ARL-TR-9997 • OCT 2024

Soldier Performance in Sociotechnical Environments (Summary Technical Report, Oct 2020–Sep 2023)

by Heather Roy, Leah Enders, Russell Cohen-Hoffing, Jonathan Touryan, Thomas Rohaly, Bianca Dalangin, Angela Jeter, Jessica Villarreal, Alex Stauff, Adam Wachsman, William Gerichs, Robert Smith, Ashley Rabin, and Stephen Gordon

NOTICES

Disclaimers

ARL-TR-9997 ● OCT 2024

Soldier Performance in Sociotechnical Environments (Summary Technical Report, Oct 2020–Sep 2023)

Heather Roy, Leah Enders, Russell Cohen-Hoffing, and Jonathan Touryan
DEVCOM Army Research Laboratory

Thomas Rohaly, Bianca Dalangin, Angela Jeter, Jessica Villarreal, Alex Stauff, Adam Wachsman, William Gerichs, Robert Smith, Ashley Rabin, and Stephen Gordon
DCS Corporation

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED START DATE	END DATE
October 2024	Summary Technical Report	9/30/2020	9/29/2023

4. TITLE AND SUBTITLE
Soldier Performance in Sociotechnical Environments (Summary Technical Report, Oct 2020–Sep 2023)

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
CO2-02 (62184)		

6. AUTHOR(S)
Heather Roy, Leah Enders, Russell Cohen-Hoffing, Jonathan Touryan, Thomas Rohaly, Bianca Dalangin, Angela Jeter, Jessica Villarreal, Alex Stauff, Adam Wachsman, William Gerichs, Robert Smith, Ashley Rabin, and Stephen Gordon

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
DEVCOM Army Research Laboratory ATTN: FCDD-RLA-FC Aberdeen Proving Ground, MD 21005	ARL-TR-9997

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCIDs: Jonathan Touryan, 0000-0001-7343-7869; Russell Cohen Hoffing, 0000-0002-3478-8196

14. ABSTRACT
The U.S. Army Futures Command's Soldier Lethality Cross-Functional Team and U.S. Army Combat Capabilities Development Command Army Research Laboratory (ARL) initiated the research program Soldier Performance in Sociotechnical Environments in 2020 to increase Warfighter lethality and situational awareness (SA). The goal of this program was to conduct foundational research and develop technologies to quantify, predict, and enhance shared SA and situational understanding across volatile, uncertain, complex, and ambiguous operating environments. To achieve this goal, ARL developed the Tactical Awareness via Collective Knowledge (TACK) concept leveraging the science of opportunistic sensing to reason over the environment, enable command and control, and inform intelligent systems. Specifically, within this program, TACK applies opportunistic sensing to understand how Soldiers' behaviors, intent, cognitive state, and attention can be quantified in real time during small-unit dismount operations. TACK leverages next-generation dismount technologies to measure Soldiers as they are conducting simulated missions to make inferences about the environment and convey actionable information to a commander and/or autonomous agent, improving mission outcomes. This report details several TACK research thrusts and test beds falling under this effort. While the concepts and technologies developed under this program focused on dismounted operations, the science behind TACK is generalizable and can be applied across domains.

15. SUBJECT TERMS
Humans in Complex Systems, Summary Technical Report, opportunistic sensing, situational awareness, lethality, simulation test bed, physiology, eye tracking

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	64

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Jonathan Touryan	(410) 278-4329

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Executive Summary

The U.S. Army Futures Command's Soldier Lethality Cross-Functional Team (SL CFT) aims to meet the continuously evolving and unique Warfighter needs by taking tactical advantage of the advancement and development of emerging technologies and by narrowing capability gaps through the process of improving equipment, training, and resources.[1] The SL CFT focuses on providing capabilities to the Warfighters to increase lethality, mobility, protection, and situational awareness (SA), while countering threats or near-peer adversaries.[2] The Army's future force must be equipped to deploy and interact with advanced modern weaponry, systems, and technologies, at the command, squad, and individual Soldier levels across fast-paced and multi-domain battlefields.[1] To support these aims, in 2020 the SL CFT and U.S. Army Combat Capabilities Development Command Army Research Laboratory initiated the research program Soldier Performance in Sociotechnical Environments.

The goal of this program was to conduct foundational research and develop technologies to quantify, predict, and enhance shared SA and situational understanding across volatile, uncertain, complex, and ambiguous operating environments. To achieve this goal, the U.S. Army Combat Capabilities Development Command Army Research Laboratory developed the concept of Tactical Awareness via Collective Knowledge (TACK). TACK leverages the science of opportunistic sensing[3] to reason over the environment, enable command and control, and inform intelligent systems. Specifically, within this program, TACK applies opportunistic sensing to understand how Soldiers' behaviors, intent, cognitive state, and attention can be quantified in real time during small-unit dismount operations. Importantly, TACK extends beyond mere quantification to explore the best ways to make inferences about the environment and convey actionable information to a commander and/or autonomous agent to improve mission outcomes. Here, opportunistic sensing refers to the process of obtaining operational data required to infer mission context, train algorithms, and quantify unit effectiveness from tasks Soldiers naturally perform without increasing burden

---

[1] Thompson M. Soldier lethality team reimagines movement, vision and combat capabilities. Department of Army (US). 2022 Mar 15. https://www.army.mil/article/254732/soldier_lethality_team_reimagines_movement_vision_and_combat_capabilities.

[2] Purtiman R. Soldier lethality cross-functional team bringing next generation technologies to Soldiers. Department of Army (US). 2018 Sep 24. https://www.army.mil/article/211454/soldier_lethality_cross_functional_team_bringing_next_generation_technologies_to_soldiers.

[3] Lance BJ, Larkin GB, Touryan JO, Rexwinkle JT, Gutstein SM, Gordon SM, Toulson O, Choi J, Mahdi A, Hung CP, et al. Minimizing data requirements for soldier-interactive AI/ML applications through opportunistic sensing. Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II. 2020;11413:19–27. https://doi.org/10.1117/12.2564515.

or hindering their performance. To do this, we have leveraged current and next-generation dismount technologies to measure Soldiers as they are conducting missions. For instance, commanders can gain real-time information about how squads navigate and perceive their environment by exploiting gaze in combination with physiological data (e.g., electroencephalogram, electrodermal activity, electrocardiogram, and eye tracking) as individuals and teams are moving through real or simulated operational environments. Through the application of human state classification models, we can make inferences related to SA and cognitive state for both the individual and small unit. As one example, we have shown how gaze information and neural activity, combined across individuals, can be used to identify targets or mission-relevant objects in the environment. Furthermore, we can visualize this type of information to inform a commander or autonomous agent (e.g., unmanned aerial systems) to improve decision-making. While the concepts and technologies developed under this program focused on dismounted operations, the science behind TACK is generalizable and can be applied across domains, including a critical transition under the Next Generation Combat Vehicle CFT, where it has been applied in mounted operations.

1. Mission Description

This U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) project investigates technologies to quantify, predict, and enhance squad-level shared Soldiers' situational awareness (SA) and situational understanding across volatile, uncertain, complex, and ambiguous operating environments leading to demonstrated increases in mission effectiveness. The result of this effort will be systems that use real-time opportunistic measures of Soldier SA to adapt autonomous assets (routes, surveillance targets, etc.) based on the dynamic needs of the Soldier/squad due to the changing mission environment.

2. Research Goals

Obtaining and maintaining SA is a critical component to ensuring force protection and mission success.[1] Maintaining SA is everyone's responsibility; however, in dismount operations there is currently no unobtrusive means to assess and share SA during mission execution without an overt behavioral indicator such as explicit verbal communication or gestures. There is also no way to identify gaps in SA, or to use one Soldier's SA to augment the SA of another (Soldier or autonomous system). The goal of this program is to quantify and predict a Soldier's SA through opportunistic sensing using data from eye tracking, scene cameras, inertial measurement units (IMUs), and other wearable sensors. While these are not currently deployed in the close combat force (Infantry Brigade Combat Team), it is anticipated that these capabilities will become widely available through the future deployment of Augmented Reality (AR) technologies such as the Integrated Visual Augmentation System (IVAS). Use of eye-tracking and scene or weapon cameras to determine both where a Soldier is looking and what they are looking at make it possible to build implicitly labeled datasets of mission-critical objects for machine learning (ML) applications.[2] In addition, this information can be used to create a dynamic common operating picture (COP) for improved command-level decision making while minimizing network requirements. Here, we seek to understand (1) how these signals can be used to infer cognitive state and SA of the individual and (2) how they can be combined across individuals to provide dynamic mission context. We leverage decades of research on human perception and cognition in conjunction with state-of-the-art simulation environments to develop algorithms for continuous inference of shared SA.

3. Program Objectives

The program objectives are as follows:

- Demonstrate that real-time quantification and visualization of squad-level SA improves survivability and lethality during simulated missions.
- Demonstrate a significant increase in lethality (>25%), including measures such as react-to-contact, number of threats eliminated, and number of rounds on target.
- Demonstrate a significant increase in survivability (>25%), including measures such as mission time, number of hits taken, and duration of enemy engagements.
- Demonstrate a significant increase in battlefield awareness (>25%), including measures of area surveilled, threats detected, and combined visual coverage.

4. Current State-of-the-Art

Opportunistic sensing for cognitive state inference is a nascent field, only now beginning to be explored, and has never been used to quantify SA in operational settings. Currently, SA is maintained by the individual Soldier and manually coordinated across squad. Real-time information about Soldier or unit status is minimal or absent. Importantly, there are few if any human-in-the-loop simulation environments that combine the realism of dismount operations with the experimental control needed to explore the foundational questions around shared SA, in either human–human or human–autonomy teams. As such, it was critical to develop test beds with the capability of achieving the research goals and program objectives described above.

5. Major Partners

ARL teamed with researchers from DCS Corporation during this effort.

6. Research Efforts

This program included two foundational sciences studies. The first study (ARL 19-122) sought to isolate and quantify neuro-physiological markers of SA through a complex visual search task to understand how these markers are modulated by cognitive state (e.g., workload). This study leveraged electroencephalogram (EEG) and eye tracking during a visual search-and-navigation task to identify physiological markers of SA. Individuals from the general population and those with military experience (active duty and veterans) participated in this study. Participants were asked to navigate through a canyon-like virtual desktop environment while mentally counting an assigned target class (e.g., Humvee). Results demonstrated not only that neural activity and gaze information could be used to correctly identify target objects, but also that behavior significantly differed between civilian and active-duty populations. Specifically, using eye tracking, we found significantly increased fixations and dwell times on targets versus distractors, indicating participants were prioritizing this mission-relevant information in the environment.[3] In addition, the cognitive load induced through a secondary auditory math task significantly reduced fixations and dwell times during periods of high load.[3] As part of this study, demographic information on posttraumatic stress disorder (PTSD) was collected, providing a unique window into the effects of stress on performance, vital for understanding SA in operational environments but not typically accessible in laboratory settings. The PTSD sub-analysis revealed that those with PTSD had a greater number of fixations on the mission-relevant trail markers and that they had significantly longer individual fixation durations compared to those without PTSD.[4] These preliminary results showed that individuals with PTSD experience greater difficulty disengaging from stimuli, especially when meeting demands from multiple tasks (navigation, search, communication), thus affecting their ability to maintain SA. Importantly, these results suggest that under stress, even healthy (non-PTSD) individuals may exhibit a similar, quantifiable decrease in SA.

The follow-on second study (ARL 21-104) extended this foundational work to include an explicit stress manipulation to further identify behavioral and physiological markers of SA in operational environments. Given the promising findings related to PTSD, this experiment directly compared PTSD and non-PTSD populations, consisting of those with military experience (active duty and veterans). While it is often difficult to induce realistic stress levels within laboratory studies, including a PTSD population provided us a wider variation in response to stress and behavior in a military-relevant context. Furthermore, the current literature suggests that those with PTSD have altered processes related to SA, including hypervigilance, increased responsiveness to stimuli, memory deficits, and an attentional bias towards perceived threat. The primary goal of this study was to investigate how stress impacts behavior and physiology across both traditional laboratory and complex visual search tasks, while manipulating stress (high and low) within a given context (neutral and combat themes). During the primary visual search task, participants were asked to navigate four routes (conditions) manipulating stress and context through immersive virtual urban environments while mentally counting an assigned target class. Throughout the experiment, EEG, electrodermal activity (EDA), electrocardiogram, and eye tracking were collected. Data was collected from 41 Soldiers from a broad range of Military Occupational Specialties. Preliminary analysis demonstrated that specific biomarkers of arousal, established in the literature, can be used to quantify stress in complex real-world scenarios. These include saccade magnitude, gaze distance, and relative number of target fixations. Initial findings reported at the 23rd annual Vision Sciences Society conference indicated participants increased scanning behavior (scanned a larger area), spent longer fixating on objects, and were slower (decreased ground velocity) navigating the environment.[5] A detailed summary of these studies is provided in Appendix A.

7. Findings

Several test beds (human-in-the-loop simulation environments) were developed under this program to both explore the foundational research questions related to SA and quantify benefits realized by using the human as a sensor to understand operational environments, providing actionable information to commanders and autonomous systems. These test beds were developed within the Unreal Engine (UE) as a suite of plug-ins, collectively referred to as UETack, with the capability to simultaneously capture game-state, behavioral, and physiological data. UETack test beds described here integrate human participants (e.g., teams) with artificial intelligence (AI) capabilities, including sensing (cameras, auditory, and advanced computer vision), planning (course of action), and action (drone movement and/or directions to the team).

To explore and quantify team dynamics and the effect of sharing real-time information between the dismount team (e.g., fire team) and commander, we created the Tactical Awareness via Collective Knowledge (TACK) Commanders Interface (CI). The TACK CI both visualized opportunistically sensed information and provided a means for bidirectional communication by annotating the environment. This interface allowed a commander to drop waypoints to aid the fire team in navigating through an urban environment. The fire team was also able to mark threats (bombs) in the environment and send this information back to the commander. These annotations were enabled by means of a (simulated) Heads-Up Display (HUD) that overlaid spatially localized tactical information onto the world. The FY21 TACK CI test demonstrated the capability to complete mission objectives with up to 50% reduction in total mission time while neutralizing up to 30% more threats (i.e., bombs) when using the TACK CI, compared to without.

The FY22 TACK CI development focused on further integrating autonomy and demonstrating the effect of integrated autonomy on mission performance. Commanders were provided three unmanned aerial systems to deploy and scan the area ahead of the dismounted fire team. Commanders were tasked to dynamically navigate their team through a virtual urban environment by dropping waypoints and sharing critical threat information. The dismount team (human and unmanned aircraft systems [UAS]) was able to share opportunistically sensed information with the commander, in real time, that the commander could then use to update the team's directives. Dismounts could mark threats (e.g., enemy combatants and bombs), and the UAS could mark elements of cover (vehicles) and threats (enemy combatants). Importantly, the UAS incorporated uncertainty visualization methodologies developed under a related ARL effort.[6] Following the recommendations of that work, the drones would mark threats and elements of cover with corresponding color-coded ellipses (e.g., red for threat), where the size and orientation of the ellipse reflected the number of object detection events and thus conveyed the level of certainty related to object location. The FY22 TACK CI test demonstrated our novel approaches to combine the information obtained from the dismount team behaviors, and UAS assets enabled the dismounted fire team to increase lethality by 50% and increase battlespace awareness by 80% across a series of urban missions. This demonstration highlighted our opportunistic sensing concept: combining mission-relevant Soldier behaviors with computer vision algorithms running on autonomous systems to create a dynamic COP that allowed commanders to adjust the course of action based on the current threat picture. This approach can be applied across multiple domains of operation to enhance lethality, survivability, and adaptability of Soldier–autonomy teams.

After demonstrating the effectiveness of sharing information across the heterogenous team (human and autonomy), in FY23, our research focused on further reducing the cognitive burden on commanders. A TACK course of action (COA) route planning application and software tool was developed and integrated with the CI. This TACK CI COA application applied two common AI/ML techniques to identify and recommend routes based on distance and threat. While the FY22 CI allowed for dismount teams to send information back to the commander, which led to increases in the team's lethality and battlespace awareness through improved decision-making, the TACK CI COA automated this process. Recommended routes were displayed to a commander via a tactical map based on the mission scenario, and the route recommendations were updated in real time, incorporating incoming information from the dismount (human–autonomy) team. Information from the dismount team would be automatically displayed on the tactical map and integrated with the route planned to provide updated recommendations. Commanders could view enemy heat maps (i.e., danger zones) based on the information generated by the fire team and UAS. The shortest route, the safest route, and an alternate route (a combination of distance and safety optimization) were presented to the commander on the TACK COA and updated at regular intervals. A detailed summary of these technologies is provided in Appendix B.

Appendix C provides details on three additional use cases of UETack including a large-scale online study that leverages UETack to create a 3D experimental platform to understand how humans adapt to changing environments. This experimental platform will serve as the basis for a proof-of-concept for other ARL efforts that need to capture single or multiplayer data from human-in-the-loop simulations deployed on either local or cloud-based servers. Additionally, UETack was leveraged during external interactions through ARL's Strengthening Teamwork for Robust Operations in Novel Groups (STRONG) program at the STRONG Summer Summit and by STRONG Cycle 2 awardees at the University of Colorado and DCS Corp to understand how roles emerge within teams.

8. Recommendations

The Soldier Performance in Sociotechnical Environments program clearly demonstrated the utility of opportunistic sensing via the TACK concept. The results of the foundational research efforts demonstrated that SA and shared SA can be quantified via observation of behavior and physiological measures. This information can be used to make inferences about the environment and provide critical information about the mission context. As the future of warfare will include an increasing number of battlefield and wearable sensors, it is critical to operationalize this research by extending the algorithms and approaches developed under this program into real-world environments using next-generation Soldier systems. Ultimately, opportunistic sensing concepts such as TACK can be a means to advance human–machine integration in both mounted and dismounted contexts to achieve overmatch in the future operating environment.

Specifically, TACK demonstrated that both head and eye movements were key factors in the quantification of SA and identification of mission phase. These measures provide readily available data to enhance SA through technologies such as a CI (see Appendix B). Future HUD systems, such as IVAS, will be able to provide a subset of this information over existing tactical networks (e.g., Nett Warrior) using current visualization tools (e.g., Android Team Awareness Kit). While the TACK test beds had limited weapon movement due to the constraints of the simulation, relative weapon position and orientation were critical measures. Next-generation squad weapons will have IMUs for advanced scope systems, shot counters, or predictive maintenance. Continued science and technology (S&T) investments will be required to mature these technologies—especially passive eye tracking—and robustly integrate them across a tactical network. Likewise, inferential algorithms will need to be adapted and optimized to the limited power and compute resources available in dismount operations. However, integrating this information across the squad has the potential to improve SA and increase small unit lethality without adding hardware.

Beyond SA, research is needed to determine quantifiable metrics for Soldier readiness and engagement. Experimental test beds are necessary to understand how integrated technologies impact performance during realistic maneuvers (e.g., battle drills). Expanding to more immersive and mobile test bed environments will expand the variety of naturalistic behaviors and cognitive responses (e.g., movement, attention, stress). Through these environments, we can identify an ensemble of potential signals and sensors that would indicate level of readiness, directions of maneuver, and enemy engagement events. In turn, this information would reduce the time and resources required to deploy battlefield AI. Importantly, this will require S&T investments to understand how an ensemble of signals—from physiological (e.g., heart rate and gaze direction), biomechanical (e.g., gait and posture), and technological (e.g., weapon and armor) sources—can be integrated to enable meaningful inference. Likewise, research is needed to explore how these measures can be extracted from, and combined across, Soldier systems and autonomous technologies to further the development of human–machine integrated formations.

Finally, it is critical to develop standardized human-sensing data encapsulation and processing methodologies as part of the technology maturation process. A clear set of data requirements should be established, evaluated, and made available to researchers and transition partners, as currently, no such standard exists. This data standard should fill the gap between the human-performance and systems-development communities. It should be able to encompass time-series data from physiological monitoring devices (e.g., wearable sensors), mixed reality technologies (e.g., OpenXR HUDs), and robotics and autonomous systems (e.g., Robotic Operating System messages). It will be important to synchronize efforts across the S&T enterprise so that individual capabilities and technologies, as they emerge, can be integrated into the opportunistic sensing ecosystem for enhancing SA and mission performance.

9. Payoff / Army Impact

An overarching goal of this program, beyond building the foundational knowledge necessary to understand and exploit the key behavioral and physiological correlates of SA (see Appendix D, Publications), has been to create a modular, extensible, and multiplayer framework for human-in-the-loop simulations. Ideally, this framework would maintain laboratory-level precision, provide real-time access to data, and integrate both human and autonomous agents while establishing a standardized data schema. Work under this program has led to the creation of a Technology Readiness Level 4 (TRL 4) suite of software tools enabling opportunistic sensing research, providing the capability for real-time quantification and visualizations of shared SA to improve command-level decision-making and enhance human–machine integration. This software has been developed and refined throughout our experimentation and testing. Most recently, the UETack software transitioned in FY23 to the Crew Optimization and Augmentation Tech (COAT) S&T project, which is integrated into the Next Generation Combat Vehicle Cross-Functional Team (CFT). COAT develops flexible, adaptive interactions between Soldiers and unmanned assets, including autonomous mobility technologies and weapon engagement systems to enable robust manned–unmanned team dynamics. COAT, in conjunction with the Combat Vehicle Robotics (CoVeR) project, transition technologies to Program Executive Office Ground Combat Systems through yearly technology insertions. Under COAT, UETack has been used by the ARL Human Autonomy Teaming Essential Research Program to further annual experimentation in the Information for Mixed Squads (INFORMS) laboratory. Here, UETack has enabled rapid multimodal data collection, real-time queries, and the manipulation and export of data for postprocessing. UETack forms a critical component of the software backbone in the INFORMS laboratory and experimentation supporting two critical TRL 4 FY23 deliverables: (1) computer-vision-based adaptive aided target recognition that can be updated to recognize novel classes of targets and (2) After-Action Review software tools that can present detailed mission data back to platoon members to modify tactics, techniques, and procedures and future system behaviors.[*]

In addition to the UETack deliverable and publications, this program has provided over 30 high-level briefs and demonstrations for key military stakeholders. A central element of these briefs was to communicate the concept of opportunistic

---

[*] The UETack technical report is available via the Defense Technical Information Center's secure R&E Gateway (Nesta J, Malone D. Development of a high temperature superconducting magnetic field source for Army application. DEVCOM Army Research Laboratory (US); 2023 May. Report No.: ARL-TR-9694).

sensing and the effectiveness of leveraging the human as a sensor to infer mission context and understand the dynamic operational environment. These demonstrations also provided stakeholders with hands-on experience using the TACK CI (both desktop and AR implementations) allowing them an opportunity to provide feedback directly to the research and development team. Specifically, the TACK concept has presented to senior representatives from ARL; DEVCOM Analysis Center; Soldier Center; and Command, Control, Communication, Computers, Cyber, Intelligence, Surveillance and Reconnaissance Center (C5ISR); U.S. Army Engineer Research and Development Center; Special Operations Command; Medical Research and Development Command; the Naval Surface Warfare Center, in addition to the Soldier Lethality CFT board of directors. In addition, this program provided several educational engagement opportunities for faculty and cadets from the U.S. Military Academy at West Point. These stakeholder engagements and demonstrations not only increased the visibility of the program but also facilitated the development of collaborative relationships, including an international Cooperative Research and Development Agreement, and provided opportunities to identify transition partners for related DEVCOM efforts. Critically, the foundational research and opportunistic sensing capabilities developed here are now being leveraged by other ARL programs for both mounted and dismount applications.

10. References

1. Endsley MR, Holder LD, Leibrecht BC, Garland DJ, Wampler RL, Matthews MD. Modeling and measuring situation awareness in the infantry operational environment. U.S. Army Research Institute for the Behavioral and Social Sciences; 2000. Accession No.: ADA372709. https://apps.dtic.mil/sti/pdfs/ADA372709.pdf.

2. Lance BJ, Larkin GB, Touryan JO, Rexwinkle JT, Gutstein SM, Gordon SM, Toulson O, Choi J, Mahdi A, Hung CP, et al. Minimizing data requirements for Soldier-interactive AI/ML applications through opportunistic sensing. Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II. 2020;11413:19–27. https://doi.org/10.1117/12.2564515.

3. Enders LR, Smith RJ, Gordon SM, Ries AJ, Touryan J. Gaze behavior during navigation and visual search of an open-world virtual environment. Frontiers in Psychology. 2021;12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.681042.

4. Roy H, Enders LR, Rohaly T, Dalangin B, Villarreal J, Jeter A. Posttraumatic stress disorder and differences in eye gaze during a visual search task with cognitive load. Proceedings of the Annual Meeting of the Cognitive Science Society. 2022;44. https://escholarship.org/uc/item/2nj601jj.

5. Enders LR, Gordon SM, Roy H, Rohaly T, Dalangin B, Jeter A, Villarreal J, Boykin GL, Touryan J. Evidence of elevated situational awareness for active duty Soldiers during navigation of a virtual environment. PloS one. 2024;19(5).

6. Kusumastuti SA, Pollard KA, Oiknine AH, Dalangin B, Raber TR, Files BT. Practice improves performance of a 2D uncertainty integration task within and across visualizations. IEEE Transactions on Visualization and Computer Graphics. 2023;29(9):3949–3960. https://doi.org/10.1109/TVCG.2022.3173889.

Appendix A. Experimentation

To realize the Tactical Awareness via Collective Knowledge (TACK) concept, we sought to develop technologies for a squad-level situational awareness (SA) assessment while providing command-level decision support with communication and intervention capabilities for completing mission objectives. Towards this effort, we sought to build a strong foundation of scientific knowledge through experimentation to identify behavioral and physiological markers of SA and cognitive states that affect mission performance. To achieve this, we conducted two foundational experiments utilizing visual search paradigms. We seek to understand what insights we could gain from measures of brain activity and how this activity correlated with other physiological signals (e.g., eye-tracking, heart rate, and electrodermal activity) that are more accessible in the field (e.g., ARL Study No. 19-122 and 21-104). We advanced this further to specifically examine how differences in cognitive states, under stress, affect SA and mission performance (ARL Study No. 21-104). The following sections provide a detailed description of these studies.

A.1 Locations of Research and Soldier Touchpoint

TACK human subjects research was conducted at two locations. Studies involving recruiting civilian participants occurred at ARL-West (ARL-W) in Playa Vista, California. The ARL-W laboratory space resides in a regional office on the University of Southern California Institute for Creative Technologies campus on a secured floor. The lab space has a sound-controlled and visually isolated booth where participants completed their tasks.

A Soldier Touch Point (STP) facility was established at the ARL-Fort Sam Houston Field Element at the Joint Base San Antonio (JBSA) to recruit for studies requiring military personnel participants (active duty and veteran populations). STPs are immersive testing and feedback methods that enable Soldiers to offer insightful interpretations on how developing equipment/technologies are being used and/or useful for training or in field exercises. The JBSA lab space had a similar configuration to ARL-W and used identical recording equipment.

A.2 Human Interest Detection (ARL 19-122)

The aims of the Human Interest Detector (HID) experiment were to examine how various physiological markers (e.g., electroencephalogram [EEG], heart rate, and eye tracking) informed one's SA in a complex, dynamic, visual search task and how these markers were modulated by cognitive state. Below we describe the tasks used in HID and the primary outcomes of this study. Data was collected from the general population at ARL-W (California) (N = 45) and from military personnel at JBSA (Texas) (N = 24).

A.2.1 HID Tasks

A.2.1.1 Baseline

The baseline task was intended to provide baseline measures to visual stimuli. In the rapid serial visual presentation task, each trial consisted of an image of either a fountain or a castle, and participants were required to press the spacebar if the image was of a fountain. Each trial was displayed for 250 ms, with 1–4 s between each trial, for a total of 390 trials (312 castles, 78 fountains), equally divided into three blocks.

A.2.1.2 Search Task

Before completing the main search task, participants completed a training that familiarized them with the virtual environment and the controls. The HID virtual environment consisted of an open-world environment in which people navigated through a trail to look for and keep count of a target amongst distractor items (Fig. A-1). The target randomly assigned was a motorcycle, Humvee, furniture, or aircraft. Participants used W/A/S/D on the keyboard to move and the mouse to change perspective. Midway through the task, participants completed a math task while navigating the environment. For each set, three numbers were presented aurally to the participant, and the participant had to report aloud the sum of these three numbers to the experimenter for a total of three sets.

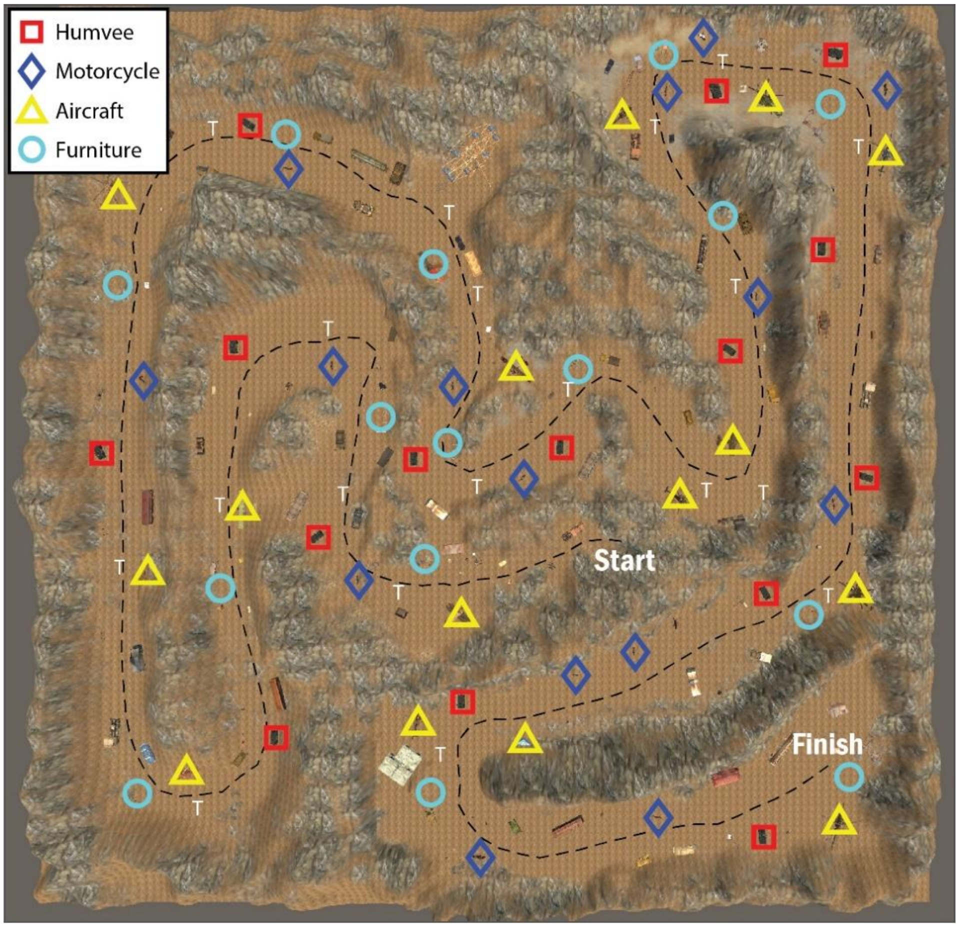

Fig. A-1 General layout of the virtual environment map.[1] The black checkered line represents an example subject's path from the starting area to the finish. Target icons are as follows: furniture (light blue circle), motorcycle (dark blue diamond), aircraft (yellow triangle), and Humvee (red square). Trail markers are present throughout the path and are indicated by a white "T."

A.2.1.4 Recall Task

The recall task was administered at the end of the experiment. Given an array of images containing a castle or a fountain, participants were asked to click on the images that they saw in the baseline task. There are 390 unique images that are spread over 19 pages, with 21 images for the first 18 pages and 12 images on the last page.

---

[1] Enders LR, Smith RJ, Gordon SM, Ries AJ, Touryan J. Gaze behavior during navigation and visual search of an open-world virtual environment. Frontiers in Psychology. 2021;12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.681042.

A.2.1.5 Questionnaires

Questionnaires measuring posttraumatic stress disorder (PTSD) (PCL-C/M), simulator sickness (SSQ[2]), and presence[3,4] were administered through an online platform on a tablet between tasks. Other questionnaires that measured motion sickness (MSSQ-Short[5]), game use (GUI; HRED protocol 17-017), and handedness[6] were administered prior to participants coming into the lab.

A.2.2 Outcomes

One aim of this study was to look at the effects of stimuli (target objects and distractor items) and cognitive load, from a secondary auditory math task, on gaze behavior in the open world search task. Using data from the ARL-W site, results showed that participants had an increased number of fixations and dwell time on targets compared to distractor items. When completing the secondary math task, participants slowed their speed, decreased gaze on objects, and increased the number of objects scanned in the environment. These findings support the use of complex virtual environments to study active visual search while maintaining high ecological validity. More details can be found in Enders et al.[1]

Virtual environments also provide dynamic visual scenes that can be used to understand how pupillary light response affects pupil size. Thurman et al.[7] looked at how visual patterns modulate pupil size, where luminance changes dynamically over time in an unconstrained 3D virtual environment. Using data from the ARL-W site, the hypothesized "blue sky effect" was found as a differential sensitivity of pupil responsiveness to blue light above the horizon.

Importantly, data collection at JBSA utilized service members, a population trained to maintain SA. Therefore, we sought to compare gaze behavior and brain activity (via EEG) between civilian (ARL-W) and service member populations (JBSA) and how these behaviors were influenced by target/distractor stimuli and the secondary

---

[2] Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. The International Journal of Aviation Psychology. 1993;3(3):203–220+. https://doi.org/10.1207/s15327108ijap0303_3.

[3] Usoh M, Catena E, Arman S, Slater M. Using presence questionnaires in reality. Presence. 2009;(5):497–503. https://doi.org/10.1162/105474600566989.

[4] Witmer BG, Jerome CJ, Singer MJ. The factor structure of the presence questionnaire. Presence: Teleoperators and Virtual Environments. 2005;14(3):298–312. https://doi.org/10.1162/105474605323384654.

[5] Golding JF. Motion sickness susceptibility. Autonomic Neuroscience. 2006;129(1):67–76. https://doi.org/10.1016/j.autneu.2006.07.019.

[6] Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971;9(1):97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

[7] Thurman SM, Cohen Hoffing RA, Madison A, Ries AJ, Gordon SM, Touryan J. Blue sky effect: contextual influences on pupil size during naturalistic visual search. Frontiers in Psychology. 2021;12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.748539.

math task. Overall, the active duty group looked at more target objects and scanned the environment faster, with larger saccade magnitudes compared to the civilian group. We examined task-relevant brain activity with convolutional neural networks[8] and found that the active duty group reduced cognitive re-processing of objects compared to the civilians. The civilian group exhibited a decrease in dwell time on each viewed object when cognitive load increased; however, there was no significant change in dwell time for the active duty group. The active duty group performed better on the primary task (reported closer to the correct number of targets total in environment) but performed worse (lower score) in the secondary math task compared to civilians. This evidence suggests that the active duty group exhibited more elements of SA in the primary task but at the cost of decreased performance in the secondary math task.[9]

In efforts to increase data analysis efficacy and foster collaboration to create more research products, a data guide was published to summarize the task and data format.[10] This data guide outlines how, in the work above, we were able to efficiently integrate EEG and eye-tracking with our virtual environment and includes a detailed description of the data.

A.3 Maladaptive Stress Response Testing (ARL 21-104)

This research program is particularly interested in improving SA for enhanced mission performance and identifying the behavioral and physiological markers associated with the perception of naturalistic stimuli and how perception may be affected by different features of the operating environment, such as stress. It is often difficult to induce realistic and natural levels of stress, especially when using military personnel, in standard laboratory tasks. However, a key feature of the Maladaptive Stress Response Testing (MAST) study is that it leverages a PTSD population to better test SA under higher stress states. Those with PTSD have altered processes that underly the acquisition and maintenance of SA, which is a critical focus of TACK. Individuals with PTSD are at a baseline heightened state of arousal, are hypervigilant, highly responsive to stimuli, experience memory

[8] Solon AJ, Lawhern VJ, Touryan J, McDaniel JR, Ries AJ, Gordon SM. Decoding P300 variability using convolutional neural networks. Frontiers in Human Neuroscience. 2019;13:201. https://doi.org/10.3389/fnhum.2019.00201.

[9] Enders LR, Gordon SM, Roy H, Rohaly T, Dalangin B, Jeter A, Villarreal J, Boykin GL, Touryan J. Evidence of elevated situational awareness for active duty Soldiers during navigation of a virtual environment. PloS one. 2024;19(5).

[10] Dalangin B, Oiknine A, Villarreal J, Jeter A, Gordon S, Hoffing RA, Thurman S, Smith R, Enders L, Roy H, et al. Using eye tracking and electroencephalography (EEG) for visual search experimentation in virtual environments: a data guide. DEVCOM Army Research Laboratory (US); 2020. Report No.: ARL-TR-9045. Accession No.: AD1109153. https://apps.dtic.mil/sti/citations/AD1109153.

deficits, and exhibit an attentional bias towards threat. By including this population in our study, along with a non-PTSD comparative sample, we were able to obtain a wider variation in task-induced stress response in military-relevant contexts. Accurate estimation of SA using physiology indicators, including variability in SA, is critical for ensuring our algorithms are robust across individuals and provides us with a better understanding of the effects of stress on performance. Furthermore, this is a population critical to our problem space in the Army. Stress, developing stress disorders, and PTSD affect Soldiers not only when they return home, but also when performing tasks in the field.

We designed MAST to specifically explore the effects of stimuli saliency and context (using neutral or combat-related cues) on arousal and SA performance within and across tasks. The MAST study sought to identify behavioral and physiological markers of SA and their dependence on cognitive state across traditional laboratory tasks and in a novel virtual urban environment using a visual search paradigm. Physiological measures including eye-tracking, heart rate, electrodermal activity, and EEG were collected across all tasks within this study. This study investigated the effects of stimuli saliency and context on arousal and performance during the visual search task.

A.3.1 Tasks

The MAST study is unique in how it leveraged a PTSD population, not just to understand and differentiate persons affected by PTSD, but to understand the heightened stress state itself. We have thus included several traditional laboratory tasks (emotional Stroop task [EST] and an emotional interference task [EIT]) to establish a baseline so we can compare our more realistic paradigm to the visual search and PTSD literature. Here we will focus on the visual search task and preliminary results for that paradigm.

A.3.1.1 Visual Search Task

The primary task within the MAST study consisted of a novel visual search task within an immersive desktop virtual environment (VE). The task was created using the Unreal Environment TACK (UETack), custom plugins for the Unreal Engine, which provides a flexible platform for designing experiments within a VE. Through lab streaming layer, we were able to synchronize data streaming across multiple behavioral and physiological data sources, making this ideal for human sensing experiments. We carefully designed our visual search task to specifically select visual and auditory stimuli to modulate stress for a PTSD population in a safe testing environment. This provided the opportunity to test stress responses within a PTSD population in a more naturalistic and complex dynamic environment that more closely mirrored that of the real world (while maintaining experimental control) as well as the opportunity to compare stress responses within the VE to more traditional lab-based tasks included in this study.

Participants were required to complete four conditions within this task that involved navigating four separate routes through the MAST environment while visually scanning for and mentally counting an assigned target class. Conditions varied by context (neutral vs. combat) and by stress level (high vs. low). The same context conditions were completed together, with the high stress condition always preceding the low stress condition, allowing us to explore recovery effects. Targets varied by context such that within the neutral conditions participants were tasked with locating a yellow pick-up truck, and within the combat conditions, a green sport utility vehicle (Fig. A-2). The targets were chosen to be roughly equivalent in size and saliency to blend in with the other environmental stimuli included in the levels (e.g., yellow construction vehicle in the neutral conditions).

Fig. A-2 Visual search construction (neutral) and combat targets

Participants were allowed to freely explore the environment using mouse and keyboard controls while generally moving between two waypoints, guided by arrows present in the environment. Each condition took approximately 5 min to navigate. Participants were asked to click with the mouse to indicate when they saw a target and were prompted to enter the total number of targets at the end of each condition. Additionally, after completing each condition (neutral vs. combat), participants responded to the same five-point Likert questionnaire used in the EIT task to indicate their perceived level of stress. Following each context block, participants completed three memory and recall tasks. Memory tasks assessed recall and recognition of objects that were placed throughout the levels.

A.3.1.4 Memory and Recall Tasks

After each condition block (blocked by context—neutral or combat) of the visual search task, participants were asked to complete a free-recall task. Participants were allotted up to 3 min to list as many objects as they recalled seeing in the environment. The second task was a recognition task. Participants were simply presented a series of images and asked yes or no as to whether an object was present in the environment. The target objects included in this task were sampled from the unique objects placed within the conditions. The final memory task included in this study was a challenging spatial awareness task. This task required participants to imagine they were standing on a given object facing another object in the environment and then indicate the directional heading towards a third object (Fig. A-3). The objects used within this task again were the unique (memorable) objects embedded in each condition.

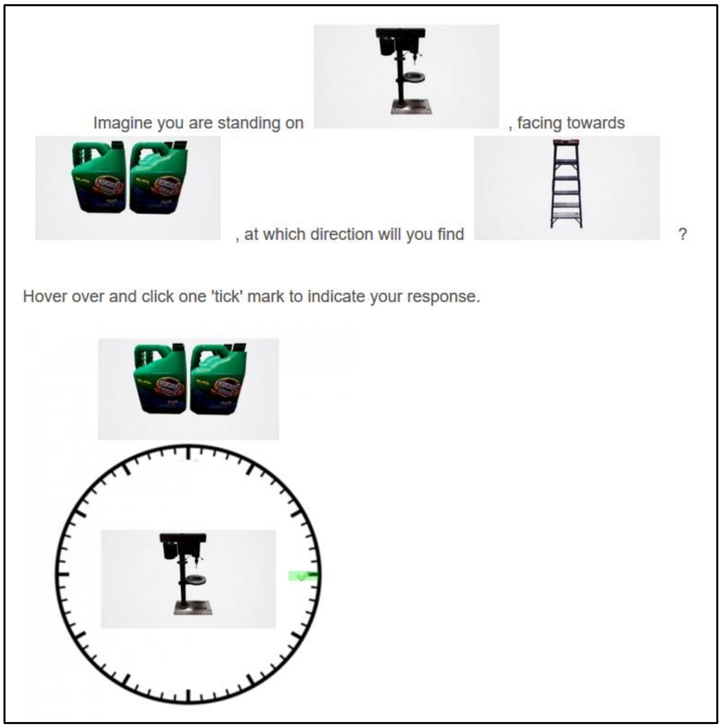

Fig. A-3 Spatial survey knowledge task

A.3.1.5 Questionnaires

In addition to completing the experimental tasks, participants were asked to complete several questionnaires providing responses related to their demographics, combat exposure (Combat Exposure Scale), personality (Big Five Inventory), presence (Ingroup-14 Presence Questionnaire), comorbid diagnoses (Adult Attention Deficit Hyperactivity Disorder Self-Report Scale), sleep (Pittsburgh Sleep Quality Index), and mild Traumatic Brian Injury (Ohio State University TBI

Identification Method). The presence of PTSD symptoms was assessed using the PTSD Checklist for DSM-IV Military.

A.3.2 Preliminary Outcomes

Preliminary analyses have focused on the effects of stress on visual search within the combat conditions, as observed through changes in eye movements. Initial results indicated that participants in the high stress environment scanned a larger area (increased saccade magnitudes), increased fixation durations, and decreased ground velocity when navigating the environment.[11] Furthermore, initial findings indicate that those with PTSD have more horizontal saccades and engage in less physical exploration of the environment (reduced path length and total mission time). These preliminary results demonstrate that innate environmental stress and a baseline heightened state of arousal impact SA. In addition to our internal research team, we have established analyses partners at Duke University and initiated a Cooperative Research and Development Agreement with the MyndBlue company in France. Additional analyses from these collaborative partners are forthcoming.

[11] Enders L, Roy H, Rohaly T, Jeter A, Villarreal J. Impacts of posttraumatic stress disorder on eye-movement during visual search in an open virtual environment under high and low stress conditions. Journal of Vision. 2023;23(9):5691. https://doi.org/10.1167/jov.23.9.5691.

Appendix B. Commanders Interface Test

The dynamics and complexities of future operations require military teams to rapidly adapt to the increased pace of technological change to process the overabundance of data. To address this challenge, we developed methodologies for autonomous systems to adapt and respond to uncertain measures of Soldier and squad state and situational awareness (SA) to enable faster human decision making and thus maximize positive mission outcomes. Specifically, we developed and demonstrated the Tactical Awareness via Collective Knowledge (TACK) XR software suite to the Soldier Lethality Cross-Functional Team (SL CFT) and related components at a Technology Readiness Level 3 (TRL 3). These software systems are key technologies that enable opportunistic sensing and the implementation of real-time quantification and visualization of shared SA to improve command-level decision making. Of these, a critical capability, showcased to the SL CFT, was the TACK Commanders Interface (CI). The TACK CI integrates behavioral and physiological data in real time to display Soldier- and squad-level information to the commander. Critically, the TACK CI interface enables information from across the human–agent team to be shared with and visualized by the commander. Using the TACK CI, a commander can receive information from the human and agent team members, monitor the team's state and progress, direct the team by dropping virtual waypoints along their route, view targets, mark confirmed targets, and propagate information back to the team. A live demonstration of the TACK CI was provided to the SL CFT along with results from the annual tests conducted in both FY21 and FY22. These annual evaluations were a key deliverable to the SL CFT; the sections below describe the scenarios and key results for the FY21 and FY22 TACK CI tests.

B.1 FY21 Scenario and Results

The purpose of the FY21 TACK CI validation test was to demonstrate to the SL CFT that the TACK CI was functional, robust, and had the capability to enhance mission speed and improve SA as designed. Specifically, the FY21 TACK CI Test (ARL Study No. 21-087) explored the effects of using the CI on mission performance across four conditions, varying the use of the CI (with and without CI), mission priority (safety vs. speed), and map. In the speed priority condition, the objective was to get to the extraction location as fast as possible, and in the safety priority condition, the objective was to get to the extraction location as safely as possible. Two maps were used to reduce learning effects, and conditions were counterbalanced across participants. For consistency, the commander's role was played by a team member. Participants were recruited for the dismount team roles. Four teams of two participated, and teams completed all conditions. Teams were tasked with navigating the virtual urban environment from a starting point to an end point, both provided to them on a map. Participants had to work together to navigate the environment to the end point while identifying and disarming bombs along their route (Fig. B-1). In the with-CI condition, the commander could drop waypoints (blue spheres) into the environment to direct the team along a route (Fig. B-1). As this test was conducted during the early stages of the pandemic, participants were granted access to the TACK virtual environment online. Participants primarily completed the task from their homes using their desktop or laptop computers. Participants were able to speak to each other and the commander via a Microsoft Teams meeting.

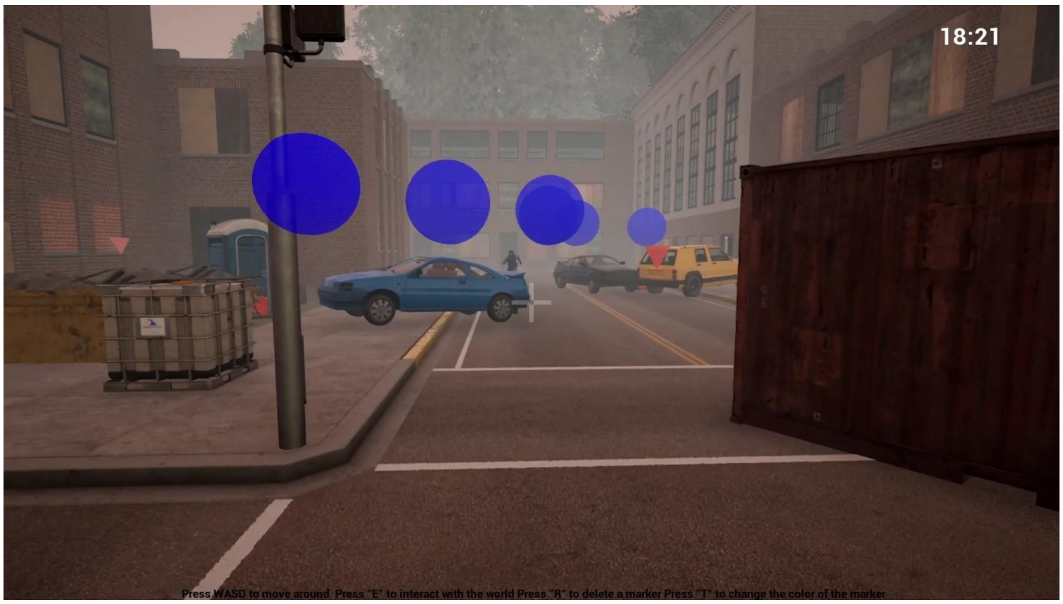

Fig. B-1 FY21 TACK CI test showing commander dropped waypoints (blue spheres) and bomb (red arrow) for defusal by dismount

The results of this test were solely intended to demonstrate that TACK had achieved the FY21 Stage Gate. Thus, results from the initial FY21 TACK CI test focused on providing summary outcomes as a proof-of-concept for the TACK CI. Analyses focused on comparing performance between with- and without-CI conditions to demonstrate whether the CI operated as intended and provided benefit towards reducing mission time and/or the number of bombs successfully defused. Comparing runs with and without CI demonstrated that teams using the CI were able to complete mission objectives with up to 50% reduced mission times while neutralizing up to 30% more threats (bombs). In the with-CI conditions, teams were faster, took a shorter path with less retread, missed fewer targets, defused more bombs, exploded fewer bombs, and all made it to the extraction location within the allotted time (20 min) for mission runs. We also compared mission priorities (speed vs. safety conditions). When prioritizing speed, with-CI teams were faster, traveled the shortest distance, and were more likely to reach the extraction point. When prioritizing speed but operating in the without-CI condition, teams traveled the longest path, missed the most targets, and defused the fewest bombs. When prioritizing safety in the with-CI condition, teams defused the most bombs. Data from this study supported the continued development of the TACK CI and informed the design and data collection of our next annual test, which focused on integrating autonomy and incorporating physiology metrics directly into the test.

B.2 FY22 Scenario

The purpose of the FY22 TACK CI validation test was to demonstrate to the SL CFT that the TACK CI was functional, robust, and capable of incorporating information from autonomous agents and improving SA as designed. The FY22 CI test compared the TACK CI with and without added autonomy. The focus of this test was to demonstrate how the added information from the autonomous drones, communicated through the CI, affected mission performance for teams. This test focused on the role of the commander. Participants were recruited from project collaborators to fill the commander role. The test also included a fire team consisting of four TACK team members. The individuals playing the fire team were kept consistent throughout the tests and maintained the same role. The fire team followed roles and basic movement techniques as outlined in TACK CI standard operating procedures. Participants recruited to fill the role of the commander were asked to lead the fire team through the virtual environment using the TACK CI interface to drop waypoints, mark threats, and view information from the team.

Teams completed two runs with each commander, one with autonomy and one without. These runs were counterbalanced across the commanders. For conditions with autonomy, commanders had three autonomous drone assets, along with their human fire team. Drones were able to scan, via computer vision, the environment for threats (enemies) and cover (vehicles) and mark the location of these items by dropping a color-coded ellipse (e.g., red for threats) onto the CI (Fig. B-2). Enemies were distributed through the environment on the roofs and at ground level. The drones were able to identify threats (both bombs and enemies) on the ground but were limited by height as to whether they could detect threats on the rooftop (e.g., threats on taller buildings were not able to be detected). The size of the ellipse provided by the drones communicated the level of certainty in terms of the number of detection events on a threat or cover object in the environment. Enemies were static agents that could fire on participants and receive fire from participants. Participants were tasked with the commander role to navigate dismount teams through the TACK virtual environment to the determined extraction point while eliminating any threats they encountered (i.e., enemies) and disarm any bombs along their route. Conditions were limited to 20 min. Eleven participants were recruited as commanders in this test.

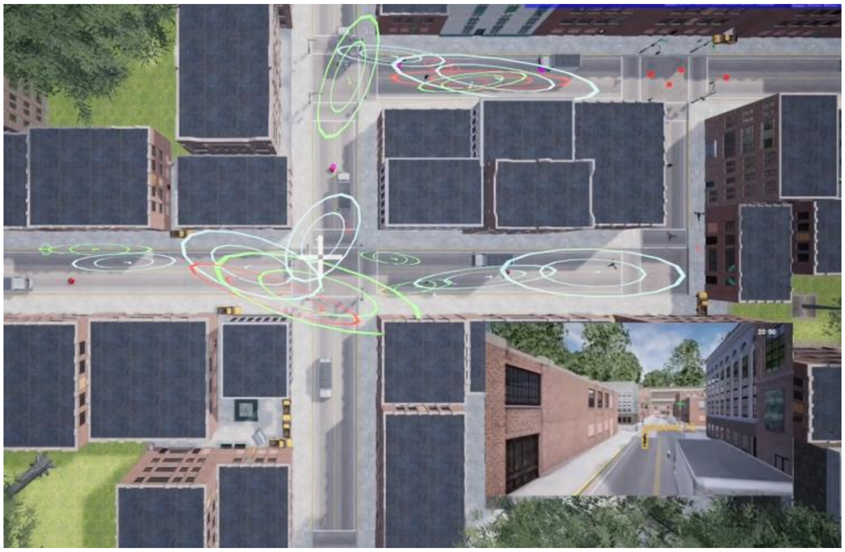

Fig. B-2 FY22 TACK CI showing overlays communicating the level of certainty and inset drone view

B.3 FY22 Autonomous Agents

The drones used in the FY22 TACK CI test used the Unreal Engine's Nav mesh for pathfinding and moving to their destination. The other drone functionality was done in Python through the pixel streaming interface. The drones are aware of all player and drone position/orientation in the environment. Each drone's destination calculation occurred every 300 frames (~5 s). The destination location is determined by translating 9000 Unreal Engine (UE) units (90 m) from the average location of the players (center of mass) in the average direction of the players (average azimuth). The drones arrange themselves in a circle with a radius of 2000 UE units, separated by 120°, around the current destination point with the first drone in the same direction as the players' average direction. The drones run the "You Only Look Once" (YOLOv3) algorithm on their forward-facing camera at approximately 6 Hz and place markers at their location and direction whenever a predefined object class is detected. The markers are grouped together in-game and displayed as ellipses with different colors representing different object classes indicating elements of threat (e.g., enemy dismounts) and cover (e.g., vehicles) based on work by ARL collaborators.[1]

[1] Kusumastuti SA, Pollard KA, Oiknine AH, Dalangin B, Raber TR, Files BT. Practice improves performance of a 2D uncertainty integration task within and across visualizations. IEEE Transactions on Visualization and Computer Graphics. 2023;29(9):3949–3960. https://doi.org/10.1109/TVCG.2022.3173889.

B.4 FY22 Behavioral Analyses

The FY22 TACK CI test highlighted our approach to fuse information passively obtained from mission-relevant Soldier behaviors with computer vision algorithms running on autonomous systems to create a dynamic Common Operating Picture that allowed commanders to adjust the course of action based on the current threat picture. Analyses of the FY22 TACK CI test largely focused on comparing performance between the CI with and without autonomy to demonstrate whether the CI operated as intended and whether incorporating autonomy input provided any benefit towards increasing the speed and effectiveness (e.g., lethality and survivability) to complete the mission and/or the number of bombs successfully defused. Results from the FY22 TACK CI test demonstrated a novel approach to combine information passively obtained from Soldier behaviors and computer vision algorithms enabling dismounted squads to increase lethality by 50% and increase battlespace awareness by 84% across a series of simulated urban missions within a desktop virtual environment. The following section goes over the behavioral analyses conducted.

A series of Paired Samples T-tests were conducted comparing behavioral metrics and performance between runs with and without autonomous drone assets. Team performance was examined across categories such as shot behavior by the team, shot behavior by the enemies, firing bouts, bomb status, mission time, and SA coverage by the team. Shot behaviors by the team examined in this analysis included the total number of shots by the team, total number of shots directed at hostiles by the team, total shots to friendlies by the team, total number of shots sustained by the team, total squad shots, and total hits by the team. Teams with drone assets on average shot more as a team ($M = 421.00$, $SD = 120.43$) and within individual roles as opposed to without drones for the team ($M = 267.73$, $SD = 35.34$). With a mean increase of 153.27, 95% CI [21.63, 284.92], drones elicited a statistically significant increase in shots taken by the team, $t(10) = 2.59$, $p = 0.027$. Using the drone assets thus resulted in a 57.24% increase in the total number of shots taken by teams. Teams with drone assets on average also shot more hostiles as a team ($M = 42.00$, $SD = 9.47$). With a mean increase of 11.72, 95% CI [2.28, 21.17], drones elicited a statistically significant increase in shots taken at hostiles by the team, $t(10) = 2.77$, $p = 0.020$, resulting in a 38.72% increase in hostiles shot. Relatedly, this resulted in a 50% increase in enemy deaths. Teams with drone assets on average neutralized more enemies ($M = 30.00$, $SD = 6.00$) compared to without drone assets ($M = 20.00$, $SD = 5.25$). With a mean increase of 10.00, 95% CI [3.49, 16.51], drones elicited a statistically significant increase in the number of enemies neutralized by the team, $t(10) = 3.42$, $p = 0.007$.

Enemy behavior was also examined between runs where teams were and were not using autonomous aids. Interestingly, for missions with drone assets, there was a greater total number of enemy shots taken (M = 172.09, SD = 67.99) compared to without drone assets (M = 105.27, SD = 46.57). With a mean increase of 66.82, 95% CI [10.23, 123.41], drones elicited a statistically significant increase in the number of enemy shots, $t(10) = 2.63$, $p = 0.025$. This resulted in a 63.46% increase in the total number of shots taken by the static enemy forces distributed in the environment. However, while teams with drone assets on average experienced a greater number of total hits from enemy shots (M = 54.18, SD = 22.46) compared to without drone assets (M = 40.73, SD = 23.78), there was no significant difference between the two conditions.

Enemy engagements were also examined in terms of the number of threat interactions and the duration of those interactions. Threat interactions were defined as the sum of all the distinct enemy actors and bombs that the team interacted with. Specifically, this metric referred to the sum of the exploded and defused bombs plus the sum of the distinct enemy actors who fired at least one shot during the run. Teams with drone assets on average experienced a greater number of total threat interactions (M = 33.91, SD = 6.46) compared to without drone assets (M = 24.45, SD = 5.32). With a mean increase of 9.45, 95% CI [2.85, 16.06], drones elicited a statistically significant increase in the number of threats interacted with by the team, $t(10) = 3.19$, $p = 0.010$. This resulted in a 38.65% increase in threat interactions for teams using drone assets. When examined alone, bomb status (i.e., exploded, defused) was not significant between conditions. However, a 39.76% increase in the number of enemy-initiated bouts did occur in teams with drone assets. Teams with drone assets on average experienced a greater number of enemy-initiated bouts (M = 9.27, SD = 2.20) compared to without drone assets (M = 6.64, SD = 2.98). With a mean increase of 2.64, 95% CI [0.42, 4.85], drones elicited a statistically significant increase in the number of enemy initiated bouts experienced by the team, $t(10) = 2.65$, $p = 0.024$.

We also examined total mission times and coverage between runs with and without autonomous drone assets. Teams with drone assets on average experienced longer mission times in seconds (M = 508.44, SD = 118.67) compared to without drone assets (M = 356.20, SD = 42.81). With a mean increase of 152.24, 95% CI [1.68, 52.87], drones elicited a statistically significant increase in shots by rooftop enemies, $t(10) = –4.73$, $p < 0.01$. This resulted in a 42.74% increase in mission times for teams using drone assets. Furthermore, teams with drone assets on average experienced greater dismount coverage (M = 21027.27, SD = 2599.40) compared to without drone assets (M = 18921.43, SD = 1824.54). With a mean increase of 2105.84, 95% CI [48.16, 4163.52], drones elicited a statistically significant increase in dismount coverage by the team, $t(10) = 2.28$, $p = 0.046$, resulting in an 11.13% increase in dismount coverage. Critically, for teams with drone assets, total coverage increased by 83.97%.

An interesting byproduct of our scenario was a trade-off between the drones being able to identify the ground level versus rooftop enemies. Because of limitations within the simulation for the drone, the drones were only able to reliably identify either the ground or rooftop enemies. We elected to prioritize the ground enemies. This, however, did introduce an opportunity to examine real-world factors such as the concept of more dynamic factors, including increased difficulty in tracking or predicting threats and system limitations (Table B-1). While teams using drones generally experienced a mean increase of 10.00, 95% CI [1.93, 18.07] and a statistically significant increase in total enemies firing, $t(10) = 2.76$, $p = 0.02$, it appears this was largely due to the increase in engagements with rooftop enemies. This is likely because commanders would receive threat information from the drones primarily related to the ground-level threats and would thus lead their teams away from these threats but inadvertently towards areas that may have been populated with more rooftop threats. Teams with drone assets appear to have engaged in a statistically significant increase in the number of rooftop ambushes ($t(10) = 3.94$, $p = 0.003$) with a mean increase of 3.45, 95% CI [1.50, 5.41], the number of rooftop enemies firing at the team ($t(10) = 3.33$, $p = 0.008$) with a mean increase of 4.45, 95% CI [1.47, 7.43], and the number of shots by rooftop enemies directed at the team ($t(10) = 2.374$, $p = 0.039$) with a mean increase of 95% CI [1.68, 52.87].

Table B-1 Rooftop enemy engagements

Rooftop enemy metrics	With drones	Without drones
Total enemies firing	(M = 25.64, SD = 6.82)	(M = 15.64, SD = 6.14)
Number of rooftop ambushes	(M = 8.45, SD = 1.57)	(M = 5.00, SD = 2.45)
Total rooftop enemies firing	(M = 11.91, SD = 2.21)	(M = 7.45, SD = 3.56)
Total shots by rooftop enemies	(M = 89.45, SD = 36.86)	(M = 62.18, SD = 26.48)

However, despite these system limitations, it is important to recall that teams with drones were able to successfully eliminate more threats in the environment. The following section provides further insights into the team's performance and gaze patterns by leveraging the opportunistically sensed physiological (eye gaze) patterns collected during the missions.

B.5 FY22 Eye Tracking Analyses

B.5.1 Statistics

A series of paired t-tests and mixed-model Analyses of Variance (ANOVAs) determined how the presence of drones (drone off/drone on) and firing condition varied (under fire/not under fire) impacted gaze patterns of the dismount fire team. Data was investigated for outliers, and data points that were greater than 3 standard deviations from the mean were removed. A correlation matrix determined if repetition of performing the navigation (two different environments, randomized) impacted variable outcomes. For this reason, order was included in the statistical models regarding variables of delay in squad returning firing, saccade angle, total number of distractor objects fixated while not under fire, and distance when fixating on bombs. Data transformations were applied to data that violated assumptions of normality or homogeneity of variance. Accordingly, a log transformation was applied to the fixation rate, individual duration of fixations, mean number of fixations, and mean dwell time data on both distractors and bombs. Specific fixation information on distractors (anything in the environment not labeled a bomb or enemy), bombs, and enemies was analyzed separately due to large variations between how many objects were present and focus shifts when under hostile fire. Enemies were often hard to see in the environment, and therefore on several runs, there was limited information on individual fixation outcomes with regard to enemy targets, especially rooftop enemies. Most information regarding enemy targets in the environment was limited to only one to three fixations.

B.5.2 Results

B.5.2.1 Movement through the Environment

Overall, the presence of the drones did not significantly impact position velocity (ground velocity) but did significantly increase the average total distance covered by the individuals on the ground (Fig. B-3). Specifically, it appeared that individuals on the ground tended to cover more distance while under gunfire, but the total distance traveled was not statistically significant ($p = 0.056$) when broken down into firing condition (under gun fire/not under gun fire). They also significantly increased the time to complete the navigation task with the drones present overall, but time under different firing conditions was not significantly different.

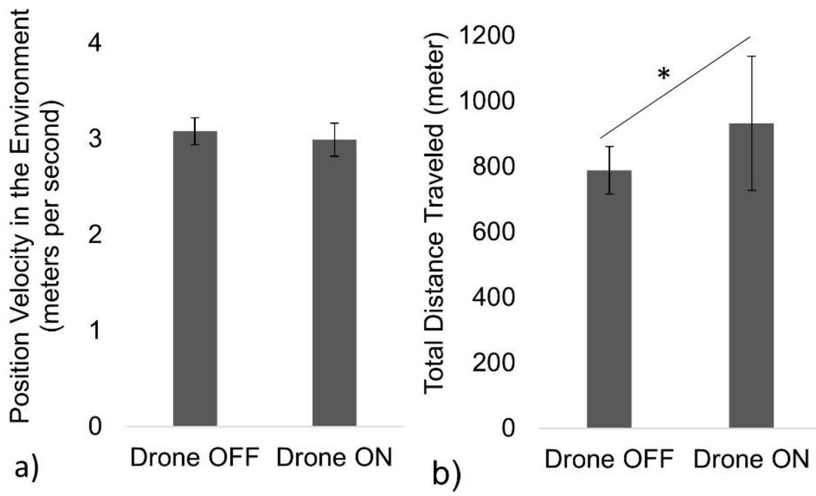

Fig. B-3 Position velocity was not significantly impacted by drone presence (a). The total distance traveled by the squad did significantly increase with drones present (b).

B.5.2.2 Enemy Encounters and Shots Fired

A repeated measures ANOVA (with order as a covariate) determined that there was no difference in the delay in squad returning fire with or without the presence of drones (Fig. B-4a, $F(1, 9) = 4.19$, $p = 0.071$). A paired t-test showed that there was no significant difference between the total (number of) enemies firing shots on the squad with or without the presence of drones (Fig. B-4b, $t = -2.76$, $p = 0.020$). A related-samples Wilcoxon signed-rank test determined that with the drones present, there was a significant increase in total successful hits on enemies by the squad with the drones present (Fig. B-4c, $Z = 2.19$, $p = 0.028$). Two separate paired t-tests showed no significant difference in the total successful hits by enemies on squad (Fig. B-4d, $t = -1.85$, $p = 0.094$) or the total number of successful hits by rooftop enemies on the squad (Fig. B-4e, $t = -1.77$, $p = 0.107$).

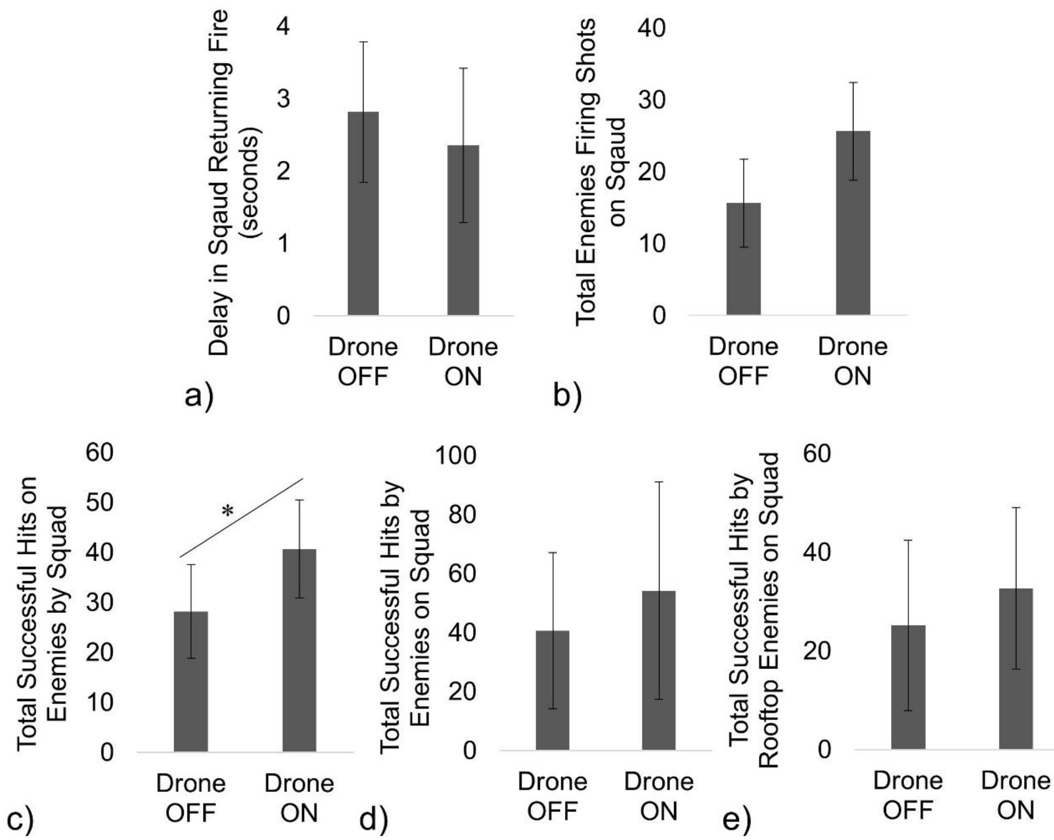

Fig. B-4 Drone presence on the squad did not significantly impact the delay in squad returning fire (a), the total enemies firing shots on the squad (b), total successful hits by enemies on squad (d), or the total successful hits by rooftop enemies on squad (e). Total successful hits on enemies by squad significantly increased with the drones (c).

B.5.2.3 Overall Gaze Pattern Statistics

Six separate mixed-model ANOVAs determined how the addition of drones impacted the squads' gaze pattern when moving under gunfire and without gunfire and disarming bombs while navigating the CI environment. The presence of drones (drone off/ drone on) providing positional information significantly decreased the object rate (the number of unique objects fixated on per second) (Fig. B-5c, $F(1, 20) = 9.94$, $p = 0.005$), saccade velocity (Fig. B-5e, $F(1, 20) = 5.15$, $p = 0.034$), and saccade magnitude (Fig. B-5f, $F(1, 20) = 6.52$, $p = 0.019$), on average from on the dismount teammates regardless of firing condition (under fire/ not under fire). Regardless of drone presence, individuals on the ground significantly decreased fixation rate (Fig. B-5a, $F(1, 20) = 7.16$, $p = 0.015$), saccade rate (Fig. B-5d, $F(1, 20) = 23.63$, $p < 0.000$), and saccade magnitude (Fig. B-5f, $F(1, 20) = 18.03$, $p < 0.000$), and individuals increased individual fixation duration (Fig. B-5b, $F(1, 20) = 57.48$, $p < 0.000$) when navigating the environment under fire compared to when navigating not under fire. Neither drone presence nor firing condition significantly impacted saccade angle, blink rate, pupil size, or position velocity (ground velocity of moving through the environment), and all other main effects and interactions were not significant ($p > 0.05$). Additionally, two separate paired t-tests showed that the presence of the drones did not impact total distance traveled for individuals regardless of if they were under fire or not under fire.

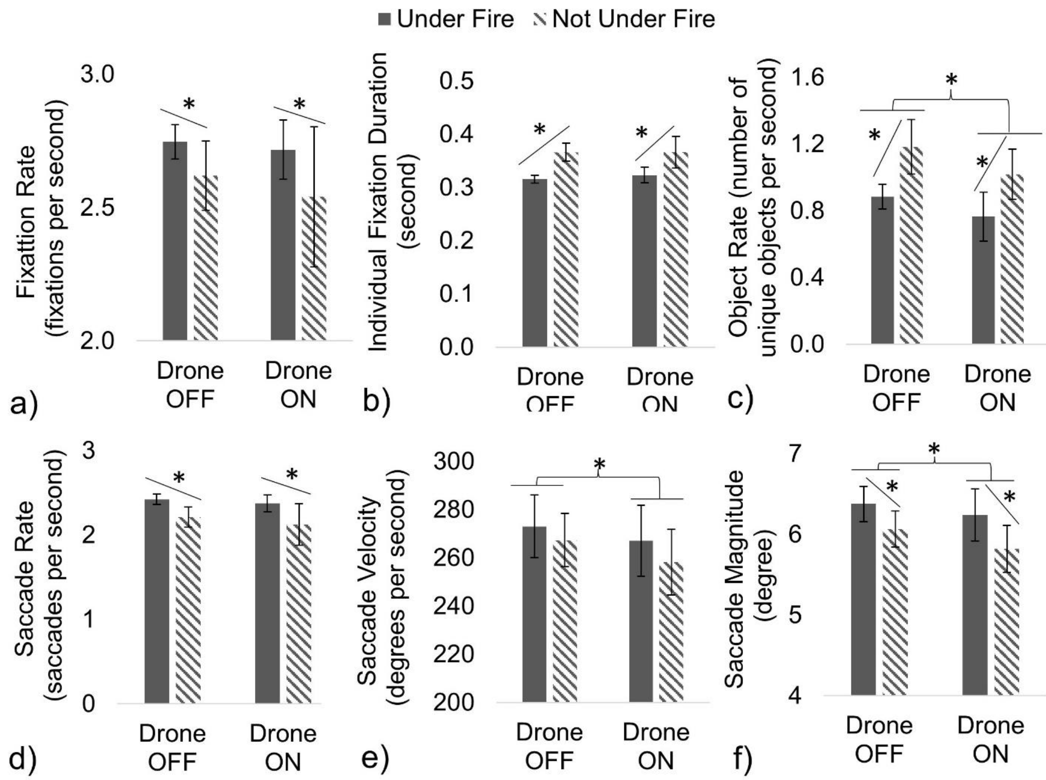

Fig. B-5 Mean squad fixation rate (a), saccade rate (d), and saccade magnitude (f) significantly decreased, and the individual fixation duration (b) and object rate (c) significantly increased during gun fire compared to when the squad was not under gun fire. Mean squad object rate (c), saccade velocity (e), and saccade magnitude significantly decreased with drone presence (f).

B.5.2.4 Squad Gaze Pattern on Distractors

Two separate repeated measures ANOVAs found that drone presence significantly increased the mean total number of distractor objects fixated while under fire (Fig. B-6b, F (1, 10) = 6.17, $p = 0.032$) but did not significantly impact the total number of distractor objects fixated on while not under fire (Fig. B-6a, F (1, 9) = 0.26, $p = 0.624$) for squads navigating the environment. A multivariate mixed model found that overall, both drone presence (F (2, 19) = 8.52, $p = 0.002$) and fire condition (F (2, 19) = 52.92, $p < 0.000$) impacted the mean number of fixations and mean dwell time on distractor objects (Fig. B-6c,d). The interaction between drone presence and fire condition was not significant (F (2, 19) = 0.564, $p = 0.578$). Follow-up separate univariate ANOVAs determined that the mean number of fixations (F (1, 20) = 11.64, $p = 0.003$) and mean dwell time (F (1, 20) = 16.79, p

= 0.001) both significantly increased when a drone was present. Additionally, both the number of fixations (F (1, 20) = 70.52, p < 0.000) and mean dwell time (F (1, 20) = 23.61, p < 0.000) significantly decreased when under gun fire compared to when not under gun fire. Distance between squad and bombs was not significantly impacted by drone presence (F (1, 19) = 1.04, p = 0.321) and fire condition (F (1, 19) = 2.83, p = 0.110).

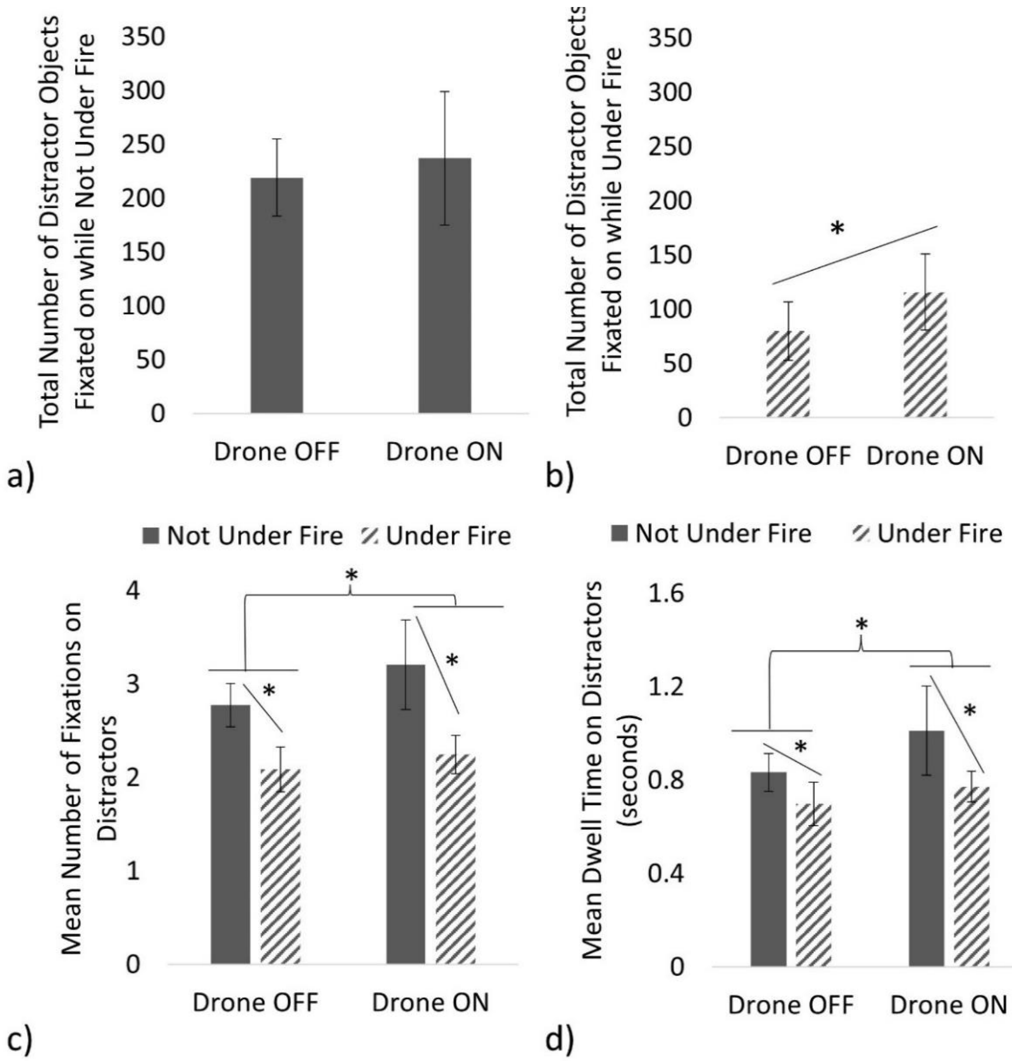

Fig. B-6 For gaze patterns on distractor objects, the total number of distractor objects fixated on in the environment was not significantly impacted by drone presence when not under fire (a), but when under gun fire, there was a significant increase in distractor objects fixated on (b). When drones were present, there was a significant increase in the mean number of fixations (c) and the mean dwell time (d) on distractors when compared to when drones were not present, and both significantly decreased (c,d) when under gun fire compared to not under gun fire.

B.5.2.5 Squad Gaze Patterns on Bombs

Two separate related-samples Wilcoxon signed rank tests determined that drone presence did not significantly affect the total number of bombs fixated on during gun fire (Fig. B-7b, Z = 1.00, $p = 0.317$) or not gun fire (Fig. B-7a, Z = 1.63, $p = 0.104$). A multivariate mixed model found that overall, both drone presence ($F(2, 18) = 4.25$, $p = 0.031$) and fire condition ($F(2, 18) = 43.83$, $p < 0.000$) impacted the mean number of fixations and mean dwell time on bombs (Fig. B-7c, d). The interaction between drone presence and fire condition was not significant ($F(2, 18) = 2.03$, $p = 0.160$). Follow-up separate univariate ANOVAs determined that the mean number of fixations ($F(1, 19) = 8.96$, $p = 0.007$) and mean dwell time ($F(1, 19) = 7.70$, $p = 0.012$) both significantly increased when a drone was present. Additionally, both the number of fixations ($F(1, 19) = 90.38$, $p < 0.000$) and mean dwell time ($F(1, 20) = 90.994$, $p < 0.000$) significantly decreased when under gun fire compared to when not under gun fire. A mixed model ANOVA showed that the distance between squad and bombs was not significantly impacted by drone presence ($F(1, 19) = 1.04$, $p = 0.321$) and fire condition ($F(1, 19) = 2.83$, $p = 0.110$).

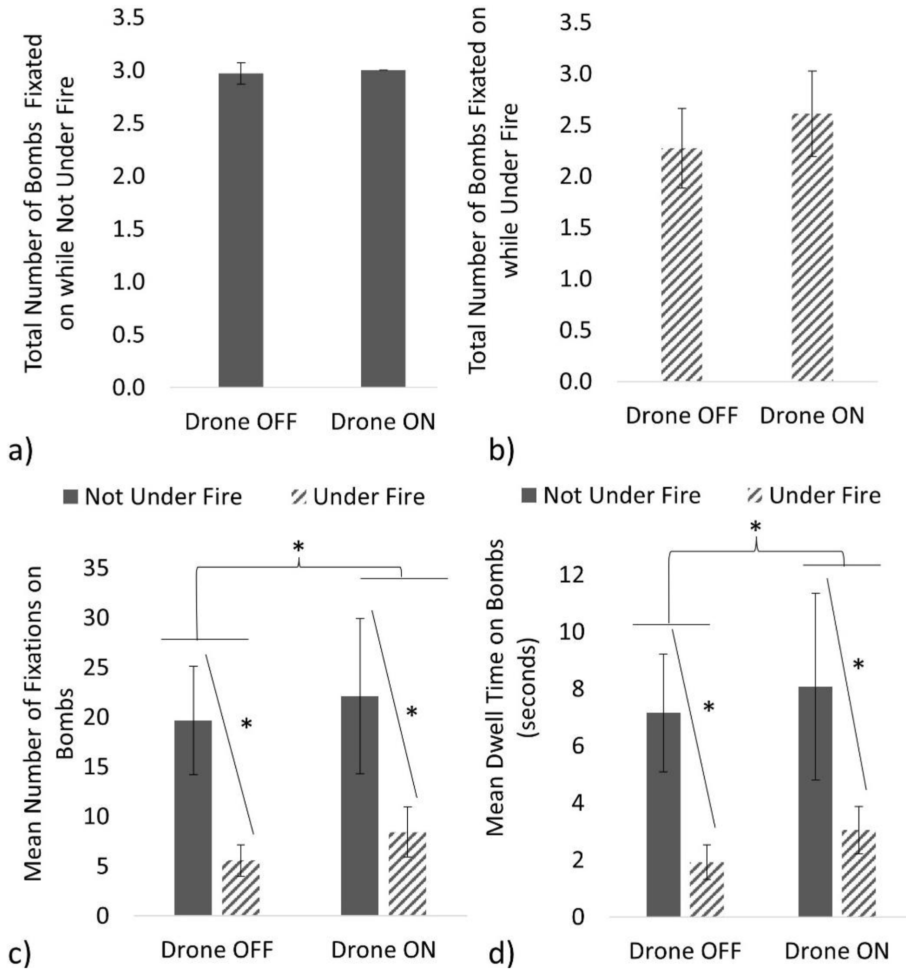

Fig. B-7 For gaze patterns on bombs, the total number of bombs fixated on in the environment was not significantly impacted by drone presence when not under fire (a) or when under gun fire (b). When drones were present, there was a significant increase in the mean number of fixations (c) and the mean dwell time (d) on bombs when compared to when drones were not present, and both significantly decreased (c,d) when under gun fire compared to not under gun fire.

B.5.2.6 Squad Gaze Patterns on Enemies

Two separate related-samples Wilcoxon signed rank tests determined that drone presence did not significantly affect the total number of enemies fixated on during gun fire (Fig. B-8b, Z = 1.00, p = 0.317) or not gun fire (Fig. B-8a, Z = 1.63, p = 0.104). A multivariate mixed model found that overall, neither drone presence (F (2, 13) = 0.53, p = 0.602), or fire condition (F (2, 13) = 0.24, p = 0.793), or the interaction between drone presence and fire condition (F (2, 13) = 1.50, p = 0.259) impacted the mean number of fixations and mean dwell time on bombs. A mixed model ANOVA showed that distance when fixating on enemies increased when the drones were present (Fig. B-8c, F (1, 15) = 6.42, p = 0.023) and was significantly reduced when enemies were present (F (1, 15) = 10.97, p = 0.005). The interaction of drone presence and firing condition was not significant (F (1, 15) = 0.874, p = 0.365).

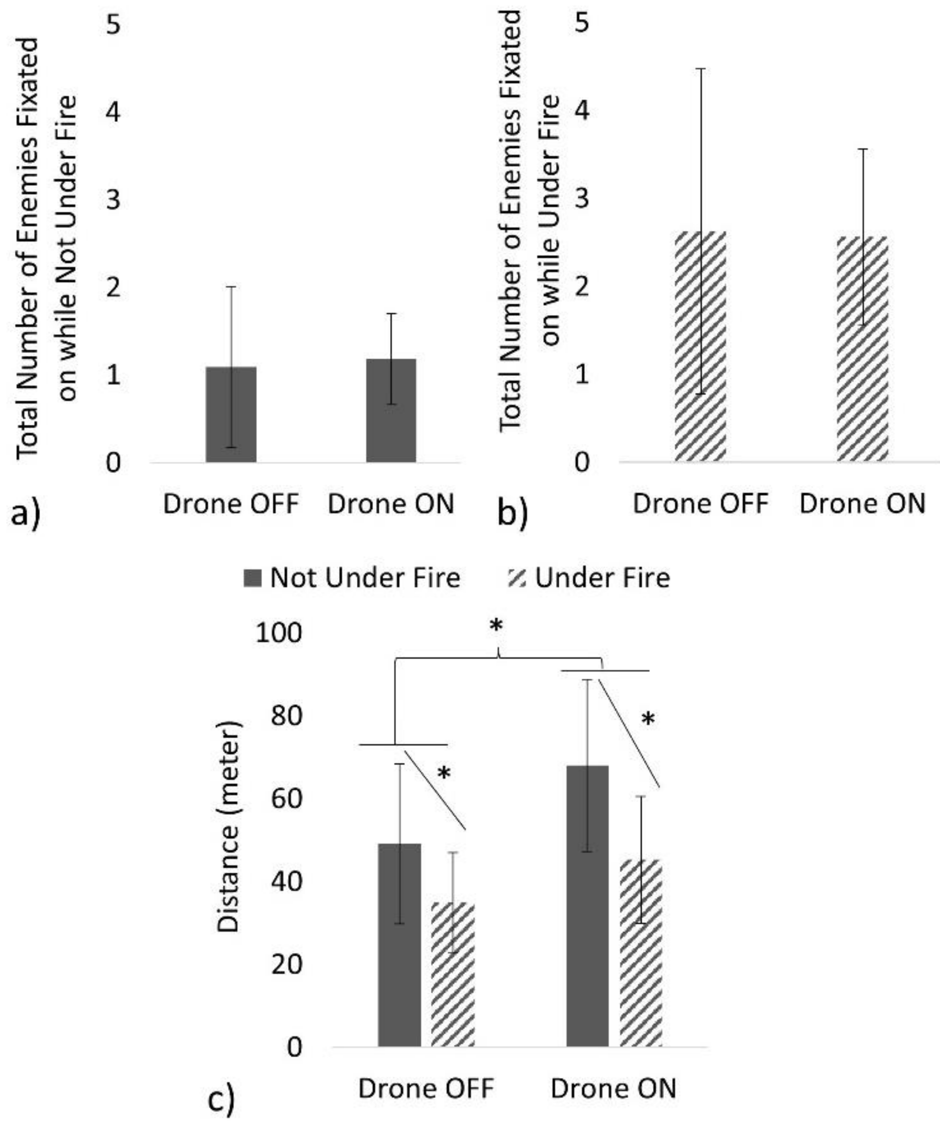

Fig. B-8 For gaze patterns on enemies, the total number of enemies fixated on in the environment was not significantly impacted by drone presence when not under fire (a) or when under gun fire (b). When drones were present, there was a significant increase in the distance by which the squad fixated on enemies compared to without the drone (c). Also, distance when fixated on enemies decreased significantly when under fire compared to when not under fire (c).

B.5.2.7 Summary

When drones were present and relaying information back to the commander, the squad covered significantly more distance in the environment without significantly changing their ground velocity. Since the drones feed enemy position information to the commander, this could be a reflection that the commanders navigated the squad around enemy encounters and, thus, total distance traveled increased. Interestingly, there does appear to be overall tendency to increase engagement with enemy fire with the drones and increased enemy hits (Fig. B-4). This could be because the drones were less successful in relaying positional information on rooftop enemies (who were also more difficult to visually detect by ground squad) and therefore could have increased a tendency for rooftop enemy engagements. In terms of general scanning of the environment, the presence of drones appeared to decrease the object rate (the number of unique objects fixated on per second), saccade velocity, and saccade magnitude. Therefore, although the squad maintained ground velocity, they appeared to scan the environment slower and visually attend (increased fixations and dwell time) on individual objects, specifically on distractors and bombs when the drones were present on the team. Interestingly, when the drone was present, the squad fixated on a significantly greater number of distractor objects during gun fire. Additionally, when the drone was present, the squad appeared to fixate (spot) enemies at significantly farther distances.

Not surprisingly, when engaged with the enemy, the squad's gaze pattern was dramatically impacted, regardless of drone presence. When under hostile threat, the squad significantly decreased fixation rate, saccade rate, and saccade magnitude and increased individual duration of fixations. Overall, the squad appeared to decrease the number of fixations and total dwell time on distractors and bombs and increase object rate. This mostly likely demonstrates the squad fixating and "aiming" at enemies (who were relatively difficult to see) and not fixating on any one object for very long. These results clearly demonstrate that eye gaze patterns are a viable metric for providing key insights into behavior and SA. Through these analyses, we were able to quantify the effect of autonomy on performance and physiology. We were able to make inferences about behavior, such as the presence of autonomy appears to reduce the team's perceptual and cognitive load, allowing them to focus on threats farther out and to be better prepared to engage in firefights.

B.6 Data Visualizer

The granularity of the data produced by UETack creates unique opportunities for visualization and additional simulation. UETack records nearly all actors (game objects), their spatial transformations, and the coordinates of the minimum bounding box that encloses all components of those entities. Thus, we developed a prototype data visualization program, TACK Visualizer in Panda3D (tvp3d) in Python using the Panda3D game engine (license https://www.panda3d.org/license/). Tvp3d leveraged actor data to recreate the scenario environment and track participants' movements throughout the world while highlighting environment actors according to the total number of gaze hits (number of times the participants' gaze vector intersected the object), all viewable with a freely moving camera. Participants' avatars (called pawns) were rendered as simple spheres with an exact rectangular frustum projection of the current display with respect to aspect ratio and field-of-view settings (Figs. B-9 and B-10). Summary statistics were also added via updating text overlays (Fig. B-10). Using the resulting visualization, an observer could intuit participants' situational awareness, changes in search behavior, and discrepancies between field of view and gaze hits (e.g., unusually low number of gaze hits on an object lying within view).

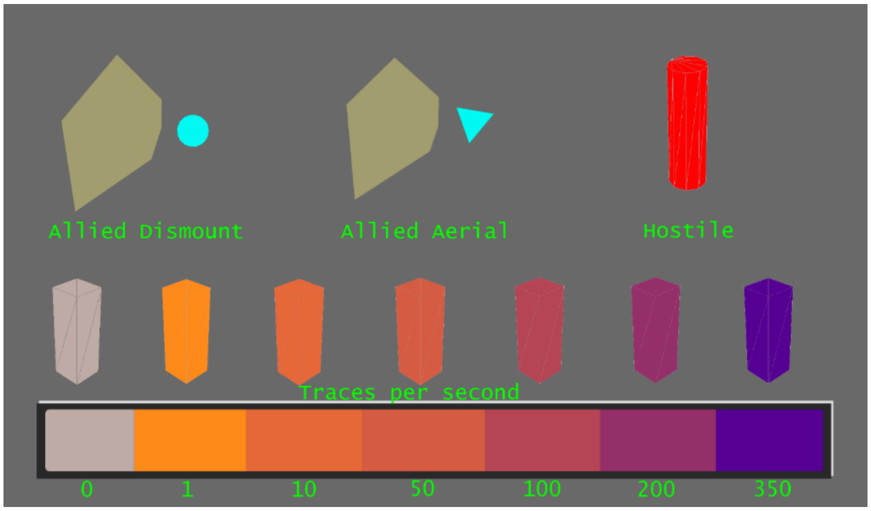

Fig. B-9 Legend for the tvp3d visualization used during the commander's interface tests. Allied dismounts were all live participants. Allied aerials included autonomous agents and the live commander. Hostiles included both static threats and computer-controlled enemy infantry. Static obstacles highlighted from orange to purple according to the number of total gaze hits during the prior second.

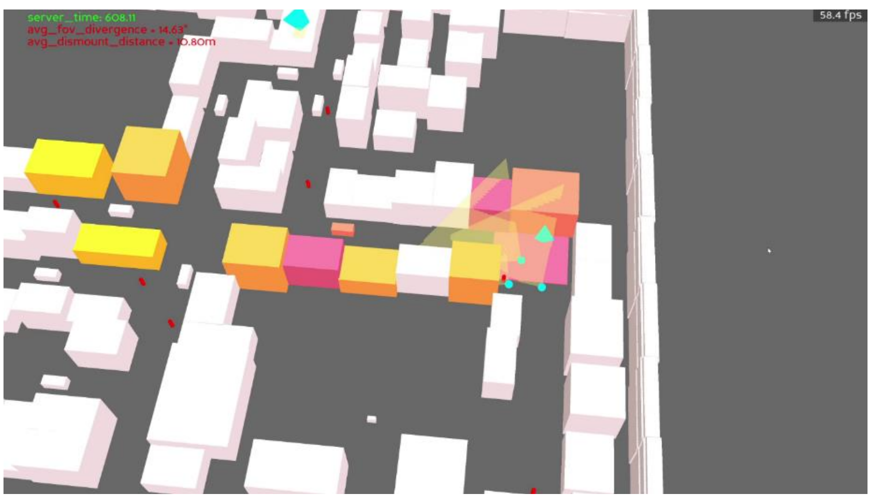

Fig. B-10 A screen capture from tvp3d during the middle of a commander's interface test, picturing three dismounts rounding a street corner alongside an autonomous aerial agent. The commander is also shown near the top of the frame. In the upper left, two dynamically updating aggregate statistics are displayed in addition to the unreal server time being visualized. "avg_fov_divergence" is the average yaw difference between all unique pairs of participants in degrees. "avg_dismount difference" is the average distance in meters across all unique pairs of participants.

While tvp3d can potentially be built into a powerful tool, it does have two major limitations at present. The greatest limitation is spatial fidelity: the axes-aligned bounding boxes that represent the space occupied by actors can overrepresent their apparent size in unreal (Fig. B-11). The second limitation is selectivity: selecting which actors to draw in the visualization and any post-hoc modifications to how they get represented requires deliberate querying and selection by someone with intimate knowledge of the simulation environment. The spatial fidelity problem could be addressed if future UETack development captures mesh components of actors; then their appearance could be recreated except for textures and transparency. Likewise, the selectivity problem can be addressed by researchers working closely with the scenario designers building the world in unreal to develop a tagging scheme to label actors according to how they should be represented in visualizations and simulations. Beyond that, actor selection is intrinsically difficult to automate. Despite these limitations, tvp3d exemplifies how UETack facilitates analyses that unite in-simulation behavior with eye tracking.

Fig. B-11 A side-by-side comparison demonstrating the spatial fidelity limitation. On the left, a curved section of wall in the commander's interface world. On the right, the tvp3d representation of the actors. Because the curve lies diagonal to the x and y world axes, the smallest axes-parallel envelope becomes wider along both the x and y axes. While the straight sections of wall are axes-aligned and accurately represented, the curved section's size becomes inflated.

Appendix C. Other Applications

C.1 STRONG

The U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) collaborative program Strengthening Teamwork for Robust Operations in Novel Groups (STRONG) aims to bring government, contractor, and academic researchers together to develop the foundation for enhanced teamwork within heterogeneous human–intelligent agent teams. The STRONG Cooperative Research Alliance is an extramural funding vehicle that encourages external researchers to propose research projects towards ARL problem spaces. This program led a Virtual Environments Workshop (VEW) and hackathon during the 2020 Annual STRONG Summer Summit.

Specifically, the VEW allowed summit attendees to explore and discuss how virtual environments are used in current research efforts within and outside of ARL. The VEW showcased the Tactical Awareness via Collective Knowledge (TACK) concept and Unreal Environment TACK (UETack) capabilities, providing a demonstration and analysis activities allowing attendees to explore, share, and discuss approaches to analyzing team-based data. Discussions focused on team processes metrics that could be extracted, appropriate and innovative analyses implementations, and discussion of results.

VEW attendees were granted access to an online custom environment (virtual laboratory) developed for this workshop. Data was collected from VEW attendees that voluntarily elected to participate in the demonstration, joining teams to explore the environment. The purpose of this exercise was to demonstrate the capabilities of the virtual environment. The data generated and collected from these sessions was then provided back to the VEW attendees for the analysis. Individuals or teams could explore the virtual laboratory for approximately 10 min while collecting letters that were scattered throughout the environment. Metrics collected included camera angle (as a proxy for eye gaze), number of letters collected, and location in the virtual environment. No personally identifiable data was collected, and numerical identifiers were applied to the data. VEW attendees were then given the opportunity to analyze the de-identified data and report their findings during the VEW feedback sessions. Results of those analyses were used to stimulate conversation during the STRONG Summer Summit.

The VEW generated interest for leveraging UETack environments from a variety of STRONG performers. One of the STRONG Cycle 2 awardees (University of Colorado and DCS Corp) is currently leveraging a virtual factory environment (Fig. C-1) to collect multimodal physiology data within dyads conducting a “Tied Team” task to understand team dynamics. Teams are tasked with finding and collaboratively disarming simulated bombs hidden throughout the virtual environment while behavioral and physiological data (e.g., eye tracking) are simultaneously collected (Fig. C-1).

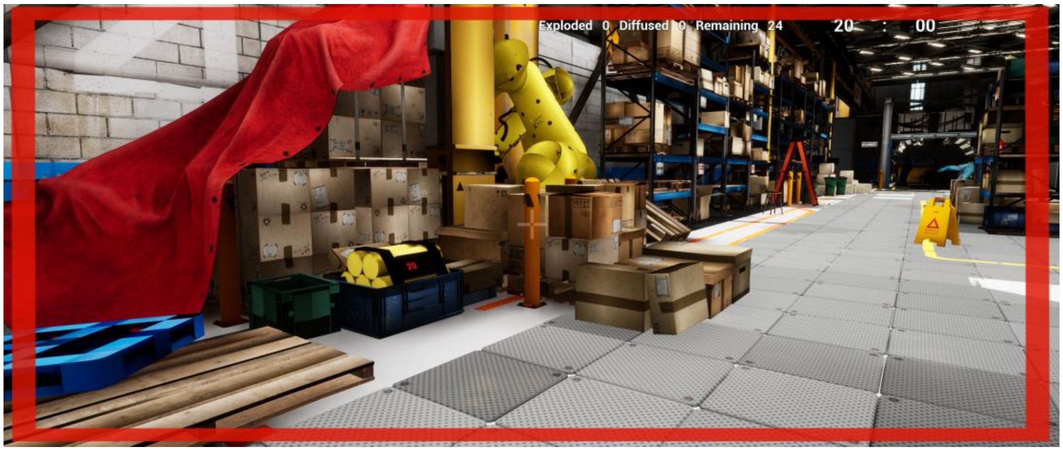

Fig. C-1 STRONG Tied Teaming virtual bomb diffusion task

By leveraging the UETack, this STRONG team was able to seamlessly collect behavioral and synchronized physiological data from a multiplayer game. This effort is focused on team dynamics and understanding how roles emerge within teams. For instance, by applying the Granger-causality algorithm to positional data, the research team was able to understand who is leading and who is following at any given moment (Fig. C-2).

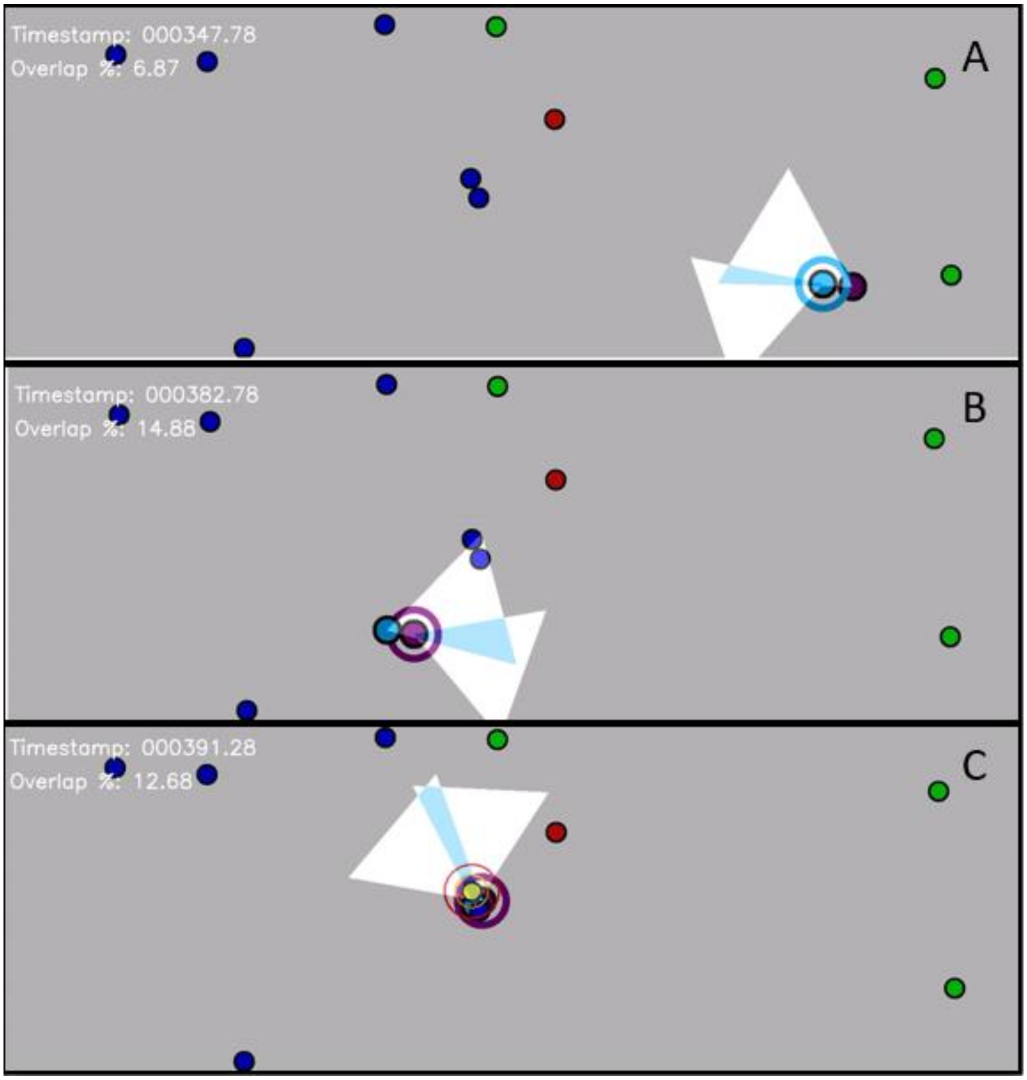

Fig. C-2 STRONG Leader–Follower Detection Model: (A) Blue player emerges first as the leader. (B) Leader position has switched to the purple player. (C) Purple leads team to diffuse bomb (highlighted in yellow).

C.2 FAALCOR

FAALCOR (Fast Adaptive Agent Learning for Coordinated Optimal Reasoning) is a 3D virtual experimental platform used to understand how humans adapt to changing environments. In this paradigm, individuals and teams control avatars to collect resources under varying constraints. The paradigm allows for manipulation of threats, movement costs, uncertainty, reward amount, and other variables. To achieve this, FAALCOR relies on the UETack software suite to enable multiple individuals to engage in the same local environment or single individuals to play via a browser on a cloud environment. Currently, we are running an experiment where individuals will navigate a virtual 3D environment collecting resources distributed throughout the environment (Fig. C-3). Participants have a limited amount of energy with which to plan their route and are instructed to collect as many resources and points as possible. We are specifically interested in understanding how individuals' behavior and physiology change in response to environmental context shifts (i.e., an increase in stamina expended while moving). Changing environmental contexts can include changes in stamina cost, resource reward amount, resource distribution, amount of uncertainty, and number of teammates. This experimental platform will serve as a proof-of-concept for other ARL efforts that need to capture single or multiplayer data from human-in-the-loop simulations deployed on either local or cloud-based servers.

Fig. C-3 First-person view of virtual environment. Vertical green bar on left depicts amount of stamina. Top right indicates team score, individual score, and overhead map of resources in the environment.

Appendix D. Publications

Arefin MS, Swan JE II, Cohen Hoffing RA, Thurman SM. Estimating perceptual depth changes with eye vergence and interpupillary distance using an eye tracker in virtual reality. 2022 Symposium on Eye Tracking Research and Applications. 2022;1–7. https://doi.org/10.1145/3517031.3529632.

Callahan-Flintoft C, Barentine C, Touryan J, Ries AJ. A case for studying naturalistic eye and head movements in virtual environments. Frontiers in Psychology. 2021;12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.650693

Cohen Hoffing R, Thurman S, Coyne J, Sibley C, Enders L, Roy H. Leveraging the pupillary light reflex for cognitive pupillometry: an initial characterization of the PLR in two data sets. Journal of Vision. 2023;23(9):5839. https://doi.org/10.1167/jov.23.9.5839.

Dalangin B, Oiknine A, Villarreal J, Jeter A, Gordon S, Hoffing RA, Thurman S, Smith R, Enders L, Roy H, et al. Using eye tracking and electroencephalography (EEG) for visual search experimentation in virtual environments: a data guide. DEVCOM Army Research Laboratory (US); 2020. Report No.: ARL-TR-9045. Accession No.: AD1109153. https://apps.dtic.mil/sti/citations/AD1109153.

Enders LR, Gordon SM, Roy H, Rohaly T, Dalangin B, Jeter A, Villarreal J, Boykin GL, Touryan J. Evidence of elevated situational awareness for active duty Soldiers during navigation of a virtual environment. PloS one. 2024;19(5).

Enders LR, Smith RJ, Gordon SM, Ries AJ, Touryan J. Gaze behavior during navigation and visual search of an open-world virtual environment. Frontiers in Psychology. 2021;12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.681042.

Enders L, Roy H, Rohaly T, Jeter A, Villarreal J. Impacts of posttraumatic stress disorder on eye-movement during visual search in an open virtual environment under high and low stress conditions. Journal of Vision. 2023;23(9):5691. https://doi.org/10.1167/jov.23.9.5691.

Gordon SM, Lawhern VJ, Touryan J. Assessing temporal variability in fixation-locked P300 responses during free-viewing visual search. 2023 11th International IEEE/EMBS Conference on Neural Engineering (NER). 2023;1–4. https://doi.org/10.1109/NER52421.2023.10123724.

Gordon SM, McDaniel JR, King KW, Lawhern VJ, Touryan J. Decoding neural activity to assess individual latent state in ecologically valid contexts. Journal of Neural Engineering. 2023;20(4):046033. https://doi.org/10.1088/1741–2552/acee20.

Hoffing RA, Thurman S. The state of the pupil: moving toward enabling real-world use of pupillometry-based estimation of human states. DEVCOM Army Research Laboratory (US); 2020. Report No.: ARL-TR-8947. Accession No.: AD1097314. https://apps.dtic.mil/sti/citations/AD1097314.

Ki JJ, Dmochowski JP, Touryan J, Parra LC. Neural responses to natural visual motion are spatially selective across the visual field, with selectivity differing across brain areas and task. European Journal of Neuroscience. 2021;54(10):7609–7625. https://doi.org/10.1111/ejn.15503.

Lance BJ, Larkin GB, Touryan JO, Rexwinkle JT, Gutstein SM, Gordon SM, Toulson O, Choi J, Mahdi A, Hung CP, et al. Minimizing data requirements for Soldier-interactive AI/ML applications through opportunistic sensing. Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II. 2020;11413:19–27. https://doi.org/10.1117/12.2564515.

Roy H, Enders LR, Rohaly T, Dalangin B, Villarreal J, Jeter A. Posttraumatic stress disorder and differences in eye gaze during a visual search task with cognitive load. Proceedings of the Annual Meeting of the Cognitive Science Society. 2022;44. https://escholarship.org/uc/item/2nj601jj.

Stankov AD, Touryan J, Gordon S, Ries AJ, Ki J, Parra LC. During natural viewing, neural processing of visual targets continues throughout saccades. Journal of Vision. 2021;21(10):7. https://doi.org/10.1167/jov.21.10.7.

Thurman SM, Cohen Hoffing RA, Madison A, Ries AJ, Gordon SM, Touryan J. Blue sky effect: contextual influences on pupil size during naturalistic visual search. Frontiers in Psychology. 2021;12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.748539.

Wu YC, Chang C, Liu W, Stevenson C, Cohen Hoffing R, Thurman S, Jung T-P. Free moving gaze-related electroencephalography in mobile virtual environments. Journal of Vision. 2023;23(9):5305. https://doi.org/10.1167/jov.23.9.5305.

Bibliography

Buck J. Maze generation: Prim’s algorithm. The Buckblog. 2011 Jan 11. https://weblog.jamisbuck.org/2011/1/10/maze-generation-prim-s-algorithm.

Day J, Davidenko N. Parametric face drawings: a demographically diverse and customizable face space model. Journal of Vision. 2019;19(11):7. https://doi.org/10.1167/19.11.7.

Endsley MR, Holder LD, Leibrecht BC, Garland DJ, Wampler RL, Matthews MD. Modeling and measuring situation awareness in the infantry operational environment. Army Research Institute for the Behavioral and Social Sciences (US); 2000. Accession No.: ADA372709. https://apps.dtic.mil/sti/pdfs/ADA372709.pdf.

Epic Games, Inc. Unreal Engine 4.27 documentation; 2021. https://docs.unrealengine.com/4.27/en-US/Resources/SampleGames/ShooterGame/.

Ettenhofer ML, Hershaw JN, Barry DM. Multimodal assessment of visual attention using the Bethesda Eye & Attention Measure (BEAM). Journal of Clinical and Experimental Neuropsychology. 2016;38(1):96–110. https://doi.org/10.1080/13803395.2015.1089978.

Golding JF. Motion sickness susceptibility. Autonomic Neuroscience. 2006;129(1):67–76. https://doi.org/10.1016/j.autneu.2006.07.019.

Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. The International Journal of Aviation Psychology. 1993;3(3):203–220+. https://doi.org/10.1207/s15327108ijap0303_3.

Khanna MM, Badura-Brack AS, McDermott TJ, Embury CM, Wiesman AI, Shepherd A, Ryan TJ, Heinrichs-Graham E, Wilson TW. Veterans with PTSD exhibit altered emotional processing and attentional control during an emotional Stroop task. Psychological Medicine. 2017;47(11):2017–2027. https://doi.org/10.1017/S0033291717000460.

Khanna MM, Badura-Brack AS, McDermott TJ, Shepherd A, Heinrichs-Graham E, Pine DS, Bar-Haim Y, Wilson TW. Attention training normalises combat-related post-traumatic stress disorder effects on emotional Stroop performance using lexically matched word lists. Cognition and Emotion. 2016;30(8):1521–1528. https://doi.org/10.1080/02699931.2015.1076769.

Mumble Project. Mumble: open source, low latency, high quality voice chat. Mumble Project, n.d. [accessed 2023 Dec 8] from www.mumble.info/.

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971;9(1):97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

Purtiman R. Soldier Lethality Cross-Functional Team bringing next generation technologies to Soldiers. Department of Army (US); 2018 Sep 24. https://www.army.mil/article/211454/soldier_lethality_cross_functional_team_bringing_next_generation_technologies_to_soldiers.

Solon AJ, Lawhern VJ, Touryan J, McDaniel JR, Ries AJ, Gordon SM. Decoding P300 variability using convolutional neural networks. Frontiers in Human Neuroscience. 2019;13:201. https://doi.org/10.3389/fnhum.2019.00201.

Thompson M. Soldier Lethality team reimagines movement, vision and combat capabilities. Department of Army (US); 2022 Mar 15. https://www.army.mil/article/254732/soldier_lethality_team_reimagines_movement_vision_and_combat_capabilities.

Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, Marcus DJ, Westerlund A, Casey B, Nelson C. The NimStim set of facial expressions: judgments from untrained research participants. Psychiatry Research. 2009;168(3):242–249. https://doi.org/10.1016/j.psychres.2008.05.006.

Usoh M, Catena E, Arman S, Slater M. Using presence questionnaires in reality. Presence. 2009;(5):497–503. https://doi.org/10.1162/105474600566989.

Witmer BG, Jerome CJ, Singer MJ. The factor structure of the presence questionnaire. Presence: Teleoperators and Virtual Environments. 2005;14(3):298–312. https://doi.org/10.1162/105474605323384654.

List of Symbols, Abbreviations, and Acronyms

3D	three-dimensional
AI	artificial intelligence
ANOVA	Analysis of Variance
AR	Augmented Reality
ARL	Army Research Laboratory
ARL-W	ARL-West
BEAM	Bethesda Eye and Attention Measure
C5ISR	Command, Control, Communication, Computers, Cyber, Intelligence, Surveillance and Reconnaissance
CFT	Cross-Functional Team
CI	Commanders Interface
COA	course of action
COAT	Crew Optimization and Augmentation Tech
COP	common operating picture
DEVCOM	U.S. Army Combat Capabilities Development Command
EEG	electroencephalogram
EIT	emotional interference task
EST	emotional Stroop tasks
FAALCOR	Fast Adaptive Agent Learning for Coordinated Optimal Reasoning
FY	fiscal year
GUI	graphical user interface
HID	Human Interest Detector
HUD	Heads-Up Display
IMU	inertial measurement unit
INFORMS	Information for Mixed Squads

IVAS	Integrated Visual Augmentation System
JBSA	Joint Base San Antonio
MAST	Maladaptive Stress Response Testing
ML	machine learning
MSSQ	measured motion sickness questionnaire
PTSD	posttraumatic stress disorder
S&T	science and technology
SA	situational awareness
SL	Soldier Lethality
SSQ	simulator sickness questionnaire
STP	Soldier Touch Point
STR	Summary Technical Report
STRONG	Strengthening Teamwork for Robust Operations in Novel Groups
TACK	Tactical Awareness via Collective Knowledge
TRL	Technical Readiness Level
UAS	unmanned aircraft system
UE	Unreal Engine
UETack	Unreal Environment TACK
VE	virtual environment
VEW	Virtual Environments Workshop
YOLO	You Only Look Once

1 (PDF) DEFENSE TECHNICAL
INFORMATION CTR
DTIC OCA

1 (PDF) DEVCOM ARL
FCDD RLB CI
TECH LIB

1 (PDF) DA HQ
DASA(R&T)

6 (PDF) USARMY AFC
L BROUSSEAU
A LINZ
S BRADY
J REGO
T KELLY
B SESSLER

2 (PDF) DEVCOM HQ
FCDD ST
C SAMMS
M HUBBARD

72 (PDF) DEVCOM ARL
FCDD RLA
C BEDELL
A SWAMI
M GOVONI
M WRABACK
H EVERITT
J CHEN
PJ FRANASZCZUK
C TEETER
A SCHOFIELD
J LANE
FCDD RLA A
S KARNA
AM RAWLETT
SE SCHOENFELD
FCDD RLA B
A WEST
FCDD RLA C
C OLIVER
G BRILL
G LARKIN
FCDD RLA J
J FOSSACECA
FCDD RLA CB
P GILLICH
FCDD RLA CC
C KRONINGER
FCDD RLA CD
J ROBINETTE
FCDD RLA CL
F FRESCONI
FCDD RLA CG
D STRATIS-CULLUM
FCDD RLA CH
B PERELMAN
FCDD RLA CV
M KWEON
FCDD RLA F
JR GASTON
FCDD RLA FC
H ROY
FCDD RLA G
ML REED
FCDD RLA H
JJ SUMNER
FCDD RLA I
D WIEGMANN
FCDD RLA L
RD DEL ROSARIO
FCDD RLA M
K CHO
FCDD RLA N
BM RIVERA
FCDD RLA PC
P PELLEGRINO
FCDD RLA PD
F FATEMI
FCDD RLA T
RZ FRANCART
FCDD RLA V
S SILTON
FCDD RLA W
TV SHEPPARD
FCDD RLB
K PLOSKONKA
FCDD RLR ET
B ASHFORD
FCDD RLB RW
P KHOOSHABEH
FCDD RLB S
K KAPPRA
J VETTEL
FCDD RLD
PJ BAKER
L VANDERHOEF
FCDD RLD A
A MOGRO
FCDD RLD C
T KINES
FCDD RLD D
J ALEXANDER
FCDD RLD I
A FINCH
A LLOPIS-JEPSEN
D ROLL

S SHIDFAR
FCDD RLD M
N ZANDER
M JOHNSON
FCDD RLR
P REYNOLDS
K RASMUSSEN
FCDD RLR A
D STEPP
FCDD RLR C
LL TROYER
FCDD RLR D
R ZACHERY
FCDD RLR DS
R FREED
FCDD RLR E
RA MANTZ
FCDD RLR EF
F GREGORY
FCDD RLR EG
H DE LONG
FCDD RLR EH
V MARTINDALE
FCDD RLR EI
SP IYER
FCDD RLR EL
J QIU
FCDD RLR EM
C VARANASI
FCDD RLR EN
JM COYLE
FCDD RLR EP
PM BAKER
FCDD RLR EV
M HAILE
FCDD RLR EW
JD MYERS
FCDD RLR G
A SCRUGGS