal events	The time it takes a team to implement a coordinated solution after perceiving an unexpected event.	Time, sender, recipients, content, off-nominal event start and end time.	Voice chat system and audio recording	2	Yes	2

Table 29 Measurement options and comparisons (continued)

Team state	Measure	Definition	Data required	Data collection equipment and procedure	Data Availability (0–3) unlikely to already available	Team level constructs	Added value (0–3) none to high
	EAST "Hidden Links" (Stanton and Harvey 2017)	The structural properties of an EAST network allow for the completion of tasks and information sharing despite the introduction of novel tasks or information being introduced into the network. "Hidden links" is similar to EAST "Broken Links" approach.	Who interacts with whom during a team task, the frequency of their interactions, and codifying the tasks that correspond with team communication	Voice chat system and audio recording	2	Yes	2

The following are three requirements for the first round of measure assessment and prioritization:

- Data availability: Can we access the data required?
- Team level constructs: Does the measure capture team level constructs?
- Added value: Does the measure add value beyond what the current HAT team is considering*?

Next are future goals for team states measures:

- Unobtrusiveness: The metrics should not interrupt the task flow in the human–agent teaming process.
- Automated analysis: There should be a way to automatically code the data to speed up data analysis.
- Real-time capture: The metrics should be able to capture real-time data to provide feedback for online adjustment.
- Context-free (generalizability): The metrics should be applicable in different domains and contexts.

4.4 Conclusion and Future Directions

This work proposes the IPSO model to evaluate team effectiveness and suggests using interaction-based measures to evaluate the essential functional level of subtasks to capture both team interaction processes and multi-criteria team outcomes. Team inputs or team composition could also be used to help predict team effectiveness. Meanwhile, team interaction processes, team states, and multi-criteria outcomes are the key components for measuring HAT effectiveness. However, team inputs, team interaction processes, team states, and team outcomes may be all connected. For example, the increasing levels of automation in combat vehicles will change interactions in battlefield scenarios, changing the subtasks of HAT. Therefore, the evaluation of team interaction processes will be changed accordingly. The interaction processes indicate team states. And then the team outcomes on completion of the subtasks will also be determined based on the interactions. Therefore, the research will have to remain cognizant of the team inputs, their determination on subtasks and capabilities, their impact on the method

* Many of these measures proposed are exploratory and need further testing and refinement and therefore the value-added estimates made below are conservative on potential value. These are preliminary as further discussions with the larger human–autonomy team are required to better assign these values.

and standards for the metrics. However, a few outcome criteria may be context-independent. For example, lives lost and property damage could be calculated across events. This work aims to generalize to various HAT scenarios.

These identified interaction-based measures need empirical tests in team tasks that involve humans, autonomies, and interactions. Then the goal is to structure and define metrics and tease out the contextual variations and interrelations between measures. Appendix E shows our initial considerations on the testbeds. Such a testbed is to be identified or developed to explore the measures, develop metrics, and identify the relationships to team effectiveness and the variables in the IPSO model in the context of the NGCV environment.

4.5 Key Points

- The team effectiveness framework consists of inputs, interaction processes, team states, and multi-criteria outcomes.
- Interactive team cognition theory embraces interaction-based metrics.
- Interaction-based measures of HAT are well suited for measuring team states in dynamic and complex battlefields.
- Team states include team SA, team workload, team resilience, and team trust, and can be evaluated using interaction-based metrics dynamically.

5. References

Alberts DS. Agility, focus, and convergence: the future of command and control. Office of the Assistant Secretary of Defense for Networks and Information; 2007.

Allen JE, Guinn CI, Horvtz E. Mixed-initiative interaction. IEEE Intelligent Systems and Their Applications. 1999;14(5), 14–23.

Baker AL, Schaefer KE, Hill SG. Teamwork and communication methods and metrics for human-autonomy teaming. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory; 2019 Oct. Report No.: ARL-TR-8844.

Baker AL, Fitzhugh SM, Forster D, Scharine A, Brewer RW, Krausman A, Schaefer KE. Team trust in human-autonomy teams: Analysis of crew communication during manned-unmanned gunnery operations. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); 2020. Report No.: AR-TR-8969.

Barth S, Schraagen JM, Schmettow M. Network measures for characterising team adaptation processes. Ergonomics. 2015;58(8):1287–1302.

Bartlett CE, Cooke NJ. Human-robot teaming in urban search and rescue. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2015;59:250–254.

Bedwell WL, Salas E, Funke GJ, Knott BA. Team workload: a multilevel perspective. Organizational Psychology Review. 2014;4(2):99–123.

Beer JM, Fisk AD, Rogers WA. Toward a framework for levels of robot autonomy in human-robot interaction. Journal of Human-Robot Interaction. 2004;3(2):74–99.

Borgatti SP, Everett MG, Johnson JC. Analyzing social networks (second edition). SAGE Publications Ltd; 2018.

Bowers CA, Braun CC, Morgan Jr BB. Team workload: Its meaning and measurement. In Team performance assessment and measurement. Psychology Press; 1997. pp. 97–120.

Bowers C, Kreutzer C, Cannon-Bowers J, Lamb J. Team resilience as a second-order emergent state: a theoretical model and research directions. Frontiers in Psychology. 2017;8:1360.

Brewer RW, Cerame E, Pursel ER, Zimmermann A, Schaefer KE. Manned-unmanned teaming: US Army Robotic Wingman Vehicles. In: Cassenti DN (ed.), Advances in Human Factors in Simulation and Modeling. Springer International Publishing; 2019. pp. 89–100.

Burke JL, Murphy RR, Coovert MD, Riddle DL. Moonlight in Miami: Field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise. Human–Computer Interaction. 2004;19(1–2):85–116.

Cannon-Bowers JA, Salas E, Converse S. Shared mental models in expert team decision making. In: Current issues in individual and group decision making. Hillsdale (NJ): Lawrence Erlbaum; 1993. pp 221–246.

Chen JY, Barnes MJ. Human–agent teaming for multirobot control: a review of human factors issues. IEEE Transactions on Human-Machine Systems. 2014;44(1):13–29.

Chen JY, Lakhmani SG, Stowers K, Selkowitz AR, Wright JL, Barnes M. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. Theoretical Issues in Ergonomics Science. 2018;19(3):259–282.

Chiou EK, Lee JD, Su T. Negotiated and reciprocal exchange structures in human-agent cooperation. Computers in Human Behavior. 2019;90:288–297. https://doi.org/10.1016/j.chb.2018.08.012.

Cooke NJ, Gorman JC. Interaction-based measures of cognitive systems. Journal of Cognitive Engineering and Decision Making. 2009;3(1):27–46.

Cooke NJ, Gorman JC, Myers CW, Duran JL. Interactive team cognition. Cognitive Science. 2013;37(2):255–285.

Cooke NJ, Shope SM. Synthetic task environments for teams: CERTT's UAV-STE. In: Handbook of human factors and ergonomics methods. CRC Press; 2004. pp. 476–483.

Corbett B, Yamaguchi T, Liu S, Huang L, Bahn S, Nam CS. Influence of haptic feedback on a pointing task in a haptically enhanced 3D virtual environment. Human-Computer Interaction. Interaction Modalities and Techniques; 2013. 561–567. https://doi.org/10.1007/978-3-642-39330-3_60.

Cummings M. Artificial intelligence and the future of warfare. Chatham House for the Royal Institute of International Affairs; 2017.

Cummings M, Huang L, Zhu H, Finkelstein D, Wei R. The impact of increasing autonomy on training requirements in a UAV supervisory control task. Journal of Cognitive Engineering and Decision Making. 2019. https://doi.org/10.1177/1555343419868917.

Dias RD, Zenati MA, Stevens R, Gabany JM, Yule SJ. Physiological synchronization and entropy as measures of team cognitive load. Journal of Biomedical Informatics. 2019;96:103250.

DOD unmanned systems integrated roadmap FY2017–2042. Washington (DC): Department of Defense (US); 2017 [accessed 20202 Aug]. https://www.defensedaily.com/wp-content/uploads/post_attachment/206477.pdf.

Dudek G, Jenkin M, Milios E. A taxonomy of multirobot systems. Robot Teams: From Diversity to Polymorphism. 2002:3–22.

Durlach PJ. Army digital systems and vulnerability to change blindness. Orlando (FL): Army Research Institute for the Behavioral and Social Science Field Unit; 2004.

Durso FT, Hackworth CA, Truitt TR, Crutchfield J, Nikolic D, Manning CA. Situation awareness as a predictor of performance for en route air traffic controllers. Air Traffic Control Quarterly. 1998;6(1):1–20.

Endsley M. Measurement of situation awareness in dynamic systems. Human Factors. 1995a;37(1):65–84.

Endsley M. Toward a theory of situation awareness in dynamic systems. Human Factors. 1995b;37(1):32–64.

Endsley M. Design and evaluation for situation awareness enhancement. Proceedings of the Human Factors Society Annual Meeting. 1988;32:97–101.

Endsley M, Jones WM. Situation awareness information dominance & information warfare. Dayton (OH): Logicon Technical Services Inc.; 1997.

Endsley MR. Direct measurement of situation awareness: Validity and use of SAGAT. In: Situational awareness. Routledge; 2017. pp. 129–156.

Entin EE, Serfaty D. Adaptive team coordination. Human Factors. 1999;41(2):312–325.

Funke GJ, Knott BA, Salas E, Pavlas D, Strang AJ. Conceptualization and measurement of team workload: a critical need. Human Factors. 2012;54(1):36–51.

Fussell SR, Kraut RE, Lerch FJ, Scherlis WL, McNally MM, Cadiz JJ. Coordination, overload and team performance: effects of team communication strategies. In: Proceedings of the 1998 ACM conference on Computer supported cooperative work; 1998 Nov. pp. 275–284.

Gao F, Cummings M. Using discrete event simulation to model multi-robot multi-operator teamwork. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2012;56:2093–2097.

Gigerenzer G, Gaissmaier W. Heuristic Decision making. Annual Review of Psychology. 2011;62(1):451–482. https://doi.org/10.1146/annurev-psych-120709-145346.

Gorman JC, Amazeen PG, Cooke NJ. Coordination dynamics. Nonlinear Dynamics, Psychology, and Life Sciences. 2010;14(3):265.

Gorman JC, Cooke NJ. Changes in team cognition after a retention interval: the benefits of mixing it up. Journal of Experimental Psychology: Applied. 2011;17(4):303.

Gorman JC, Cooke NJ, Pederson HK, DeJoode JA. Coordinated awareness of situation by teams (CAST): measuring team situation awareness of a communication glitch. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2005;49:274–277.

Gorman JC, Cooke NJ, Winner JL. Measuring team situation awareness in decentralized command and control environments. Ergonomics. 2006;49(12–13):1312–1325.

Gorman JC, Demir M, Cooke NJ, Grimm DA. Evaluating sociotechnical dynamics in a simulated remotely-piloted aircraft system: A layered dynamics approach. Ergonomics. 2019;62(5):629–643.

Guo C, Sharlin E. Exploring the use of tangible user interfaces for human-robot interaction: a comparative study. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008:121–130. https://doi.org/10.1145/1357054.1357076.

Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Hancock PA, Meshkati N (eds.), Advances in psychology, 52. Human mental workload. North-Holland; 1988. p. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Hewitt C, Woods J. Inconsistency robustness. College Publications; 2015.

Hill S, Stachowiak C, Tauson R. Soldier performance in the enclosed compartment of a moving vehicle. NATO Appl Veh Technol Panel Symp, Habitability

Combat Transp. Veh: Noise, Vibration Motion; 2004. Paper RTO-MP-AVT-110.

Hoff K, Bashir M. Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors. 2015;57(3):407–434.

Hoffman RR, Hancock PA. Measuring resilience. Human Factors. 2017;59(4):564–581.

Holder E. Defining Soldier intent in a human-robot natural language interaction context. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2017 Oct. Report No.: ARL-TR-8195.

Hollan J, Stornetta S. Beyond being there. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1992:119–125.

Hollnagel E. The four cornerstones of resilience engineering. In: Resilience Engineering Perspectives, Vol. 2. CRC Press; 2016. pp. 139–156.

Huang C-M, Cakmak M, Mutlu B. Adaptive coordination strategies for human-robot handovers. Robotics: Science and Systems; 2015.

Huang L, Cooke N, Gutzwiller RS, Berman S, Chiou E, Demir M, Zhang W. Distributed dynamic team trust in human, artificial intelligence, and robot teaming. In: Trust in Human-Robot Teaming; in press.

Huang L, Cummings M, Nneji VC. Preliminary analysis and simulation of railroad dispatcher workload. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2018;62:691–695.

Huang L, Cummings M, Ono M. A mixed analysis of influencing factors for trust in a risk-aware autonomy experiment. Human Factors and Erognomics Society Annual Conference; in review.

Huang L, Johnson C, Cooke N, Holder E. Interaction analysis for the development of NGCVs. Task 1 in this Report. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); 2019.

Hughes AM, Hancock GM, Marlow SL, Stowers K, Salas E. Cardiac measures of cognitive workload: a meta-analysis. Human Factors. 2019;61(3):393–414.

Hutchins S, Cosenz, K, Barnes M, Feng T, Pillalamarri K. Soldier-robot teaming: Effects of multimodal collaboration on team communication for robot reconnaissance. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2010. Report No.: ARL-TR-5385.

Ilgen DR, Hollenbeck JR, Johnson MD, Jundt DK. Teams in organizations: From input-process-output models to IMOI models. Annual Review of Psychology; 2005. https://doi.org/10.1146/annurev.psych.56.091103.070250.

Jenkins Q, Jiang X. Measuring trust and application of eye tracking in human robotic interaction. IIE Annual Conference. Proceedings; Norcross. 2010:1–6.

Jeon Y, Nam CS, Kim Y-J, Whang MC. Event-related (De)synchronization (ERD/ERS) during motor imagery tasks: Implications for brain–computer interfaces. International Journal of Industrial Ergonomics. 2011;41(5):428–436. https://doi.org/10.1016/j.ergon.2011.03.005.

Jian J-Y, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics. 2000;4(1):53–71.

Johnson C, Lematta G, Huang L, Holder E, Bhatti S, Cooke N.. An Interaction taxonomy of human–agent teaming in next generation combat vehicle systems. Proceedings of the AHFE 2020 International Conference on Human Factors in Robots and Unmanned Systems; in press.

Johnson M, Bradshaw JM, Feltovich PJ, Jonker CM, Van Riemsdijk MB, Sierhuis M. Coactive design: designing support for interdependence in joint activity. Journal of Human-Robot Interaction. 2014;3(1):43–69.

Klein G, Woods DD, Bradshaw JM, Hoffman, RR, Feltovich PJ. Ten challenges for making automation a "team player" in joint human-agent activity. IEEE Intelligent Systems, 2004;19(06):91–95. https://doi.org/10.1109/MIS.2004.74.

Lee JD, See KA. Trust in automation: Designing for appropriate reliance. Human Factors. 2004;46(1):50–80. https://doi.org/10.1518/ hfes.46.1.50_30392.

Lematta GJ, Johnson CJ, Chiou EK, Cooke NJ. Does team interaction exploration support resilience in human autonomy teaming? Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2019;63:1866–1866.

Lenné MG, Hoggan BL, Fidock J, Stuart G, Aidman E. The impact of auditory task complexity on primary task performance in military land vehicle crew. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2014;58:2185–2189.

Lester JC, Converse, SA, Kahler SE, Barlow ST, Stone BA, Bhogal RS. The persona effect: affective impact of animated pedagogical agents. CHI, 97. 1997:359–366.

Marks MA, Mathieu JE, Zaccaro SJ. A temporally based framework and taxonomy of team processes. The Academy of Management Review. 2001;26(3):356–376. JSTOR. https://doi.org/10.2307/259182.

Mathieu J, Maynard MT, Rapp T, Gilson L. Team effectiveness 1997–2007: a review of recent advancements and a glimpse into the future. Journal of Management. 2008;34(3):410–476.

Mayer RE. Multimedia learning. Cambridge University Press; 2001.

Meyer J, Wiczorek R, Günzler T. Measures of reliance and compliance in aided visual scanning. Human Factors. 2014;56(5):840–849.

Milner A, Seong DH, Brewer RW, Baker AL, Krausman A, Chhan D, Thomson R, Rovira E, Schaefer KE. Identifying new team trust and team cohesion metrics that support future human-autonomy teams. Proc. Applied Human Factors and Ergonomics Society. 2020.

Mitchell DK. Workload analysis of the crew of the Abrams V2 SEP: Phase I baseline IMPRINT model. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2009 Nov. Report No.: ARL-TR-5028.

Murphy RR, Burke JL. The safe human-robot ratio. In: Human-robot interactions in future military operations. CRC Press; 2016. pp. 51–70.

Norman DA. The psychology of everyday things.(The design of everyday things). Basic Books; 1988.

Ono M, Pavone M, Kuwata Y, Balaram J. Chance-constrained dynamic programming with application to risk-aware robotic space exploration. Autonomous Robots. 2015;(4):555–571.

Orasanu J, Salas E. Team decision making in complex environments. In Klein GA, Orasanu J, Calderwood R, Zsambok CE (eds.), Decision making in action: Models and methods. Ablex Publishing; 1993. p. 327–345.

Patterson ES, Roth EM, Woods DD, Chow R, Gomes JO. Handoff strategies in settings with high consequences for failure: lessons for health care operations. International Journal for Quality in Health Care. 2004;16(2):125-132.

Plott BM, McDermott PL, Barnes M. Advanced Video Activity Analytics (AVAA): human performance model report. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2017 Dec. Report No.: ARL-TR-8255.

Rajivan P, Cooke N. Impact of team collaboration on cybersecurity situational awareness. In: Lecture notes in computer science (including subseries Lecture notes in artificial intelligence and lecture notes in bioinformatics) (Vol. 10030). Springer Verlag; 2017. pp. 203–226. https://doi.org/10.1007/978-3-319-61152-5_8.

Reina D, Reina M, Hudnut D. Why trust is critical to team success. Center for Creative Leadership; 2017 [accessed 2020 Jan 8]. https://www.ccl.org/wp-content/uploads/2017/05/why-trust-is-critical-team-success-research-report.pdf.

Rouse WB, Morris NM. On looking into the black box: prospects and limits in the search for mental models. Psychological Bulletin. 1986;100(3):349.

Russell SJ, Norvig P. Artificial intelligence: a modern approach. Malaysia: Pearson Education Limited; 2016.

Salas E, Cooke NJ, Rosen MA. On teams, teamwork, and team performance: discoveries and developments. Human Factors. 2008;50(3):540–547.

Salas E, Sims DE, Burke CS. Is there a "big five" in teamwork? Small Group Research. 2005;36(5):555–599.

Salmon PM, Stanton N, Jenkins DP. Distributed situation awareness: theory, measurement and application to teamwork. CRC Press; 2017.

Save L, Feuerberg B, Avia E. Designing human-automation interaction: a new level of automation taxonomy. Proc. Human Factors of Systems and Technology. 2012.

Schaefer KE, Straub ER, Chen JY, Putney J, Evans III AW. Communicating intent to develop shared situation awareness and engender trust in human-agent teams. Cognitive Systems Research. 2017;46:26–39.

Schaefer KE, Baker AL, Brewer RW, Patton D, Canady J, Metcalfe JS. Assessing multi-agent human-autonomy teams: US Army robotic wingman gunnery operations. In: Proceedings of the SPIE Defense + Commercial Sensing, Micro- and Nanotechnology Sensors, Systems, and Applications XI Conference; 2019 May; Baltimore, MD. doi://10.1117/12.2519302.

Sellers J, Helton WS, Näswall K, Funke GJ, Knott BA. Development of the team workload questionnaire (TWLQ). Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2014;58:989–993.

Shattuck LG, Woods DD. Communication of intent in military command and control systems. In: The Human in Command. Springer; 2000. p. 279–291.

Sheridan TB. Risk, human error, and system resilience: fundamental ideas. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2008;50(3):418–426. doi: 10.1518/001872008X250773.

Singh J. Medical issues in aviation: flight crew, circadian rhythms, fatigue, alertness, & long haul operations. 2009. https://slideplayer.com/slide/13330492/.

Spiegelhalter D. Risk and uncertainty communication. Annual Review of Statistics and Its Application. 2017;4:31–60.

Spiker VA, Holder EW, Walls WF, Campsey WM, Bruce PD. SamePage: development of a team training tool to promote shared understanding. Santa Barbara (CA): Anacapa Sciences Inc.; 2007.

Stanton N. Representing distributed cognition in complex systems: how a submarine returns to periscope depth. Ergonomics. 2014;57(3):403–418.

Stanton N, Harvey C. Beyond human error taxonomies in assessment of risk in sociotechnical systems: a new paradigm with the EAST 'broken-links' approach. Ergonomics. 2017;60(2) 221–233.

Stanton N, Salmon PM, Walker GH, Salas E, Hancock PA. State-of-science: Situation awareness in individuals, teams and systems. Ergonomics. 2017;60(4):449–466.

Stanton N, Salmon P, Walker G. Systems thinking in practice: applications of the event analysis of systemic teamwork method. CRC Press; 2018.

Stephens RJ, Woods DD, Branlat M, Wears RL. Colliding dilemmas: Interactions of locally adaptive strategies in a hospital setting. 4th Resilience Engineering International Symposium; 2011.

Stimpson AJ, Ryan JC, Cummings ML. Assessing pilot workload in single-pilot operations with advanced autonomy. Proc. Human Factors and Ergonomics Society Annual Meeting. 2016;60:675–679. http://journals.sagepub.com/doi/abs/10.1177/1541931213601155.

Taylor RM. Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Situational awareness in aerospace operations AGARD-CP-478); 1990; Neuilly Sur Seine, France. NATO-AGARD. pp. 3/1–3/17.

US Army. FM 7-8: Infantry rifle platoon and squad. Washington (DC): Headquarters, Department of the Army; 1992.

US Army. FM 3-20.21: Heavy Brigade Combat Team (HBCT) gunnery. Fort Knox (KY): Directorate of Training, Doctrine, Combat Development, and Experimentation, Armor Center (US); 2008.

US Army. ATP 3-20.15. Tank platoon. Washington (DC): Headquarters, Department of the Army; 2019.

USNRC. PRA procedures guide: a guide to the performance of probabilistic risk assessments for nuclear power plants (NUREG/CR-2300). US Nuclear Regulatory Commission; 1983. http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr2300/.

Walker GH, Gibson H, Stanton N, Baber C, Salmon P, Green D. Event analysis of systemic teamwork (EAST): a novel integration of ergonomics methods to analyse C4i activity. Ergonomics. 2006;49(12–13):1345–1369.

Wickens CD. Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science. 2002;3(2):159–177.

Wickens CD. Multiple resources and mental workload. Human Factors; 2008;50(3):449–455.

Wickens CD, Hollands JG. Signal detection, information theory, and absolute judgment. Engineering Psychology and Human Performance. 2000;2:24–73.

Woods DD, Tittle J, Feil M, Roesler A. Envisioning human-robot coordination in future operations. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2004;34(2):210–218.

Woods DD. Four concepts for resilience and the implications for the future of resilience engineering. Reliability Engineering & System Safety.2015;41:5–9.

Woods DD. The theory of graceful extensibility: basic rules that govern adaptive systems. Environment Systems and Decisions. 2018;38(4):433 –457.

Yanco HA, Drury JL. Classifying human-robot interaction: an updated taxonomy. 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583). 2004;3:2841–2846 vol.3. https://doi.org/10.1109/ICSMC.2004.1400763.

Appendix A. Structured Interview Questions for the Field Trip

Date and Location: March 19th, 2019 at Ft. Huachuca, AZ

Questions for SMEs with MTC experience: Prior mechanized platoon leader/sergeants/vehicle commanders

- Can you briefly describe your role(s) during MTC?
  - Eric mentioned that these people have served in different roles before. So after we ask about their roles, we need to be specific about which role we are interested in to ask more questions.
- What are the typical scenarios in MTC have you encountered?
  - Going through the mission, nothing happens; Actions on contact (Eight forms of contact: Direct, Indirect, Non-Hostile, Obstacle, CBRN, Aerial, Visual, Electronic Warfare), a transition to a different task
  - Provide some scenarios ahead of time and also ask them for input
- What tasks are involved in your role in these scenarios?
  - What is the procedure for the task?
  - What information is needed to do these tasks?
  - How critical are these tasks? Criticality depends on the severity of the consequences.
  - How often do you do these tasks?
  - How long does each task last?
- What are the greatest (mental, physical, or team) challenges in these tasks? Please give some examples you have experienced.
  - Examples of incidents
- What tools have you used for your tasks during MTC?
  - What do you like about the tools/functions?
  - What do you dislike about the tools/functions?
  - What tools or functions of a tool you would prefer to have to assist you in completing your tasks?
- Who did you interact with to conduct these tasks?
  - Who did you provide information or orders to/from?
  - What are their roles in the tasks? How do they impact your tasks?
  - Exercises: Under a scenario, draw interactions between agents
- Communication/interactions:
  - What tools do you use to communicate with people inside the vehicle and elsewhere?
  - Under what circumstances, did you need to communicate with operators of other vehicles?

- Teamwork
    - What did you see good examples of good teamwork during MTC?
    - What are examples due to the failure of communication or teamwork?
- Performance Measurement
    - What performance criteria matters the most?
    - How do you evaluate individual performance?
    - How do you evaluate team performance?
- Any other special issues about the tasks you want to talk about?
    - Is there any trust issue in the tasks?

Appendix B. Task Analysis

Task Class	Task	Subtask	Taskwork	Teamwork	Essential Interaction
1 Gunnery	1.1 Prepare and maintain weapon systems	1.1.1 Zero/calibrate weapon sights	X		
		1.1.2 Communicate status of zero/sight calibration		X	
		1.1.3 Conduct pre-fire checks for weapon, sight, and fire control system serviceability	X		
		1.1.4 Communicate status of pre-fire checks		X	
		1.1.5 Load ammunition	X		
		1.1.6 Communicate status of ammunition		X	
		1.1.7 Identify weapon or sight malfunctions	X		
		1.1.8 Communicate weapon or sight malfunctions		X	
		1.1.9 Repair weapon or sight malfunctions	X		
	1.2 Detect surroundings for potential targets/ conduct surveillance	1.2.1 Operate weapons, sights, and/or turret to sense surroundings	X		
		1.2.2 Communicate to operate weapons, sights, or turret to sense surroundings		X	
		1.2.3 Designate sectors of responsibility to observe and scan for targets	X		
		1.2.4 Communicate sectors of responsibility		X	X
		1.2.5 Establish sector limits based on direction or reference points	X		
		1.2.6 Communicate sector limits		X	
		1.2.7 Monitor within sectors of responsibility using weapons, sights, and/or turret	X		
		1.2.8 Communicate status of sector of responsibility		X	
		1.2.9 Direct crew to change sector monitoring		X	
		1.2.10 Establish search technique, areas, scan rate, and/or search pattern	X		
		1.2.11 Communicate search technique, area, scan rate, and/or search pattern		X	

		1.2.12 Detect potential targets through indirect signals (e.g., dust, smoke) and direct target signals (e.g., people, vehicles)	X		
		1.2.13 Communicate detection of potential targets		X	X
		1.2.14 Determine location of potential targets (distance, direction, movement direction)	X		
		1.2.15 Communicate location of potential targets		X	X
		1.2.16 Move weapons, sights, or turret to potential targets	X		
		1.2.17 Communicate to move weapons, sights, or turret to potential targets		X	
	1.3 Identify targets	1.3.1 Classify potential target (e.g., tank, PC, truck, dismounts)	X		
		1.3.2 Communicate potential target classification		X	X
		1.3.3 Identify potential target by features (e.g., model, markings, activity e.g. T-80)	X		
		1.3.4 Communicate potential target identity		X	X
		1.3.5 Classify target alignment (i.e., friendly, enemy, neutral)	X		
		1.3.6 Communicate target alignment		X	X
	1.4 Determine target engagement	1.4.1 Determine target priority based on target types, orientations, and activities	X		
		1.4.2 Communicate target priority based on target types, orientations, and activities.		X	
		1.4.3 Determine range to target	X		
		1.4.5 Communicate range to target		X	X
		1.4.6 Determine engagement method (weapon, ammo, technique)	X		
		1.4.7 Communicate target engagement method (weapon, ammo, technique)		X	X
		1.4.8 Decide whether or not to engage the target	X		
		1.4.9 Communicate about decision to engage the target		X	X

	1.5 Engage targets	1.5.1 Give command to fire		X	X
		1.5.2 Fire weapon	X		
		1.5.3 Confirm firing execution (verbally)		X	X
		1.5.4 Identify misfires / malfunctions	X		
		1.5.5 Communicate misfires / malfunctions		X	
		1.5.6 Correct misfires / malfunctions	X		
	1.6 Call for indirect fire support	1.6.1 Identify need for indirect fire support on a specific target or location	X		
		1.6.2 Communicate need for fire support (within crew or unit)		X	X
		1.6 3 Communicate target information (within crew or unit)		X	X
		1.6.4 Contact higher HQ to request fire support with target's information (e.g., location, model, activity)		X	X
	1.7 Assess targets	1.7.1 Assess where rounds impact	X		
		1.7.2 Communicate round impact (e.g. short, left, over)		X	
		1.7.3 Assess the status of the target	X		
		1.7.4 Communicate the status of target		X	
		1.7.5 Re-engage target	X		
		1.7.6 Give re-engagement fire command		X	
		1.7.7 Give cease fire command		X	X
2 Mobility	2.1 Prepare and maintain vehicle systems	2.1.1 Maintain consumables (e.g., fuel, oil, battery, etc.)	X		
		2.1.2 Communicate consumable status		X	
		2.1.3 Assess vehicle for physical damage and malfunctions	X		
		2.1.4 Communicate physical damage or malfunctions		X	
		2.1.5 Repair the vehicle	X		

		2.1.6 Request resources to repair		X	
	2.2 Drive	2.2.1 Move along covered and concealed routes	X		
		2.2.2 Communicate route (local, where we are going right now)		X	X
		2.2.3 Occupy battle positions	X		
		2.2.4 Communicate battle positions		X	
		2.2.5 Avoid hazards and obstacles along route	X		
		2.2.6 Communicate hazards and obstacles along route		X	
		2.2.7 Maintain heading	X		
		2.2.8 Communicate heading		X	
		2.2.9 Maintain speed	X		
		2.2.10 Communicate speed		X	
		2.2.11 Move vehicle on a variety of terrain types (roads, cross-country, dirt, mud, sand, rocks)	X		
		2.2.12 Communicate with driver to move vehicle over difficult terrain		X	
		2.2.13 Move vehicle in road traffic	X		
		2.2.14 Communicate with driver to move vehicle in traffic		X	
		2.2.15 Move within a formation	X		
		2.2.16 Communicate with driver to move within a formation		X	
		2.2.17 Communicate with adjacent units or crews to synchronize movement (e.g. bounding)		X	
		2.2.18 Move vehicle across water fords	X		
		2.2.19 Communicate with driver to move vehicle across water fords		X	
	2.3 Navigation	2.3.1 Read maps or aerial photographs utilizing map symbols, grids, overlays, and marginal map information	X		

		2.3.2 Localize the vehicle by resection or intersection methods	X		
		2.3.3 Determine possible routes via map reconnaissance	X		
		2.3.4 Select a route utilizing OAKOC information	X		
		2.3.5 Perform terrain association by comparing map information with environmental and terrain information	X		
		2.3.6 Perform dead reckoning by localizing a polar coordinate	X		
		2.3.7 Communicate vehicle location		X	
		2.3.8 Communicate distance or time to checkpoints or objectives		X	
		2.3.9 Communicate selected route		X	
		2.3.10 Communicate about obstacles		X	
		2.3.11 Recognize objective	X		
		2.3.12 Communicate the location of the objective		X	
		2.3.13 Communicate route progress		X	
3 Actions on contact	3.1 Identify and report contact	3.1.1 Identify contact type and location	X		
		3.1.2 Communicate contact type and location (within crew)		X	X
		3.1.3 Report contact type and location to higher HQ (SALUTE / SPOTREP)		X	X
	3.2 React to direct or indirect fire contact	3.2.1 Move to covered battle position that provides observation and fields of fire	X		
		3.2.2 Communicate battle position		X	
		3.2.3 Communicate movement		X	X
		3.2.4 If under fire, use direct or indirect fire to suppress enemy	X		
	3.3 Develop situation and choose a course of action	3.3.1 Identify enemy size, composition, and capabilities	X		
		3.3.2 Identify enemy location, orientation, and direction of movement	X		
		3.3.3 Communicate enemy size, composition, and capabilities		X	

		3.3.4 Communicate enemy location, orientation, and direction of movement		X	
		3.3.5 Assess impact of obstacles and terrain on situation	X		
		3.3.6 Communicate impact of obstacles and terrain on situation		X	
		3.3.7 Assess friendly situation (location, strength, capabilities)	X		
		3.3.8 Communicate friendly situation (location, strength, capabilities)		X	
		3.3.9 Identify available courses of action	X		
		3.3.10 Communicate available courses of action		X	
		3.3.11 Select a course of action	X		
		3.3.12 Communicate selected course of action to supervisor		X	X
		3.3.13 Change course of action	X		
		3.3.14 Communicate changes to course of action		X	
	3.4 Execute selected course of action	3.4.1 Communicate course of action to crew / unit		X	X
		3.4.2 Execute selected course of action	X		
		3.4.3 Continue mission or transition from Movement to Contact	X		
		3.4.4 Communicate continue mission or transition from Movement to Contact		X	X
4 Crew /automation management	4.1 Prepare/ maintain communications equipment	4.1.1 Communicate required communication system settings		X	
		4.1.2 Assess required communication settings	X		
		4.1.3 Set-up radio	X		
		4.1.4 Set-up intercom	X		
		4.1.5 Set-up text chat	X		
		4.1.6 Set-up digital map system	X		
		4.1.7 Conduct communications checks		X	X

		4.1.8 Identify communication system malfunctions	X		
		4.1.9 Communicate communication system malfunctions		X	
		4.1.10 Repair communication system malfunctions	X		
	4.2 Monitor communications equipment	4.2.1 Monitor radio channel		X	
		4.2.2 Monitor intercom channel		X	
		4.2.3 Monitor text chat channel		X	
		4.2.4 Monitor digital maps system channel		X	
		4.2.5 Communicate about communication channels		X	
	4.3 Monitor crew status	4.3.1 Request crew report		X	
		4.3.2 Communicate status of crewmembers		X	
		4.3.3 Communicate status of vehicle		X	
		4.3.4 Communicate status of weapons, sights, and ammunition		X	
		4.3.5 Monitor crew status (injuries, hydration, fatigue)		X	
	4.5 Provide or request medical support	4.5.1 Identify need for medical support	X		
		4.5.2 Request medical support		X	
		4.5.3 Conduct individual self-aid	X		
		4.5.4 Conduct crewmember first aid	X		
		4.5.5 Communicate medical evacuation (CASEVAC/ MEDEVAC)		X	X
		4.5.6 Move to casualty evacuation point	X		
	4.6 Share mission	4.6.1 Receive the mission from supervisor		X	X
		4.6.2 Issue warning order (early guidance on mission, intent, and execution timeline)		X	
		4.6.3 Issue the operations order (situation, mission, execution, sustainment, communications plan)		X	X

		4.6.4 Conduct back brief to confirm understanding of mission		X	
		4.6.5 Conduct mission rehearsals to enhance understanding of mission		X	
	4.7 Mobility mode selection	4.7.1 Identify need to change mobility autonomy mode	X		
		4.7.2 Communicate need to change mobility autonomy mode		X	
		4.7.3 Select mobility autonomy mode	X	X*	X
		4.7.4 Communicate mobility autonomy mode		X	
		4.7.5 Monitor mobility autonomy status		X	
	4.8 Gunnery mode selection	4.8.1 Identify need to change gunnery autonomy mode	X		
		4.8.2 Communicate need to change gunnery autonomy mode		X	
		4.8.3 Select gunnery autonomy mode	X	X*	X
		4.8.4 Communicate gunnery autonomy mode		X	
		4.8.5 Monitor gunnery autonomy status		X	
	4.9 Change role	4.9.1 Identify need to change role	X		
		4.9.2 Communicate need to change role		X	
		4.9.3 Select new role	X	X*	X
		4.9.4 For between agent role transfers execute role transfer procedures		X	X
		4.9.5 Monitor role status		X	
	4.10 Gain control of RCV	4.10.1 Identify need to gain control	X		
		4.10.2 Communicate need to gain control		X	
		4.10.3 Select RCV	X	X*	X
		4.10.4 Monitor control of RCV		X	

* The designation as teamwork centers on the fact that we are focusing on human-autonomy teams and this is an interaction between the human and an intelligent autonomous agent.

Appendix C. Taxonomy Diagrams

Human–autonomy teaming interaction taxonomy consists of three components: team composition, communication, and tasks (Figs. C-1 through Fig. C-4). Each component has a hierarchical structure of dimensions. Under the team composition, the human–vehicle ratio varies by the team composition, and therefore, does not have discrete categories. The agent formal roles are unknown at this time, but are anticipated to be defined roles similar to current "gunner", "tank commander", and so on in the NGCVs. The task analysis (see Appendix B) also attached to this report provides further detail on the task category.

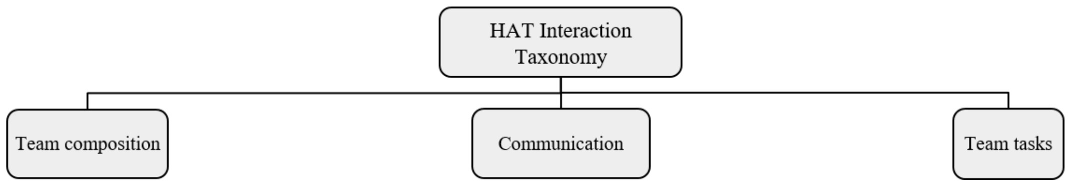

Fig. C-1 Human–autonomy teaming interaction taxonomy

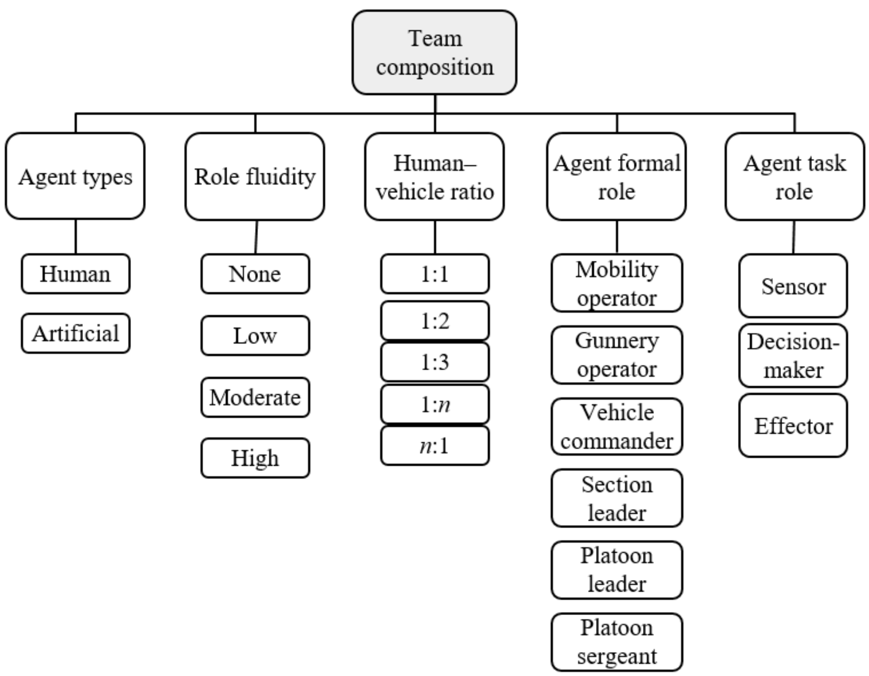

Fig. C-2 Taxonomy of task composition

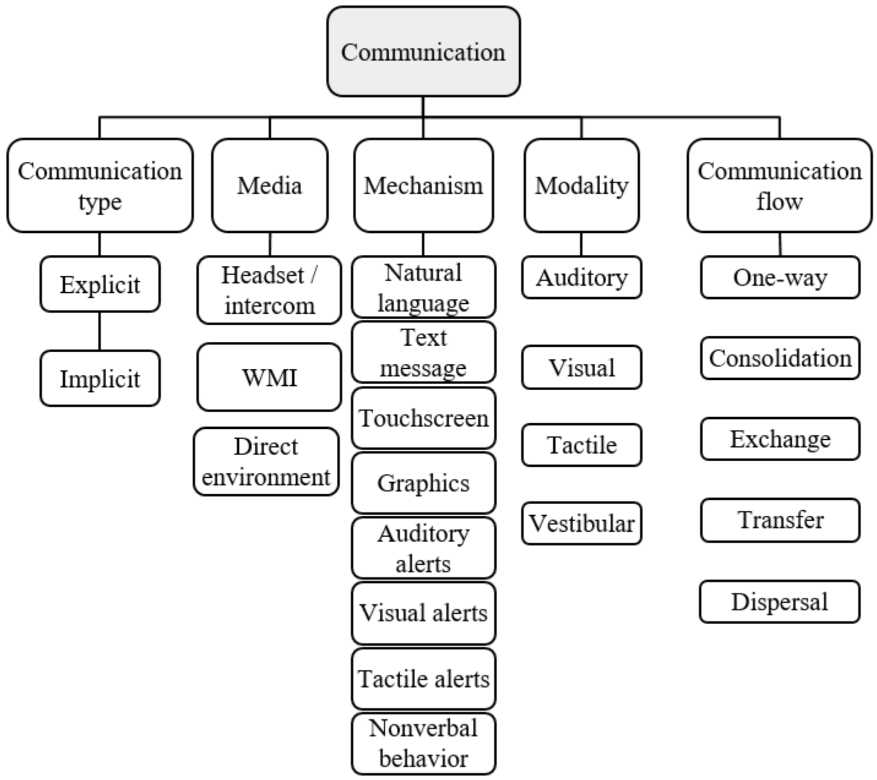

Fig. C-3 Taxonomy of team communication

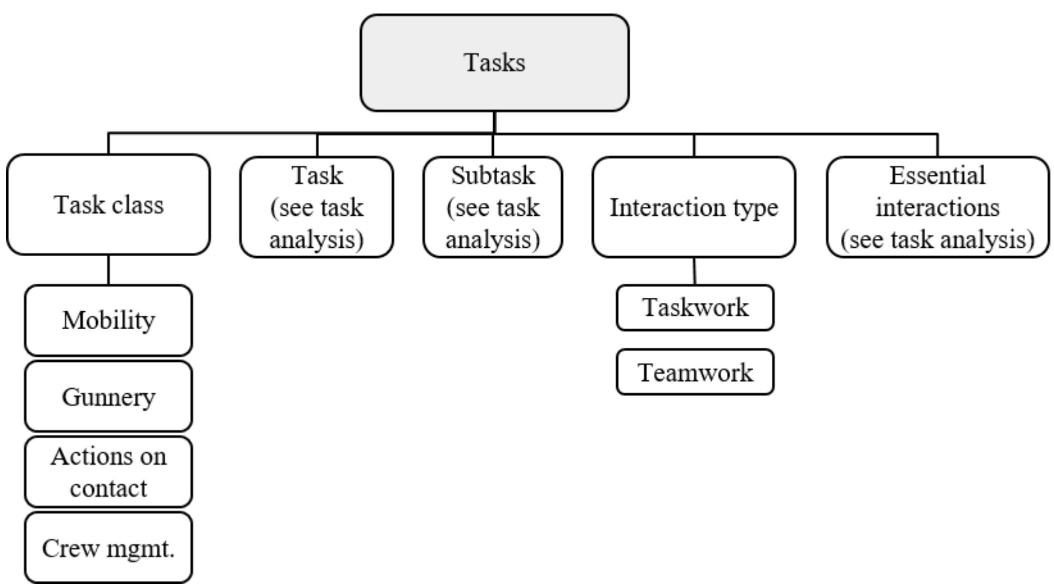

Fig. C-4 Taxonomy of team tasks

Appendix D. The Applied Human Factors and Ergonomics (AHFE) 2019 Conference

Mr. Ralph W. Brewer (CCDC ARL), a session chair for the session of Human–Agent Teaming: Soldiers, Sailors, Airmen, and Marines, in the affiliated conference of Human Factors in Robots and Unmanned Systems (HFRUS) at the 10th International Conference on Applied Human Factors and Ergonomics (AHFE 2019) invited Dr. Nancy Cooke to submit an abstract to his session, with a full paper as an option. We submitted an abstract in December 2018 without a full paper because we just started with the project and did not have any data or solid works at the submission deadline.

Huang, L., Johnson, C.J., Cooke, N.J., & Holder, E. (2019). Human–Agent Teaming with Next Generation Combat Vehicles (NGCVs): Interaction Metrics and Models. Abstract presented at *the AHFE 2019 International Conference on Human Factors in Robots and Unmanned Systems*, July 24–28, 2019, Washington D.C., USA.

Objective: The current study seeks to analyze human–agent teaming within the context of the Next Generation Combat Vehicles (NGCVs) system and to provide design guidance to facilitate and measure teaming performance.

Significance: Traditional armored vehicles (e.g., M1 Abrams Tank and M2 Bradley Fighting Vehicle) have been demonstrated to be an effective means of providing mobile, protected firepower on the battlefield. However, the risk of losing a human life in combat is a critical issue and should be minimized. With the Next Generation Combat Vehicle, two unmanned robotic combat vehicles (RCVs) are used as wingmen to protect the lives in the Manned Combat Vehicles (MCVs). However, controlling an unmanned vehicle when sitting in a manned vehicle creates a new task procedure that requires appropriately reallocating functions and providing effective interfaces to support human decision making and team performance. A solution is to understand the mission, identify germane specific tasks and procedures, facilitate good communication, design interfaces that respect individuals' workload and staffing constraints, and train people adequately for the new task environment. As a step in this direction, a taxonomy of human–agent interactions was developed with the goal of associating interactions with optimal interaction methods.

Method: This study focused on the tactical task *Movement to Contact (MTC)*—moving to develop the situation and establish or regain contact. Based on materials related to traditional armored vehicles and the NGCVs, the researchers created a taxonomy of human–agent teaming interactions regarding NGCVs in MTC. Then the researchers will interview soldiers and military researchers to verify the accuracy of the taxonomy and refine it.

Results: *Task Analysis and Taxonomy.* NGCVs are different from the traditional tank platoon in terms of 1) *staffing*—a transition from a 4-person crew (tank: tank commander, gunner, driver, loader) to a 2-person crew (NGCV: potentially driver/navigator and tank commander/gunner,) in a MCV; and 2) operation *angles*—direct control vs. teleoperation; viewing sources and angles also differ. These differences will impact human–agent interactions at various subtasks in MTC, such as navigating vehicles and searching for a target in surroundings. In navigating and driving an RCV in NGCVs, the combined driver/navigator role will need to complete tasks used to be done by two people. In searching for and confirming targets, the two RCV crew view from a remote-controlled angle rather than three different direct views from the vehicle they locate.

The taxonomy of human–agent interactions identified communications (i.e., directions, contents, and approaches) and operations (e.g., manual vs. supervisory control) between human–human, human–RCV, and human–MCV under the subtasks in MTC. Action initiator, receiver, and interaction content (i.e., information/action) for each interaction were listed.

Design guidance and evaluation metrics. Repetitive interactions that do not necessarily require human involvement should be automated to allow operators to do high-level decision-making tasks that cannot be automated. Technology should also provide additional information feeds to compensate for the loss of situation awareness when tele-operating the vehicles. Individuals' workload, shared mental model, task completion rate and time, and the effectiveness of communication were used to measure human–agent teaming.

Appendix E. Testbeds

The second step of this project requires a suitable testbed to empirically verify the teamwork metrics. Points listed in this section are part of ongoing development that will be refined and carried out in the following years.

E.1 Testbed Requirements

Below is a general set of requirements for the testbed to capture the key dynamics.

- A supervisor/coordinator role to align overall activities to the mission/goal
- At least one additional smaller team, possibly 2 working a task that requires internal coordination within that smaller team with inputs and outputs relevant to others in the larger team but also coordination that is not. The other member of that smaller team or a second team could be an automated agent or confederate or scripted but information exchange has to be required.
- A way to examine how "nets" of information impact coordination and performance
  - Information specific to task members only (e.g., search, find, respond, maneuver vehicle and sensors) with a reporting/reporter relationship to the larger team when relevant
  - Information that applies to the whole team
- Synchronized movement of separated entities in time and space (geo-located coordination and references of multiple moving parts)
  - This could be various combinations of air and ground and control vehicle
- The entities at a minimum have to include unmanned remotely controlled assets with the ability to vary the levels of autonomous behaviors possible
  - that the Roboleader study done on the MIX simulation included an intelligence agent to plan and coordinate the Unmanned vehicle (air and ground) routes
- The information for the remote entities or their actions has to be critical to mission success. They have to be an essential part of the mission/team.

E.2 Possible Independent Variables

- Role of the coordinator/leader
  - tasked or not

    - Amount of coordination required
- Task organization
    - Set roles
    - Fluid roles
    - The number of members working together on a task
- Distribution of information by nets
    - How they are set up (assigned, participant customized)
    - How information is shared to different nets (tracking the flow)
- Impact of variations in the autonomy provided
    - Obstacle avoidance
    - Route selection
    - Automated task completion: sensor employment, target recognition, movement in formations, etc.
- Variations in the information content
    - Reliability (source, temporal delay, etc.)
    - Timeliness
    - Amount of detail
    - Portrayal as text, audio, graphic, etc.

E.3 Dependent Variables

* Team states (e.g., team trust, team situation awareness, team workload, team cohesion, team resilience, etc.)
* Team outcomes (e.g., mission success, performance score)
    - Interactions and information flow as indicators of team performance
    - Objective measures of team performance related to mission (examples to be adapted to context):
        - Target ID: Targets found/dealt with vs missed
- Errors/Accidents (civilian or friendly injured, navigation errors, crashes, etc.)

- Timing measures

E.4 Testbed Options

During the first contract year, we have examined several options of testbeds. Michael Barnes connected us to Julia Wright and Shan Lakhmani at ARL Florida, as well as Daniel Barber at the University of Central Florida, regarding the Mix platform. The options listed in Table E. are still undergoing evaluation.

Table E-1 Testbed options

ID	testbeds	Accessibility	Validity	Modification	Resolution	Other notes	Decision
1	Call of Duty video game	Easy	Teamwork not required	No	Low-medium		Rejected
2	UCF Mix Platform	Medium	Teamwork not required	Yes, but too much to be changed by someone we do not have direct control.	Low	Hard to coordinate with the team to work on our and the other ARL project.	Not likely
3	APG INFORMS lab	Extremely difficult	Most authentic	Not needed	High		Not likely
4	APG Wingman system	Difficult	5-Man wingman is close	No (resource)	Medium-high		Not likely
5	Minecraft	Easy	Constraints and capabilities explored.	Yes	Low	Modification greatly improves data accessibility and customizability.	Promising
6	ARMA III video game	Easy	Constraints and capabilities unexplored.	Yes	Medium	High learning curve for participants; high cost	Less likely but unsure
7	ASU customized testbed (donated physical environment + Pheeno cars + Galileo interface)	Easy	Will be customized for our 7-man context	Yes	Medium	Mixed reality; onsite	Promising

List of Symbols, Abbreviations, and Acronyms

3D	three-dimensional
ACE	ammunition, casualty, and equipment report
AiTR	aided target recognition
ARL	Army Research Laboratory
ATR	automatic target recognition
BDA	battle damage assessment
CAST	coordinated awareness of situation by teams
CCDC	US Army Combat Capabilities Development Command
CITV	commander's independent thermal viewer
CFV	Cavalry Fighting Vehicle
COP	common operating picture
%DET	percentage of determinism
DOD	Department of Defense
EAST	Event Analysis of Systemic Teamwork
FBCB2	Force XXI Battle Command Brigade and Below
FM	Field Manual
FRAGO	fragmentary order
GOTWA	(G) where I'm Going, (O) Others I'm taking, (T) Time of my return, (W) What to do if I don't return, (A) Actions to take if I'm hit or Actions to take if you're hit
GPS	global positioning system
HAT	human–autonomy teaming
HHQ	higher headquarters
IED	improvised explosive device
IMO	Input-Mediator-Output
IMPRINT	Improved Performance Research Integration Tool
IMOI	Input–Mediator–Output–Input
IPSO	Inputs–Processes–States–Outcomes
iSAFT	incorrect SA in Failed Team Tasks

ITC	interactive team cognition
MCV	manned combat vehicle
METT–TC	Mission, Enemy, Terrain, Troops available, Time, and Civil considerations
MFV	maned fighting vehicle
MUMT	manned–unmanned teaming
MTC	movement-to-contact
NGCV	Next Generation Combat Vehicle
NASA	National Aeronautics and Space Administration
NASA-TLX	NASA Task Load Index
OPORD	Operations Orders
PoC	point of contact
RCV	robotic combat vehicle
ROEs	rules of engagement
SA	situation awareness
SAGAT	Situation Awareness Global Assessment Technique
SALUTE	size, activity, location, unit identification, time, and equipment
SART	Situation Awareness Rating Technique
SITREPS	situational reports
SME	subject-matter expert
SNA	Social Network Analysis
SOPs	standard operating procedures
SPAM	Situation Present Assessment Method
TARDEC	US Army Tank Automotive Research Development and Engineering Center
TTPs	tactics, techniques, and procedures
TWLQ	Team Workload Questionnaire
UAV	unmanned aerial vehicle
WARNO	warning order
UAV-STE	Uninhabited Aerial Vehicle–Synthetic Task Environment

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 CCDC ARL
(PDF) FCDD RLD DCI
TECH LIB

1 CCDC ARL
(PDF) FCDD RLH B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

1 CCDC ARL
(PDF) FCDD HSI
J THOMAS
6662 GUNNER CIRCLE
ABERDEEN PROVING
GROUND MD
21005-5201

1 USAF 711 HPW
(PDF) 711 HPW/RH K GEISS
2698 G ST BLDG 190
WRIGHT PATTERSON AFB OH
45433-7604

1 USN ONR
(PDF) ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 USA NSRDEC
(PDF) RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 OSD OUSD ATL
(PDF) HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

6 ARIZONIA STATE
(PDF) UNIVERSITY
L HUANG
N COOKE
C JOHNSON
G LEMATTA
S BHATTI
M BARNES

ABERDEEN PROVING GROUND

10 CCDC ARL
(PDF) FCDD RLH
J LANE
Y CHEN
P FRANASZCZUK
A MARATHE
K MCDOWELL
K OIE
FCDD RLH F
J GASTON (A)
FCDD RLH FA
A DECOSTANZA
FCDD RLH FB
D BOOTHE (A)
FCDD RLH FC
K COX (A)