UNIVERSITY OF VIRGINIA

Eyes on Target: A Real-Time Eye Tracking Approach for Enhanced Performance and Interruption Management of Teams in UAV C2 Operations

A DISSERTATION

submitted in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Systems Engineering

by

Jad A. Atweh

May 2025

Approval Sheet

This Dissertation

is submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Systems Engineering)

Author: Jad A. Atweh

This Dissertation has been approved by the following examining committee:

Advisor: Sara Riggs (Systems and Information Engineering)

Chair: Matthew Bolton (Systems and Information Engineering)

Committee Member: Rupa Valdez (Systems and Information Engineering)

Committee Member: Seongkook Heo (Computer Science)

Committee Member: Shannon McGarry (Naval Research Laboratory)

Accepted for the School of Engineering and Applied Science:

Jennifer L. West, Dean of the School of Engineering and Applied Science

May 2025

To Mom, Dad, Mostafa, & Maya

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

They say it takes a village—mine spanned across continents, disciplines, and countless conversations, each voice shaping this work in ways big and small. First and foremost, I want to express my deepest gratitude to my advisor, Dr. Sara Riggs. From the moment I was an undergraduate intern in your lab, you believed in me, and you never stopped. The last four years would not have been possible without your unwavering guidance, support, and ability to respond to my texts in under 60 seconds (yes, I counted). Your dedication to your students is unmatched, and I am beyond grateful for your mentorship and the countless hours you have invested in me.

I am also incredibly thankful to my committee members. Dr. Matthew Bolton, thank you for always challenging my thinking with your unique perspectives and thought-provoking conversations. Your ability to reframe problems in ways I hadn't considered has made me a better researcher. Dr. Rupa Valdez, I deeply appreciate your mentorship in qualitative methods and macroergonomics and the Zoom calls whenever I had qualitative coding questions. Dr. Seongkook Heo, your insights during my proposal and our discussions along the way have been invaluable. And finally, a special thank you to Dr. Shannon McGarry, a fellow Riggs Lab alum, who has supported me from the start. Our long conversation in that hotel lobby in Arizona will always stand out as a pivotal moment in shaping Chapter 6 of this dissertation.

To current and previous Riggs Lab colleagues—thank you for the support, the laughs, and the countless hours spent growing together. I especially want to thank my lab cohort who joined with me. Our conversations on research and life have been just as important as the work itself. I also want to thank Jawad Alami for redesigning the simulator used in my studies, which was instrumental in conducting this research. A special thank you to my undergraduate interns—Rana Hazimeh, Jackie Al Hayek, Ihab Tabbara, Ghady Nasrallah, and Noah Simsic—for their contributions to this work. Mentoring you all has been a rewarding experience, and I have learned just as much from you as you have from me.

To my collaborators and colleagues, thank you for the opportunity to work on exciting projects that have shaped my research journey. I especially want to thank Dr. Nadine Moacdieh for guiding me through research during my undergraduate years, shaping my early understanding of the field and instilling in me the curiosity and drive that led me to pursue this path. I also want to extend my gratitude to the Diversity Programs Office, where I had the privilege of working for two summers.

I am also grateful to the SIE department at UVA for providing the resources and technical support necessary for my research. A huge shoutout to Chantel Gross—genuinely, without you, the department would fall apart. I hope you know how much we all appreciate everything you do. Moreover, this work was made possible through the generous financial support of the NSF and SIE, and I am incredibly grateful for the funding that allowed me to pursue this research.

Additionally, I extend my thanks to Amazon and the colleagues I met there for the Research Scientist internship experience that broadened my perspective and deepened my understanding of industry applications. Many of the insights I gained during my time there directly informed and strengthened aspects of this dissertation which I will bring back to my students and the classroom.

To my friends in Lebanon and the U.S., thank you for being my support system. My SIE friends—Nora Scheer, Jessica Arora, Lauren Horde, Skye Taylor, Nicholas Gardella, Ekim Koca—your friendship has been invaluable. A special thank you to Nora and Jessica, who have been the closest thing to family in Charlottesville—I will never forget our little Camellias runs, where work and life blended seamlessly. To my UVA friends—Archit Uniyal, Lama Khraibani, Joelle Youness, Leah Shaheen, Juliet Wallemacq, and Anne Zhang—I am so grateful for the time we spent together. I also want to thank Seungbeen (Daniel) Lee for taking care of Flappy during my faculty visits and while I was presenting key findings from this dissertation at conferences. Our paths crossing has been a privilege, and I am grateful for your friendship. And to my Lebanese Club family, thank you for keeping a little piece of home with me throughout this journey.

To my best friend, Rawan Akl—we have been through some of the toughest times together, supporting each other through challenges that felt insurmountable. Despite being 6,000 miles apart, our friendship has remained unwavering, untouched by time or distance. You have been an important source of comfort through every triumph and setback. True friendships like ours are rare, and I am endlessly grateful for the bond we share.

Finally, to my family. From the very beginning, you supported my dream to pursue my Ph.D. in the U.S. wholeheartedly, encouraging me to take this leap despite the challenges. This path was never mine alone—it was paved by your sacrifices, your belief in me, and your endless encouragement. My mom, Fatima Ayad, and my dad, Ali Atweh, I am endlessly grateful for the sacrifices you have made, from before I was born until this very moment. Without your sacrifices and without you by my side, I would not be the person I am today. Mom, you have taught me everything I know about resilience, hard work, and compassion. Dad, your strength and unwavering support mean more to me than I can express. To my siblings, Mostafa and Maya, thank you for always answering my calls (even when they were just to rant) and for being my constant source of laughter and perspective.

And, of course, to Flappy, who has been my steadfast furry companion through the long nights of writing and analysis, providing much-needed stress relief (and frequent keyboard interference).

To everyone who has been part of this journey—thank you. I could not have done this alone.

VITA

Jad A. Atweh

EDUCATION

Doctor of Philosophy in Systems Engineering **2025**
University of Virginia *Charlottesville, VA*

Master of Engineering in Systems Engineering **2023**
University of Virginia *Charlottesville, VA*

Bachelor of Engineering in Industrial Engineering **2021**
American University of Beirut *Beirut, Lebanon*

RESEARCH EXPERIENCE

Graduate Research Assistant **2021–2025**
University of Virginia *Charlottesville, VA*

Research Scientist II Intern **2023**
Amazon *Seattle, WA*

TEACHING EXPERIENCE

Graduate Teaching Assistant **2021–2025**
University of Virginia *Charlottesville, VA*

REFEREED JOURNAL PUBLICATIONS

Eyes Over Ears? Exploring the Interaction of Gaze Sharing and Team Verbal Communication in UAV C2 Operations **2025**
Journal of *Applied Ergonomics*

Quantifying Collaborative Strategies and Identifying Performance Breakdowns of UAV C2 Teams Using Multidimensional Cross-Recurrence Quantification Analysis **2025**
International Journal of Human-Computer Studies

Designing and Implementing Exoskeleton Devices for Nurses with Musculoskeletal Disorders **2025**
Journal of *Applied Ergonomics*

Gaze Sharing, a Double-Edged Sword: Examining the Effect of Real-Time Gaze Sharing Visualizations on Team Performance and Situation Awareness **2024**
Journal of *Human Factors*
Editors' Pick

Does It MultiMatch? What Scanpath Comparison Tells us About Task Performance in Teams **2023**
Journal of *Cognitive Engineering and Decision Making*

SELECTED CONFERENCE PROCEEDINGS

Predicting Mental Demand of Teammates Using Eye Tracking Metrics: A Machine Learning Approach **2025**
Proceedings of the *ACM Symposium on Eye Tracking Research and Applications (ETRA)*

From Disruptions to Decisions: The Role of User-Controlled Gaze Sharing in Team Interruption Recovery Across Task Complexities **2025**
Proceedings of the *Human Factors and Ergonomics Society Annual Meeting*

Open Communication with Real-Time Gaze Sharing Enhances Team Performance by Reducing Saccadic Activity and Workload **2025**
Proceedings of the *Human Factors and Ergonomics Society Annual Meeting*

Can We See Team Situation Awareness? Predicting Team SA Using Eye Tracking and Machine Learning **2025**
Proceedings of the *Human Factors and Ergonomics Society Annual Meeting*

Display Design Shapes How Eye Tracking and Workload Predict Team Performance **2025**
Proceedings of the *IEEE Systems and Information Engineering Design Symposium (SIEDS)*

Real-Time Gaze Sharing Techniques and Their Influence on Shared Situation Awareness of Teammates in UAV C2 Operations **2024**
Proceedings of the *Human Factors and Ergonomics Society Annual Meeting*

Second-Place Winner of the HFES Cognitive Engineering and Decision Making (CEDM) Technical Group Best Student Paper Award

ScanMatch Versus MultiMatch: A Comparison of the Sensitivity of Scanpath Similarity Metrics to Changes in Workload **2024**
Proceedings of the *IEEE SIEDS*

Winner of the Optimization, Data, and Systems Design Track Best Paper Award

Reducing Barriers in Analyzing Eye Tracking Data: The Development of a GUI to Preprocess Eye Tracking Data **2024**
Proceedings of the *IEEE SIEDS*

Can Real-Time Gaze Sharing Help Team Collaboration? A Preliminary Examination of its Effectiveness with Pairs **2023**
Proceedings of the *Human Factors and Ergonomics Society Annual Meeting*

Quantifying Visual Attention of Teams During Workload Transitions Using AOI-Based Cross-Recurrence Metrics **2023**
Proceedings of the *Human Factors and Ergonomics Society Annual Meeting*

Identifying Individual-, Team-, and Organizational-Level Factors that Affect Team Performance in Complex Domains Based on Recent Literature **2022**
Proceedings of the *Human Factors and Ergonomics Society Annual Meeting*

ABSTRACT

Teams operating in complex systems, such as Unmanned Aerial Vehicle (UAV) command-and-control (C2) environments, face significant challenges in maintaining situation awareness (SA), managing workload, and ensuring effective coordination. Understanding how these teams interact and adapt under varying conditions is critical to optimizing performance and resilience. This dissertation explores the use of eye tracking technology as a real-time, unobtrusive tool to study and enhance team dynamics in these high-stakes settings.

The first phase of this research validates the use of scanpath similarity techniques for assessing team performance under different workload conditions. Both (1) Multidimensional Cross-Recurrence Quantification Analysis and (2) MultiMatch metrics have shown to be sensitive to workload changes and correlate with team performance, offering a promising approach to monitoring team states in real-time. However, the application of real-time metrics to understand and support team-level performance remains largely uncharted territory.

Building on these findings, this dissertation investigates the potential of gaze sharing as a novel intervention to improve team collaboration. By enabling team members to visualize their partner's gaze in real-time, gaze sharing might address inefficiencies such as redundant overlapping visual attention during tasks. This research evaluates various gaze sharing visualization techniques, including fixation dots and trails, to determine their impact on workload, SA, scanning behavior, and task performance. Additionally, this work explores the integration of gaze sharing with communication strategies and its role in helping teams recover from interruptions. The findings reveal that gaze sharing using the fixation trail not only complements verbal and non-verbal communication, but also serves as a valuable tool for enhancing coordination during disruptions.

This dissertation bridges gaps in the theory of team collaboration within UAV C2 environ-

ments, contributing to both cognitive systems engineering and human-computer interaction. It advances theoretical understanding by identifying metrics and mechanisms that define effective team dynamics under varying conditions. Moreover, it provides actionable guidelines for integrating gaze sharing technologies into operational systems and lays the foundation for adaptive displays that monitor and support team performance in real-time. The contributions of this research extend beyond UAV operations, offering insights and methodologies that can be applied to other domains such as air traffic control and emergency response. By addressing the challenges of team coordination and workload management in complex systems, this work provides a critical step toward designing technologies that enhance both the safety and efficiency of collaborative operations in the modern world.

Chapter 1

Introduction, Motivation, and Overview of Dissertation

1.1 Teams in Complex Systems

Teams are the foundation of many organizations and corporations, where a team is formally defined as two or more people who have precise roles and rely on one another to accomplish a common objective (Salas et al., 1992). From building the ancient pyramids in Egypt to modern day space exploration, collaboration and coordination among team members have been essential for achieving goals that cannot be accomplished by individuals working alone (Salas et al., 2008).

In many environments, teamwork occurs within what are referred to as "complex systems." Complex systems are characterized by high interdependence, dynamic and often unpredictable variables, and a need for rapid decision-making under significant constraints (Hollnagel & Woods, 2005). Examples include air traffic control, emergency medical response, and power plants, where operators must manage a high volume of information and make quick adjustments to meet situational demands (Malakis & Kontogiannis, 2023). In these systems, teamwork is not only essential for efficiency and accuracy but also for safety, as failures in coordination can have significant consequences (Waterson et al., 2015). In one instance, an unmanned aerial vehicle (UAV) crashed into the ground, with the accident later attributed to a lack of coordination between the operators handling the UAV (Williams, 2006).

Therefore, understanding what factors affect teamwork, and how this can be supported through real-time display design, is an important human factors topic. Working within complex systems presents unique challenges for teams that may not arise in simpler task environments. The dynamic nature of these systems demands rapid decision-making under time constraints and high cognitive load can make it difficult for team members to fully communicate their intentions or understand those of others in real-time (Stanton et al., 2013).

However, it is still not clear how best to analyze the attention allocation of people working in teams (Atweh et al., 2022). There is a need for quantitative measures that can be used in real-time and at a fine-grained level of analysis. To better understand and support teamwork in complex systems, researchers have turned to eye tracking as a method for capturing real-time data on attention and focus. Eye tracking allows researchers to observe where team members are directing their gaze, offering insights into team coordination. By analyzing gaze patterns, researchers can gain a better understanding of how team members align their attention, especially during high-stakes or high-workload moments (Atweh & Riggs, 2025; Lobo et al., 2016; Moacdieh et al., 2020). Eye tracking data can thus serve as a valuable, unobtrusive, and real-time measure of team cognition, helping to reveal the subtleties of workload distribution and transitions within complex systems. While the field is still developing, eye tracking shows promise as a tool for both research and practical applications in supporting effective team collaboration in complex systems (Hirshfield et al., 2023).

1.2 Motivation and Research Questions

The motivation behind this dissertation is driven by the pressing need to enhance team performance in complex, high-stakes environments, such as those encountered in UAV command and control (C2) operations (Lu et al., 2021; Mangaroska et al., 2022). Teams working in such environments are required to process large amounts of information quickly, make real-time decisions under pressure, and recover from interruptions (Papamitsiou et al., 2020). While significant strides have been made in understanding team dynamics, critical gaps remain in how we quantify collaboration, identify performance breakdowns, and optimize team performance (Avvenuti & Vecchio, 2009; Chen et al., 2021).

Eye tracking has emerged as a valuable tool for quantifying cognitive and attentional processes in real-time, offering insights into team coordination strategies. By analyzing gaze behaviors, researchers can assess how teams align their visual attention, particularly under varying workload conditions. Gaze-based measures such as scanpath similarity metrics (e.g., ScanMatch, MultiMatch, and Multidimensional Cross-Recurrence Quantification Analysis) could provide a means of evaluating how well teammates track shared visual information. However, it remains unclear how these metrics change with workload demands and whether they can serve as reliable predictors of team performance breakdowns. Addressing this gap is critical for advancing real-time assessment techniques and designing adaptive systems that enhance team effectiveness.

Beyond detection, countermeasures are needed to mitigate coordination failures in high-stakes environments (Fall et al., 2018; Myers et al., 2019). One promising approach is the integration of *Gaze Sharing*, which allows teammates to view each other's eye movements in real-time. Gaze sharing has been proposed as a method to reduce coordination costs in task-specific domains. However, the impact of different gaze sharing visualizations (e.g., dot, trail) on team performance, workload, and communication dynamics in more complex systems remains an open question. Additionally, while gaze sharing *might* serve as a communication aid, its interaction with verbal communication—whether as a complement or replacement—requires further investigation to ensure its effective integration into collaborative systems.

Furthermore, dynamic environments like UAV C2 operations are often disrupted by interruptions, requiring operators to switch between tasks and later reorient themselves to the primary mission. Such disruptions pose significant challenges for maintaining shared awareness and team coordination. A key question is whether user-controlled gaze sharing (i.e., the ability to toggle gaze sharing on or off) can improve team performance by allowing flexibility in when and how visual attention is shared. Examining how gaze sharing influences interruption recovery and whether its effectiveness differs based on task complexity is essential for determining its utility in real-world applications.

To address these challenges, this dissertation explores three central research aims:

- **Aim 1:** Understand how eye tracking can be leveraged to quantify team collaboration and identify team performance breakdowns in UAV C2 tasks.
  - RQ 1.1: How do scanpath similarity metrics (e.g., ScanMatch, MultiMatch, Multidimensional Cross-Recurrence Quantification Analysis) change as workload increases in UAV C2 tasks?
  - RQ 1.2: How do scanpath similarity metrics correlate with team performance measures (e.g., team score, response time) across different workload conditions, and can these correlations help identify performance breakdowns?
- **Aim 2:** Investigate how different gaze sharing displays (dot, trail, no gaze sharing) influence team performance, workload, and communication dynamics in complex systems.

- RQ 2.1: How does gaze sharing influence team collaboration in more complex systems, such as UAV C2 operations?
- RQ 2.2: How do different gaze sharing visualization techniques (dot, trail, no gaze sharing) affect team scanning techniques, situation awareness, workload, and performance?
- RQ 2.3: How do verbal and non-verbal communication techniques (e.g., gaze sharing) interact in UAV C2 teams, and under what circumstances do teams perceive one technique as a replacement for or a complement to another?

- **Aim 3:** Examine how user-controlled gaze sharing (via an on/off toggle) influences team collaboration and performance in UAV C2 operations, particularly in the context of frequent interruptions and varying task complexity.
  - RQ 3.1: How does user-controlled gaze sharing affect team performance compared to continuous gaze sharing and no gaze sharing displays?
  - RQ 3.2: How does gaze sharing influence teams' ability to recover from interruptions, and does its effect differ based on task complexity (simple vs. complex tasks)?

1.3 Summary of Chapters

Chapter 2 presents the background and foundational topics relevant to this dissertation. This dissertation is composed of four studies that address the aims and research questions outlined in Section 1.2. Each study explores different facets of team collaboration and performance, with a focus on real-time metrics and novel display technologies in UAV C2 environments.

- Study 1: The first study addresses Aim 1 and is presented in Chapter 3, which focuses on the use of eye tracking to quantify team collaboration and identify performance breakdowns in dynamic UAV C2 environments. By employing eye tracking techniques, this study aims to measure how teams coordinate their attention, and how these patterns may signal performance issues in high-stakes tasks.

- Studies 2 and 3: The second and third studies, presented in Chapters 4 and 5, collectively address Aim 2, which explores the impact of gaze sharing displays on team communication, workload, and performance. Chapter 4 answers RQs 2.1 and 2.2 and Chapter 5 answers RQ 2.3. These studies investigate the design and effects of gaze sharing visualizations (such as fixation dots and trails) on how teams manage data overload, adapt to changing conditions, and whether such displays can complement or even replace verbal communication in time-critical decision-making.
- Study 4: The fourth study tackles Aim 3 and is presented in Chapter 6, which examines the impact of user-controlled access to gaze sharing on team collaboration and performance, particularly in the context of frequent interruptions. This study focuses on understanding whether providing team members with the ability to toggle gaze sharing on and off can improve their ability to manage interruptions and maintain effective communication in high-pressure environments.

Chapter 7 concludes the dissertation by summarizing its key findings, discussing its intellectual contributions to the academic literature, and reflecting on its broader implications for society.

1.4 References

Atweh, J. A., Moacdieh, N. M., & Riggs, S. L. (2022). Identifying Individual-, Team-, and Organizational-Level Factors that Affect Team Performance in Complex Domains

Based on Recent Literature. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1795–1799. https://doi.org/10.1177/1071181322661213

Atweh, J. A., & Riggs, S. L. (2025). Predicting Mental Demand of Teammates Using Eye Tracking Metrics: A Machine Learning Approach. *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*. https://doi.org/10.1145/3715669.3726802

Avvenuti, M., & Vecchio, A. (2009). Mobile Visual Access To Legacy Voice-Based Applications. *Proceedings Of The 6th International Conference On Mobile Technology, Application & Systems*, 1–6. https://doi.org/10.1145/1710035.1710097

Chen, L., Chen, P., Zhao, S., Luo, Z., Chen, W., Pei, Y., & Yin, E. (2021). Adaptive Asynchronous Control System Of Robotic Arm Based On Augmented Reality-Assisted Brain–Computer Interface. *Journal Of Neural Engineering*, *18*(6), 066005. https://doi.org/10.1088/1741-2552/ac3044

Fall, C. L., Quevillon, F., Blouin, M., Latour, S., Campeau-Lecours, A., Gosselin, C., & Gosselin, B. (2018). A Multimodal Adaptive Wireless Control Interface For People With Upper-Body Disabilities. *IEEE Transactions On Biomedical Circuits And Systems*, *12*(3), 564–575. https://doi.org/10.1109/TBCAS.2018.2810256

Hirshfield, L. M., Wickens, C., Doherty, E., Spencer, C., Williams, T., & Hayne, L. (2023). Toward Workload-Based Adaptive Automation: The Utility of fNIRS for Measuring Load in Multiple Resources in the Brain. *International Journal of Human–Computer Interaction*, 1–27. https://doi.org/10.1080/10447318.2023.2266242

Hollnagel, E., & Woods, D. D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. CRC Press.

Lobo, J. L., Del Ser, J., De Simone, F., Presta, R., Collina, S., & Moravek, Z. (2016). Cognitive Workload Classification Using Eye-Tracking and EEG Data. *Proceedings of the International Conference on Human-Computer Interaction in Aerospace*. https://doi.org/10.1145/2950112.2964585

Lu, Y., Adrados, J. S., Chand, S. S., & Wang, L. (2021). Humans Are Not Machines—Anthropocentric Human–Machine Symbiosis For Ultra-Flexible Smart Manufacturing. *Engineering*, *7*(6), 734–737. https://doi.org/10.1016/j.eng.2020.09.018

Malakis, S., & Kontogiannis, T. (2023). Team Adaptation and Safety in Aviation. *Safety Science*, *158*, 105985. https://doi.org/10.1016/j.ssci.2022.105985

Mangaroska, K., Sharma, K., Gašević, D., & Giannakos, M. (2022). Exploring Students' Cognitive And Affective States During Problem Solving Through Multimodal Data: Lessons Learned From A Programming Activity. *Journal Of Computer Assisted Learning*, *38*(1), 40–59. https://doi.org/10.1111/jcal.12590

Moacdieh, N. M., Devlin, S. P., Jundi, H., & Riggs, S. L. (2020). Effects of Workload and Workload Transitions on Attention Allocation in a Dual-Task Environment: Evidence From Eye Tracking Metrics. *Journal of Cognitive Engineering and Decision Making*, *14*, 132–151. https://doi.org/10.1177/1555343419892184

Myers, C. M., Furqan, A., & Zhu, J. (2019). The Impact Of User Characteristics And Preferences On Performance With An Unfamiliar Voice User Interface. *Proceedings Of The 2019 CHI Conference On Human Factors In Computing Systems*, 1–9. https://doi.org/10.1145/3290605.3300277

Papamitsiou, Z., Pappas, I. O., Sharma, K., & Giannakos, M. N. (2020). Utilizing Multimodal Data Through FsQCA To Explain Engagement In Adaptive Learning. *IEEE Transactions On Learning Technologies*, *13*(4), 689–703. https://doi.org/10.1109/TLT.2020.3020499

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an Understanding of Team Performance and Training. *Teams: Their training and performance*, *3*(1), 3–29. https://www.jstor.org/stable/24531239

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Human Factors: The Journal of the Hu-*

man Factors and Ergonomics Society, *50*, 540–547. https://doi.org/10.1518/001872008X288457

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2013). *Human Factors Methods: A Practical Guide for Engineering and Design* (2nd). CRC Press. https://doi.org/10.1201/9781315587394

Waterson, P., Robertson, M. M., Cooke, N. J., Militello, L., Roth, E., & Stanton, N. A. (2015). Defining the Methodological Challenges and Opportunities for an Effective Science of Sociotechnical Systems and Safety. *Ergonomics*, *58*(4), 565–599. https://doi.org/10.1080/00140139.2015.1015622

Williams, K. (2006). Human Factors Implications of Unmanned Aircraft Accidents: Flight-Control Problems. Elsevier. https://doi.org/10.1016/S1479-3601(05)07008-6

Chapter 2

Background

2.1 Current State of UAV C2 Research

2.1.1 Overview of UAV C2 Operations

Unmanned Aerial Vehicles (UAVs), commonly known as drones, are remotely piloted or autonomously operated aircraft used in various applications, including military operations, disaster response, and commercial logistics (Hildmann & Kovacs, 2019). The versatility of UAVs, combined with their ability to perform tasks in environments that are hazardous or inaccessible to humans, has driven their widespread adoption (Ateş et al., 2022). Central to the effective operation of UAVs is the concept of Command and Control (C2), which refers to the systems, processes, and human operators responsible for managing and directing UAV operations. C2 systems form the backbone of UAV operations, ensuring seamless communication between operators and UAVs for effective mission execution. A typical UAV C2 system comprises a Ground Control Station (GCS), communication links, and onboard systems, all designed to enable the operator to plan, monitor, and adjust mission parameters in real-time (Harinarayana et al., 2024).

UAV C2 encompasses the complete range of activities related to the control of the vehicle's flight path, navigation, payload management, and mission execution (Debnath et al., 2024; Kwak & Sung, 2018). These tasks often involve a complex interplay between human operators and automated systems, requiring robust coordination and communication mechanisms. In military and high-risk environments, UAV C2 systems are tasked with maintaining operational effectiveness under adverse conditions, such as contested airspace or electronic warfare scenarios (Benhassine et al., 2024). This requires resilience through redundant systems and robust fail-safes. Modern UAV C2 designs also incorporate autonomous capabilities, allowing UAVs to execute tasks such as target tracking or waypoint navigation with minimal human intervention, which significantly reduces the cognitive load on operators (Debnath et al., 2024).

Despite the advancements, challenges such as cybersecurity, real-time data processing, and maintaining connectivity in degraded environments persist. Moreover, the increasing use of UAVs in civilian airspace raises regulatory and safety concerns that necessitate further research into standardized and reliable C2 protocols (Karam et al., 2024). These issues highlight the ongoing need for innovation in UAV C2 systems to meet the demands of complex and dynamic operational domains.

2.1.2 Human Factors Research in UAV C2

Research in UAV C2 is diverse, spanning multiple domains including technical system design, automation, communication networks, and human factors. The primary focus in the technical research domain has been on improving the reliability and performance of the communication links between UAVs and their GSCs. Recent advancements in UAV C2 systems have been driven by technological innovations such as improved sensors, enhanced communication links, and sophisticated data processing capabilities. These advancements have expanded the operational scope and efficiency of UAVs.

Beyond the technical aspects, human factors research plays a pivotal role in ensuring that the systems are designed with the operators' needs and limitations in mind. Human factors research in this domain can be broadly categorized into physical and cognitive aspects. Physical ergonomics in UAV C2 research primarily focuses on the ergonomics of control stations, including the layout of controls, seating arrangements, and the design of input devices. These studies aim to optimize the physical interaction between operators and the UAV C2 systems, minimizing fatigue and discomfort during extended operations.

Studies demonstrate that poorly designed workstations can increase physical strain and negatively impact performance, particularly in scenarios requiring sustained attention and precision (Arrabito et al., 2010). One key area of research involves the spatial organization of controls and displays within the GCS. Zhao et al. (2023) found that intuitive layouts, which align with natural body movements and minimize repetitive strain, contribute significantly to reducing operator fatigue and errors. Seating ergonomics, particularly for long-endurance missions, has also been a focal point, with designs emphasizing lumbar support and adjustable configurations to accommodate diverse body types (Arnold, 2016).

While physical ergonomics is a vital component of UAV C2 research, it is often overshadowed by cognitive engineering research due to the high cognitive demands of UAV operations. The physical design of the GCS is intricately linked to cognitive performance, as discomfort and fatigue can amplify cognitive load and degrade decision-making capabilities (Golightly et al., 2020). Cognitive engineering research in UAV C2 operations aims to address the complexities of human-machine interaction by designing systems that support effective decision-making, reduce operator workload, and enhance overall mission performance. UAV C2 systems often involve intricate tasks such as multi-UAV coordination, real-time data interpretation, and dynamic re-planning under time constraints. The cognitive demands placed on operators are substantial, necessitating a systematic approach to display design, workload management, and error mitigation (Tuncal, 2024).

Managing cognitive workload is a cornerstone of UAV C2 research. Zhang et al. (2024) investigated how interface designs impact operator workload in UAV control systems. Their study demonstrated that adaptive displays, which modify information presentation based on real-time assessments of cognitive load, significantly enhance operator performance. For instance, when an operator's workload exceeds optimal levels, the interface can prioritize critical information and suppress non-essential details.

Decision-support systems (DSS) play a critical role in reducing the cognitive burden of UAV operators, particularly in time-sensitive or high-stakes situations. Lim et al. (2018) explored the integration of AI-driven decision aids within UAV C2 systems. Their research demonstrated that DSS tools enhance situation awareness by aggregating and analyzing data from multiple UAVs, presenting operators with actionable insights rather than raw data. For example, AI algorithms can identify potential threats, recommend optimal flight paths, or prioritize mission objectives, allowing operators to focus on strategic decision-making. This research underscores the importance of decision-support systems in improving both efficiency and accuracy in UAV operations.

The complexities of managing multiple UAVs simultaneously have spurred research into collaborative human-machine displays. Donath et al. (2010) proposed a cognitive assistant system designed to support operators in multi-UAV environments. Their system leverages human behavior models to anticipate operator needs, providing preemptive recommendations and streamlining decision-making processes. This proactive approach minimizes the cognitive load associated with managing multiple UAVs, enabling operators to allocate their attention more effectively across tasks. The study further highlights the importance of designing displays that promote intuitive interaction and seamless collaboration between humans and machines.

While significant progress has been made in understanding how humans interact with these complex systems, much of the existing work focuses on single-operator models or human-machine dyads, with limited emphasis on the dynamics of team-based operations. This is a critical gap, as UAV operations often occur in collaborative environments where multiple operators work together within centralized UAV command centers. These teams must coordinate not only with each other but also with automated systems, creating layers of complexity that extend beyond individual interactions. Donath et al. (2010) highlighted the challenges of managing multi-UAV environments, yet their approach predominantly addressed individual operator assistance rather than team-level dynamics.

Team-based UAV C2 operations introduce unique challenges, including the need for shared situation awareness, effective communication, and coordinated decision-making (Katna et al., 2025). Since team research in complex systems like UAV C2 operations is underexplored, we conducted a systematic review of the existing literature on team dynamics and performance in such systems. This review aims to identify gaps, synthesize findings, and provide a foundation for understanding how to enhance teamwork in these environments. The findings from this review, including recommendations for where research should focus, are discussed in the following section (Section 2.2).

2.2 Factors that Affect Team Performance

We conducted a systematic review of the team literature from the past five years to identify key factors affecting team performance (Atweh et al., 2022). The review highlighted several individual-level, team-level, and organizational-level factors that are currently of interest in the research community. Here we provide the results of the systematic literature review (i.e., the main factors identified at each level); however, more details on the methodology can be found in the full published paper (Atweh et al., 2022).

2.2.1 Individual-Level Factors

Emotions and Attitudes

Based on our systematic review of recent literature, emotions and attitudes are significant factors affecting team performance. Research highlights the influence of positive emotions and attitudes on team dynamics. C. P. Lin et al. (2017) illustrated that a positive team affective tone is positively associated with team performance. Meneghel et al. (2016) demonstrated that the dissemination of positive emotions among team members enhances their ability to think broadly and develop positive meanings amidst challenges. Furthermore, C. Lee and Wong (2017) found that increased emotional intelligence among teammates led to reduced task conflicts, and more recent studies, such as Michinov and Michinov (2020), have confirmed that higher average levels of individual emotional intelligence correlate with better team performance. Moreover, effective communication of these emotions alongside constructive feedback can positively impact team effectiveness, as noted by Momeny and Gourgues (2019).

Situation Awareness

Situation awareness (SA)—the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future—is crucial for team performance (Endsley, 1988). Hamilton et al. (2017) suggested that individual SA is essential for task completion, though there are times when aspects of SA should be minimized to regulate behavior effectively. McNeese et al. (2017) examined SA, but with the presence or absence of a synthetic teammate as part of a group of three. They found that synthetic teams performed as well at the mission levels as all-human teams but processed targets less efficiently. The results reflected the weakness of the synthetic teammate when it comes to team SA and coordination strategies.

Our review highlights the need to further explore SA across different domains. While current findings provide insights into SA in various settings, more research is necessary to understand how these dynamics apply to different contexts, such as aviation or healthcare. Section 2.4 provides more details on individual and team SA.

Workload and Stress

The relationship between stress and workload is well-documented. Higher physical and mental workloads can lead to increased stress and reduced performance in larger teams as compared to smaller teams, and this subsequently results in more mistakes made by the team and deterioration in the team's overall performance (Galdikiene et al., 2016). Within healthcare, Sonoda et al. (2018) demonstrated that mental stress among circulating nurses impacted their sense of teamwork performance. Our review underscores the importance of understanding these factors in different domains to determine whether findings are generalizable or specific to certain environments.

2.2.2 Team-Level Factors

Communication and Coordination

Effective communication and coordination are critical for team performance, especially in complex and dynamic environments. Research by Daggett et al. (2016) showed that teams excelling in face-to-face verbal communication and system interaction performed better in information discovery tasks. D'Angelo and Begel (2017) confirmed that teams using verbal and gestural references improved coordination and performance. However, Hoogeboom and Wilderom (2019) found that excessive team communication could hinder information sharing in non-routine contexts.

Our review reflects a growing interest in effective communication strategies and coordination structures, particularly in fields requiring rapid responses, such as emergency landings or surgery. Training may be necessary to optimize knowledge sharing and communication approaches (Olaisen & Revang, 2017).

Knowledge and Expertise Sharing

Knowledge sharing among team members is crucial for enhancing performance and fostering innovation. Jamshed et al. (2018) and Olaisen and Revang (2017) highlighted the positive impact of expertise sharing on team collaboration and future innovations. However, there is a need for systematic examination of how to best share information and integrate diverse knowledge to maximize team performance across various domains.

2.2.3 Organizational-Level Factors

Diversity

Diversity in teams, including expertise, background, and gender, has been shown to positively influence team performance and innovation. Kim et al. (2017) found a correlation between team diversity and collective intelligence. Garcia Martinez et al. (2017) also examined diversity within the team structure, and their findings suggest that diversity is a valuable strategy for an organization to pursue as it provides greater cognitive ability. However, their research highlights that high diversity in gender or skills can also deteriorate performance. Our review emphasizes the need for a balanced approach to diversity, considering both its benefits and potential drawbacks in team performance.

Leadership

Recent studies suggest that teams usually employ structural elements to guide or coordinate their work. For instance, managers are likely to elect a leader to monitor individual team member work, divide collective work among team members, and set rules or procedures for teamwork, including deadlines for tasks (Meyer et al., 2015, 2016). Research shows a strong connection between different leadership and team performance. Chiniara and Bentein (2018) demonstrated that servant leadership, which prioritizes the well-being of employees, improves team performance. Gyanchandani (2017) illustrated that transformational leaders who share their vision and ideas enhance team effectiveness and satisfaction.

Team Size

In the literature, participants have typically been recruited and divided into teams of different sizes, depending on the domain. Teamwork could involve as little as two (D'Angelo & Begel, 2017; Devlin et al., 2020) or three (McNeese et al., 2017) people, and as many as 40 (Garcia Martinez et al., 2017). Salas et al. (2017) found that team size affects performance, with larger teams sometimes experiencing decreased efficiency. Our review underscores the critical role of team size in influencing performance across various contexts, highlighting the need for further research to identify optimal team size configurations. Such insights are essential for designing teams that maximize efficiency, foster collaboration, and adapt to the unique demands of specific environments.

2.2.4 Summary

The literature review provides a nuanced understanding of factors affecting team performance across individual, team, and organizational levels. The findings reveal a broad interest in studying team performance in various domains, with specific insights into the impact of emotions, SA, workload, communication, and diversity. The integration of these factors into Mickan and Rodger (2000)'s framework underscores the need for continued research across different contexts to enhance our understanding of team dynamics and performance.

The next three sections of Chapter 2 present the background on relevant topics from the review and of importance to this dissertation such as shared mental models (section 2.3), situation awareness (section 2.4), and eye tracking (section 2.5).

2.3 Team Cognition and Shared Mental Models

2.3.1 Team Cognition

Team cognition refers to the mental processes and shared understanding that occur within a team to facilitate effective collaboration and decision-making (Hutchins & Kendall, 2011). It encompasses how team members think together, interpret information, and work towards common goals. Team cognition, much like individual cognition, encompasses processes such as learning, planning, reasoning, decision-making, problem-solving, memory, and assessing situations. However, while individual cognitive processes are often internal and less observable, team cognition is typically externalized through interactions and behaviors, making it more readily observable.

In high-stakes environments, effective team cognition is critical for successful outcomes. According to Wildman et al. (2014), team cognition involves the collective processes of team members' knowledge and thinking patterns that support the execution of team tasks. These processes include shared understanding of goals, strategies, and roles, which are crucial for coordinating complex activities. Good team cognition enables teams to adapt to changing circumstances, make informed decisions, and respond effectively to unforeseen events.

Moreover, Cooke et al. (2013) hypothesize that team cognition is a dynamic, emergent activity that cannot be reduced to any single team member or even to the collective shared cognition of the team members. Instead, it emerges from the interactions within the team as it responds to a complex and ever-changing environment. This perspective emphasizes the role of the team as a functional unit, distinct from the sum of its individual parts, in processing information, coordinating, and performing effectively.

In their framework, team cognition is viewed as a process rather than a state, emphasizing the dynamic and adaptive nature of team interactions (i.e., Interactive Team Cognition; Cooke et al., 2013). It is not simply about static shared knowledge or mental models but about how teams actively and continuously integrate and adapt their cognitive resources in response to the demands of their environment. This perspective is particularly relevant in sociotechnical systems, where teams interact with complex systems and technologies, such as in UAV operations. Here, the fluid and emergent nature of team cognition plays a crucial role in determining team performance and effectiveness.

Understanding team cognition through the lens of interactive team cognition shifts the focus from individual team members to the team as a whole, providing insights into how collective cognitive activities emerge and function. This approach is critical for improving team performance in complex, dynamic environments, where effective coordination and adaptive behavior are paramount.

Research on team cognition has highlighted several key aspects that influence team performance. Chou et al. (2012) emphasized that team cognition is closely linked to team performance and effectiveness. They noted that teams with well-developed cognitive structures are better at handling dynamic tasks and achieving higher levels of performance. Similarly, Chen and Kanfer (2024) explored how cognitive and motivational processes within teams impact their ability to perform complex tasks. Their findings suggest that team cognition involves both shared knowledge and interdependent cognitive processes that enhance team effectiveness.

2.3.2 Shared Mental Models

Shared mental models (SMMs) are cognitive representations that team members hold about the team's tasks, goals, and roles (Bierhals et al., 2007). SMMs are crucial for effective team decision-making and performance in various fields, including healthcare and construction, and represent overlapping knowledge and assumptions among team members (Gisick et al., 2018). They can be categorized into task-specific, task-related, teammate knowledge, and attitudes/beliefs, which fall under task-work and team-work domains (Yusoff & Salim, 2020). In healthcare, SMMs enhance clinical competency committee decisions and resident assessments (Edgar et al., 2021). Measuring SMMs involves qualitative and quantitative methods, with cognitive task analysis being a specific evaluation technique (Gisick et al., 2018; Wu et al., 2023). Developing and maintaining SMMs requires clear communication, shared understanding of goals, and well-defined expectations among team members (Edgar et al., 2021).

SMMs can be categorized into different types, including task-specific models and team-specific models. Task-specific mental models relate to the understanding of the task at hand, such as the procedures and strategies required to complete it. Team-specific mental models involve knowledge about team members' roles, skills, and preferences (Cannon-Bowers et al., 1993).

SMMs contribute to effective teamwork by providing a common framework for understanding and interacting with each other. Cannon-Bowers et al. (1993) highlighted that SMMs enhance team performance by improving communication, reducing misunderstandings, and facilitating coordination. Similarly, Mohammed et al. (2010) found that SMMs are associated with better team performance and higher levels of team cohesion. They argue that when team members share a common understanding of their goals and roles, they are better able to synchronize their efforts and achieve successful outcomes.

The development and maintenance of SMMs involve continuous communication and interaction among team members. Teams develop SMMs through collaborative experiences, feedback, and iterative adjustments to their understanding of tasks and roles (Klimoski & Mohammed, 1994). Maintenance of these models requires ongoing communication and adaptation to changes in the team's environment or tasks (DeChurch & Mesmer-Magnus, 2010). Effective communication and feedback mechanisms are essential for sustaining SMMs and ensuring that all team members remain aligned in their understanding. Maintaining SMMs in dynamic and complex environments can be challenging. Marks et al. (2001) indicated that discrepancies in mental models among team members can lead to coordination problems and reduced performance. Teams must be vigilant in updating their mental models to reflect changes in the task or environment to avoid these issues (Floren et al., 2018).

In the context of UAV C2 operations, effective team cognition and SMMs are critical for achieving high performance. UAV C2 operations often involve complex, high-pressure situations where team members must work together seamlessly to manage multiple UAVs and respond to dynamic scenarios. The development and maintenance of SMMs in UAV C2 teams are essential for ensuring that all team members have a common understanding of their roles, objectives, and the operational environment. McNeese et al. (2017) demonstrated that well-developed SMMs contribute to better coordination and performance in UAV C2 environments. By fostering effective team cognition and ensuring that SMMs are maintained, UAV C2 teams can improve their SA, decision-making, and overall operational effectiveness.

2.4 Individual, Team, and Shared Situation Awareness

Situation awareness (SA) is defined as the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley, 1988). Measuring SA has been found to provide valuable insights into the cognitive state of people while they are performing a wide variety of tasks (Endsley, 1988). Researchers have attempted to measure SA in several ways (Endlsey, 2021):

a) *Process measures.* These include communication techniques, verbal protocols, workload measures (e.g., NASA-TLX), and physiological measures such as eye tracking that are continuous in nature. Process measures can provide information on processes, strategies, and types of assessments made. However, little research to date exists to support the validity of physiological measures in assessing SA (Endlsey, 2021).

b) *Performance measures.* Performance measures include as response time and accuracy can usually be gathered without input from the operator; however, these measures can incorrectly correlate SA with performance, which can be affected by other factors such as sensitivity (i.e., differentiating between cues or stimuli) and diagnosticity (i.e., determining what cues, stimuli, information means about the state of a task or system).

c) *Subjective measures.* Subjective measures require the participants to self-rate their level of SA. Likert scales or SART (Situation Awareness Rating Technique; Taylor, 1990) are among the few direct measures of SA. Subjective measures are typically easy to collect and can be used in various domains; however, participants might be biased in their self-assessments of performance using SART (i.e., over/under confidence).

d) *Objective measures.* These measures require participants to correctly answer predetermined questions about the scenario and include SPAM (Situation Present Assessment

Method) and SAGAT (Situation Awareness Global Assessment Technique; Endsley, 2021). SPAM queries people on relevant SA knowledge of the past, present, and future. Administering SPAM does not require the experimenter to freeze the simulation but adds to the task load and could negatively affect performance. On the other hand, SAGAT directly measures SA by having participants answer questions during short simulation freezes (Endsley, 1995). SAGAT is based on unbiased sampling across scenario and avoids end-of-trial memory dependence (Endlsey & Rodgers, 1994). The use of SAGAT has been adopted in several studies and successfully used in a wide variety of domains including aviation, air traffic control, driving, health care, maritime/military operations, and power systems (Durso et al., 1998; Gardner et al., 2017; Hogan et al., 2006; Matthews & Beal, 2002).

Team Situation Awareness (TSA) refers to the degree to which every team member possesses the SA needed for his or her job. Measuring TSA presents a challenge due to the limited availability of comprehensive assessment methods. While various tools and techniques exist for evaluating individual SA, there has been limited work extending these measures to teams. Current assessment approaches often focus on individual contributions, which may not fully capture the dynamic and interdependent nature of a team. Furthermore, TSA should ideally consider interactions among team members, SMMs, communication dynamics, and collaborative decision-making processes. Amongst the aforementioned measures, SAGAT can assess team SA to determine whether the team exhibits an aggregate understanding of the critical aspects of the scenario (Endlsey, 2021).

Shared Situation Awareness (SSA) refers to the degree to which team members have the same SA on shared SA requirements. Unlike TSA, which evaluates the team as a whole, SSA focuses on the alignment and overlap of individual team members' situation awareness. High SSA is characterized by consistency in understanding critical information, task goals, and environmental changes across team members. This shared understanding enables teams to coordinate effectively, anticipate one another's needs, and adapt dynamically to changing circumstances.

However, researchers have raised challenges to SAGAT regarding the freeze method (de Winter et al., 2019; Flach, 1995). For example, de Winter et al. (2019) outlined six fundamental limitations of SAGAT that include time delays between the freeze moment and query response, task interruption/disruption, and the necessity to bring the situation to conscious memory. Despite the critiques raised, a common quantifiable measure of TSA is using a SAGAT-TSA score (Hultin et al., 2019; Sulistyawati et al., 2008). SAGAT-TSA is calculated using the mean SA score for each team based on the accuracy of their individual SAGAT answers (Endlsey, 2021). An overall TSA score can be tabulated alongside a score for each of the three levels of SA.

2.5 Eye Tracking in Individuals and Teams

Eye tracking technology is increasingly being used to study team performance in complex domains (J. Y. Lee et al., 2020). This technology provides a trace of people's eye movements, which enables researchers to monitor visual attention in real-time (Y. Lin et al., 2004). By tracking a pair's eye movements simultaneously, researchers can gain insights into joint attention and collaboration in complex domains (Damacharla et al., 2019). Eye tracking technology allows researchers to collect objective, quantitative data on how teammates collaborate in real-time. It provides valuable insights into the cognitive processes involved in collaborative problem-solving, including the distribution of attentional resources and the coordination of attention between teammates. This information can be used to identify areas for improvement in team performance and develop strategies for optimizing collaboration in high-stress, data-rich environments.

Eye tracking is the process of measuring and recording the movement of a person's eyes and understanding where they are looking or gazing (Poole & Ball, 2006). Eye tracking has been extensively used in human factors research, with applications in various fields such as aviation (e.g., Mao et al., 2021), driving (e.g., Kapitaniak et al., 2015; Palinko et al., 2010), programming (e.g., Goswami et al., 2016), military (e.g., Shree et al., 2018), website design (e.g., Wang et al., 2014), air traffic control (e.g., Alonso et al., 2013), and medicine (e.g., Harezlak & Kasprowski, 2018).

This technology utilizes devices called eye trackers, which can be desktop-mounted near a display or head-mounted on a person. In either setup, the eye tracker produces raw eye location data referred to as points of regard (POR), indicating where the person is looking on the screen. These PORs can then be used to identify the two fundamental components of eye tracking studies: fixations and saccades. Fixations are moments characterized by a stable location and duration, representing periods during which visual processing occurs (Findlay, 2004; Figure 2.1). The swift eye movements between consecutive fixations are known as saccades, during which visual processing is typically suppressed (Yarbus, 1967). A scanpath is the sequence of fixations and saccades, offering a way to visualize eye movements (Noton & Stark, 1971). Lastly, an area of interest (AOI) is a designated region of the display where eye tracking data analysis is focused. These elements serve as the foundational components of eye tracking research.

From the analysis of fixations and saccades, a substantial body of research has emerged to deepen our understanding of attention, workload, and other cognitive processes. Various metrics have been developed to quantify these eye movements (Skaramagkas et al., 2021). For fixations, researchers often examine the number of fixations and fixation duration (Bylinskii et al., 2015). For saccades, researchers focus on the number of saccades, saccade duration and velocity (Krauzlis, 2013). These metrics, when analyzed together, offer a detailed picture of how attention is allocated, cognitive workload is managed, and visual information is processed. The number and duration of fixations, along with saccade metrics, can indicate where and for how long attention is allocated (Skaramagkas et al., 2021). More fixations or longer durations might suggest increased attentional load or difficulty in processing information (Yang et al., 2013). Higher numbers of fixations, longer fixation durations, and slower saccade velocities can be associated with increased cognitive workload and fatigue in individuals (Škvareková et al., 2020; Tsai et al., 2007). The number of fixations and saccades, along with their durations and amplitudes, are crucial in understanding how individuals scan and search for visual targets. Patterns such as a higher number of fixations in cluttered displays or longer fixation times on relevant items can provide insights into search efficiency and strategies.

Eye tracking technology also allows researchers to study workload variations in real-time, which can help them understand how these transitions impact team performance (Devlin et al., 2020). Studies that use eye tracking to study pair's performance and attention allocation often use gaze coupling/overlap which refers to moments when teammates are looking at the same AOI, a specific area or region on the screen that researchers have identified to be relevant for analysis purposes. Previous work has shown that the coupling of gaze between collaborating partners may improve the quality of interaction and comprehension (Richardson & Dale, 2005), but this is not always the case (Villamor & Rodrigo, 2018). To date, the focus has been on the percentage of gaze overlap (i.e., when both teammates look at the same point) and similarities between teammates' gaze trajectories (i.e., shared patterns in their eye movement paths; (Devlin et al., 2019). While these analyses are needed, it is also important to explore the percentage of identical scanpath segments between two people over time (i.e., portions of eye movement paths that match each other between teammates) and the average duration the teammates are synchronized or aligned, especially during changes in workload (Silva et al., 2015). In complex domains, scanning large amounts of data is a crucial component of team collaboration, and quantitative measures of scanning can provide insights into how teams process and analyze information (Atweh & Riggs, 2025). These

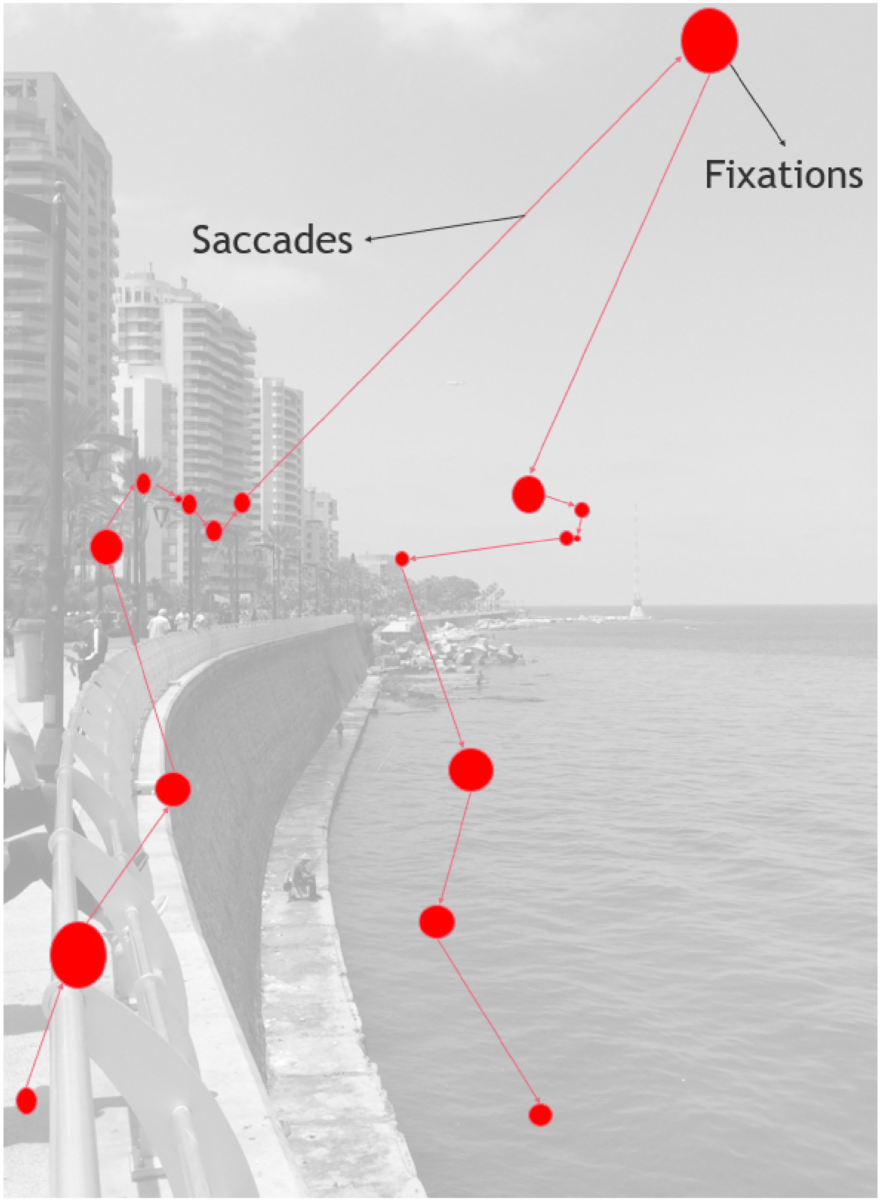


Figure 2.1: Fixations are usually depicted as a circle whose diameter is proportional to fixation duration. Saccades are represented as lines between two successive fixations. All the fixations and saccades together create a scanpath.

measures can indicate how much attention is being paid to specific information and identify potential challenges in the scanning process. By analyzing quantitative data, researchers can identify bottlenecks, inefficiencies, or areas where teams excel.

2.6 Current Gaps in Teams Research

The literature on team dynamics and performance is extensive, yet several critical gaps remain unaddressed. One significant gap is the limited understanding of the real-time processes that underlie effective team collaboration. Much of the existing research relies on retrospective self-reports and post-task debriefs, which can be subject to memory biases and do not capture the fluid, dynamic nature of teamwork. Real-time data collection and analysis methods, such as eye tracking and physiological monitoring, are emerging but are not yet widely integrated into team studies. These methods hold the promise of providing richer, more nuanced insights into how teams coordinate and make decisions in real-time. The role of technology in mediating team interactions also represents a significant gap in the literature. With the increasing reliance on digital tools and remote collaboration, it is essential to understand how these technologies affect team processes and outcomes. Current research has only begun to scratch the surface of how tools like virtual reality, collaborative software, and automated systems impact team communication, coordination, and decision-making. Additionally, there is a need to explore the design of these technologies to support and enhance team performance, especially in high-pressure and dynamic environments. Based on a comprehensive review we conducted (Atweh et al., 2022), we have identified five overarching areas for future research:

2.6.1 The Need for Longitudinal Studies in Team Development

A significant gap in the current literature is the scarcity of longitudinal studies that examine team development over time (Ilgen et al., 2005; Mathieu et al., 2008). Most team research provides snapshots of team performance at single points in time, which limits our ability to understand how teams evolve, adapt, and develop competencies over extended periods (Hackman, 2002; Kozlowski & Bell, 2003). This short-term focus overlooks the dynamic and ongoing nature of team processes, which are influenced by continuous interactions and changing circumstances (Edmondson, 1999; Marks et al., 2001).

Longitudinal studies are crucial for uncovering the processes of team learning, adaptation, and resilience, especially in the face of changing tasks and environments (Burke et al., 2006; Waller et al., 2004). Teams are often required to adapt to new challenges, integrate new members, and refine their strategies over time (Kozlowski & Bell, 2003). Understanding these processes requires observing teams over extended periods, capturing the fluctuations and trends in their performance and interactions (Mathieu et al., 2008). Such research would provide valuable insights into the lifecycle of teams, including how they form, develop, mature, and potentially disband (Tannenbaum & Cerasoli, 2013).

Moreover, longitudinal studies can reveal the factors that contribute to sustained team effectiveness. For instance, they can help identify the practices and conditions that facilitate continuous improvement and learning within teams (Edmondson, 1999; Ilgen et al., 2005). By tracking teams over time, researchers can examine how initial team configurations, leadership styles, and communication patterns influence long-term outcomes (Hackman, 2002; Marks et al., 2001). This perspective can also highlight the critical moments or turning points that significantly impact team trajectories, such as major successes, failures, or changes in team composition (Burke et al., 2006; Mathieu et al., 2008).

Additionally, longitudinal research can inform interventions aimed at supporting team development (Tannenbaum & Cerasoli, 2013). By understanding the natural progression of team dynamics, practitioners can design targeted interventions that address specific developmental stages or challenges (Waller et al., 2004). For example, early interventions might focus on building trust and cohesion, while later interventions could emphasize advanced coordination strategies or conflict resolution skills (Hackman, 2002; Kozlowski & Bell, 2003). Tailoring interventions to the evolving needs of teams can enhance their effectiveness and ensure that they provide timely support (Edmondson, 1999).

Furthermore, longitudinal studies can contribute to the development of theories and models that more accurately reflect the complexity of team dynamics. Current theories often rely on static assumptions and may not fully capture the iterative and emergent nature of team processes. Longitudinal data can provide the empirical foundation needed to refine and expand these theories, leading to a more comprehensive understanding of team behavior (Burke et al., 2006).

2.6.2 How Teams Adapt to Changing Workload Over Time in Complex Environments

Understanding how teams adapt to changing workload conditions, particularly in dynamic and high-stress environments, is critical for improving their performance and resilience. While significant research has been dedicated to individual responses to stress, team-level dynamics, which are essential for collective success, remain underexplored (Deacon, 2020; Singh, 2024). Teams, as integrated systems, must navigate fluctuating demands by effectively coordinating, redistributing tasks, and maintaining communication. The ability to manage workload fluctuations in real-time determines not only the efficiency of the team but also its long-term cohesion and adaptability (Dietz et al., 2017; Entin & Serfaty, 1999).

High workload conditions can significantly degrade performance by inducing errors, reducing situation awareness, and increasing stress levels for both individuals and the team as a whole (Michel et al., 2021; Sonoda et al., 2018). Teams under stress must exhibit adaptability, including reallocating resources, reshaping task assignments, and maintaining effective communication during crises (Golden et al., 2018; Hagemann et al., 2012). Effective adaptation to such conditions requires teams to exhibit flexibility in how they allocate resources, adjust task assignments, and communicate under pressure. Research has shown that communication patterns, particularly those that reinforce mutual understanding and mental models, can mitigate many of the adverse effects of high workload environments (Entin & Serfaty, 1999). Moreover, teams with strong pre-existing cohesion and robust communication strategies are better equipped to sustain performance even during periods of extreme stress (Hagemann et al., 2012; Sorensen et al., 2024).

Task redistribution is a common strategy to manage workload imbalances, alleviating pressure on overloaded team members. However, this approach often introduces its own challenges, such as increased coordination overhead and potential task misallocation (Burke et al., 2018; Mu et al., 2024). Balancing the immediate benefits of redistribution with the risks of disrupting established workflows and expertise is a critical aspect of effective workload management. For instance, while redistribution may improve short-term adaptability, research suggests it can negatively impact team cohesion if not executed carefully, leading to long-term inefficiencies (Sassenus et al., 2024).

In high-stress environments like emergency response or space exploration, real-time adaptation becomes even more critical. Teams operating under these conditions often rely on technological aids, such as predictive analytics and physiological monitoring, to make faster and more accurate decisions (Dietz et al., 2017; Michel et al., 2021). These tools not only reduce cognitive overload but also enable teams to identify and address stress points before they escalate into performance issues. Extreme environments, such as polar expeditions or deep-space missions, further illustrate the importance of maintaining team adaptability over time, as sustained exposure to high workload conditions requires resilience and ongoing recalibration of team dynamics (Vessey & Landon, 2017).

The implications of understanding how teams adapt to changing workload conditions are profound. By examining team dynamics longitudinally, researchers can identify patterns and interventions that support adaptability and resilience. Tailored training programs that emphasize task redistribution, communication under duress, and cohesion-building strategies are essential for teams in high-stress environments. Future research must delve deeper into these dynamics, integrating insights from psychology, organizational science, and computational modeling to enhance our understanding of team adaptation over time (Deacon, 2020; Wickens & Huey, 1993). Such efforts will provide invaluable guidance for designing teams and systems that thrive under the most demanding conditions.

2.6.3 How to Build Effective Teams and Leaders

A persistent research challenge is determining the best strategies for building effective teams, from their initial formation to the overall structure of larger teams (Martin Jr, 1997). Questions remain about whether it is more beneficial for experts to work with other experts or novices, and how experts should share their knowledge (Jamshed et al., 2018). Teams composed of experts may excel in performance due to their high level of skill and experience. However, such teams may also face challenges, such as dominance of certain members and reduced opportunities for learning and innovation. Conversely, mixed teams of experts and novices can benefit from diverse perspectives and learning opportunities, but they may also struggle with coordination and knowledge transfer (Coakes et al., 2008).

Current literature provides limited guidance on structuring teams based on task requirements, geographical location, and communication strategies. Future research should address these gaps to develop a more nuanced understanding of how to form and structure teams for optimal performance (Larson et al., 2023). For instance, research could explore the impact of team size and composition on performance across different types of tasks. Additionally, studies could examine how remote and co-located teams differ in their dynamics and outcomes, particularly in terms of communication and collaboration.

Effective team building also involves establishing clear roles, responsibilities, and goals. Research should investigate the best practices for defining and communicating these elements to ensure that all team members are aligned and working towards a common objective.

Furthermore, understanding how different team structures and communication strategies affect performance can provide valuable insights for practitioners. For example, research could explore the effectiveness of different leadership styles, decision-making processes, and conflict resolution techniques in various team contexts.

Moreover, implementing training programs for both team members and leaders on understanding and managing emotions is crucial. Such training would enhance their ability to interact effectively during complex, time-sensitive tasks (Van der Hoek et al., 2018). Training programs should focus on developing skills such as empathy, active listening, and emotional regulation. These skills can help team members understand and support each other, particularly in high-stress situations. Additionally, training should address how to handle negative emotions and conflicts constructively, ensuring that they do not undermine team performance.

2.6.4 Quantifying Team Collaboration Using Real-Time Metrics

The ability to quantify team collaboration has become increasingly critical as organizations and research efforts seek to enhance team performance in complex, dynamic environments. Despite substantial progress in understanding team dynamics, a significant gap persists in developing robust, real-time metrics that accurately capture the multifaceted nature of collaboration. Historically, studies have relied on qualitative methods, such as self-reports and observational data, which, while valuable, are subjective and limited in scope (Steitz et al., 2020).

Team collaboration is inherently complex, encompassing processes like communication, coordination, decision-making, and mutual support. Effective collaboration involves not only the frequency and clarity of communication but also nuanced factors like responsiveness, adaptability, and shared understanding among team members (Burnett & Lisk, 2021;

Wiltshire et al., 2024). Researchers have emphasized the need for integrated frameworks that combine multiple data sources, such as task performance metrics, physiological indicators, and communication patterns, to create a comprehensive picture of team collaboration (Damacharla et al., 2018; Walton & Gilbert, 2022).

Emerging technologies have paved the way for innovative approaches to quantifying collaboration. For instance, real-time eye-tracking data provide insights into team members' shared attention and visual coordination, while physiological monitoring can measure stress levels and cognitive load (Budacu & Pocatilu, 2018; Leshed et al., 2009). These methods have demonstrated potential for capturing critical aspects of collaboration, but they remain underdeveloped and require further validation to become widely adopted (Gorman et al., 2020; Škec et al., 2017).

Additionally, advances in human-machine teaming have introduced new dimensions to collaborative metrics. Tools such as AI-driven analytics and natural language processing can analyze communication patterns in real-time, offering actionable feedback to enhance collaboration (Burnett & Lisk, 2021; Damacharla et al., 2018). Such technologies also facilitate the integration of qualitative and quantitative data, allowing teams to monitor performance metrics and make informed adjustments dynamically (Steitz et al., 2020; Wiltshire et al., 2024).

Frameworks for evaluating collaboration must address scalability and context-specific challenges. Many existing metrics are tailored to small-scale studies or controlled environments, making their application to larger, more diverse teams difficult (Walton & Gilbert, 2022; Woodcock, 2005). Future efforts should focus on developing scalable, adaptable methods that can be applied across a range of industries and team structures, such as healthcare, engineering, and education (Leifer, 1998; Škec et al., 2017).

Furthermore, real-time collaboration metrics should shift beyond outcome-based measures to process-oriented insights. Metrics that capture the dynamics of team behaviors, such as adaptability during crises or the evolution of trust over time, provide a deeper understanding of the factors driving successful collaboration (Gorman et al., 2020; Leshed et al., 2009). Process-oriented metrics are particularly valuable in identifying areas for intervention and designing targeted strategies to improve team dynamics and outcomes (Burnett & Lisk, 2021; Wiltshire et al., 2024).

In conclusion, the field of team collaboration measurement is evolving rapidly, driven by technological advancements and a growing recognition of the limitations of traditional methods. Real-time metrics that integrate diverse data sources hold the promise of transforming how teams are analyzed and supported, ultimately fostering more effective and adaptable collaboration in both traditional and human-machine teams.

2.7 Conclusion

In this chapter, we explored key gaps in the literature on UAV C2 operations, particularly those related to cognitive processes and team collaboration. We discussed the challenges teams face in UAV environments, such as maintaining SA, SMMs, and effective coordination under dynamic and changing workload conditions. These challenges underscore the critical need to quantify team collaboration and develop real-time metrics to support teamwork in UAV C2 environments.

Eye tracking emerged as a promising approach to address these gaps by providing real-time insights into team dynamics and cognitive processes. Specifically, the ability to monitor and display real-time changes in gaze behavior offers potential to enhance team cognition, SMMs, TSA, and SSA. These metrics can inform the design of adaptive systems that support better collaboration and mitigate performance breakdowns in UAV teams.

To address these needs, the next chapter, Chapter 3, delves into a comprehensive study where eye tracking was employed to quantify team collaboration and identify potential breakdowns. This study utilized scanpath similarity analysis to examine how teams adapt to workload changes, providing actionable insights into enhancing team performance in UAV C2 operations.

2.8 References

Alonso, R., Causse, M., Vachon, F., Parise, R., Dehais, F., & Terrier, P. (2013). Evaluation of Head-Free Eye Tracking as an Input Device for Air Traffic Control. *Ergonomics*, *56*(2), 246–255. https://doi.org/10.1080/00140139.2012.744473

Arnold, K. P. (2016). The UAV Ground Control Station: Types, Components, Safety, Redundancy, and Future Applications. *International Journal of Unmanned Systems Engineering*, *4*(1), 37. https://doi.org/10.14323/ijuseng.2016.4

Arrabito, G. R., Ho, G., Lambert, A., Rutley, M., Keillor, J., Chiu, A., Au, H., & Hou, M. (2010). *Human Factors Issues for Controlling Uninhabited Aerial Vehicles* (tech. rep.). Defence RD Canada.

Ateş, S. S., Uzgör, M., & Yüksek, K. (2022). UAV Tracking Module Proposal Based on a Regulative Comparison Between Manned and Unmanned Aviation. *Journal of Airline and Airport Management*, *12*(1), 29–47. https://doi.org/10.3926/jairm.206

Atweh, J. A., Moacdieh, N. M., & Riggs, S. L. (2022). Identifying Individual-, Team-, and Organizational-Level Factors that Affect Team Performance in Complex Domains Based on Recent Literature. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1795–1799. https://doi.org/10.1177/1071181322661213

Atweh, J. A., & Riggs, S. L. (2025). Predicting Mental Demand of Teammates Using Eye Tracking Metrics: A Machine Learning Approach. *Proceedings of the 2025 Symposium* *on Eye Tracking Research and Applications (ETRA '25)*. https://doi.org/10.1145/3715669.3726802

Benhassine, M., Quinn, J., Stewart, D., Arsov, A. A., Ianc, D., Ivan, M., & Utterbeeck, F. V. (2024). Advancing Military Medical Planning in Large Scale Combat Operations: Insights From Computer Simulation and Experimentation in NATO's Vigorous Warrior Exercise 2024. *Military Medicine*, *189*(Supplement$_3$), 456–464. https://doi.org/10.1093/milmed/usae152

Bierhals, R., Schuster, I., Kohler, P., & Badke-Schaub, P. (2007). Shared Mental Models—Linking Team Cognition and Performance. *CoDesign*, *3*(1), 75–94. https://doi.org/10.1080/15710880601170891

Budacu, E. N., & Pocatilu, P. (2018). Real Time Agile Metrics For Measuring Team Performance. *Informatica Economica*, *22*(4). https://doi.org/10.12948/issn14531305

Burke, C. S., Priest, H. A., & Salas, E. (2018). Stress and Teams: How Stress Affects Decision Making at the Team Level. *Taylor & Francis Handbook*, 276–290. https://www.taylorfrancis.com

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding Team Adaptation: A Conceptual Analysis and Model. *Journal of Applied Psychology*, *91*(6), 1189–1207. https://doi.org/10.1037/0021-9010.91.6.1189

Burnett, J. R., & Lisk, T. C. (2021). The Future Of Employee Engagement: Real-Time Monitoring And Digital Tools For Engaging A Workforce. In *International perspectives on employee engagement* (1st Edition, pp. 117–128). Routledge.

Bylinskii, Z., Borkin, M. A., Kim, N. W., Pfister, H., & Oliva, A. (2015). Eye Fixation Metrics for Large Scale Evaluation and Comparison of Information Visualizations. *IEEE Workshop on Eye Tracking and Visualization*. https://api.semanticscholar.org/CorpusID:18126165

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared Mental Models in Expert Team Decision Making. In J. Castellan (Ed.). Lawrence Erlbaum Associates, Inc.

Chen, G., & Kanfer, R. (2024). The Future of Motivation in and of Teams. *Annual Review of Organizational Psychology and Organizational Behavior*, *11*(1), 93–112. https://doi.org/10.1146/annurev-orgpsych-111821-031621

Chiniara, M., & Bentein, K. (2018). The Servant Leadership Advantage: When Perceiving Low Differentiation in Leader-Member Relationship Quality Influences Team Cohesion, Team Task Performance and Service OCB. *The Leadership Quarterly*, *29*(2), 333–345. https://doi.org/10.1016/j.leaqua.2017.05.002

Chou, H. W., Lin, Y. H., & Chou, S. B. (2012). Team Cognition, Collective Efficacy, and Performance in Strategic Decision-Making Teams. *Social Behavior and Personality: An International Journal*, *40*(3), 381–394. https://doi.org/10.2224/sbp.2012.40.3.381

Coakes, E. W., Coakes, J. M., & Rosenberg, D. (2008). Co-Operative Work Practices And Knowledge Sharing Issues: A Comparison Of Viewpoints. *International Journal Of Information Management*, *28*(1), 12–25. https://doi.org/10.1016/j.ijinfomgt.2007.10.004

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive Team Cognition. *Cognitive Science*, *37*(2), 255–285. https://doi.org/10.1111/cogs.12009

Daggett, M., O'Brien, K., Hurley, M., & Hannon, D. (2016). Predicting Team Performance Through Human Behavioral Sensing and Quantitative Workflow Instrumentation. In *Advances in human factors and system interactions* (pp. 245–258). https://doi.org/10.1007/978-3-319-41956-5_22

Damacharla, P., Javaid, A. Y., Gallimore, J. J., & Devabhaktuni, V. K. (2018). Common Metrics To Benchmark Human-Machine Teams (HMT): A Review. *IEEE Access*, *6*, 38637–38655. https://doi.org/10.1109/ACCESS.2018.2853560

Damacharla, P., Javaid, A. Y., & Devabhaktuni, V. K. (2019). Human Error Prediction Using Eye Tracking to Improvise Team Cohesion in Human-Machine Teams. *Advances in Intelligent Systems and Computing*, *778*, 47–57. https://doi.org/10.1007/978-3-319-94391-6_5

D'Angelo, S., & Begel, A. (2017). Improving Communication Between Pair Programmers Using Shared Gaze Awareness. *Conference on Human Factors in Computing Systems - Proceedings*, 6245–6255. https://doi.org/10.1145/3025453.3025573

de Winter, J. C., Eisma, Y. B., Cabrall, C. D. D., Hancock, P. A., & Stanton, N. A. (2019). Situation Awareness Based on Eye Movements in Relation to the Task Environment. *Cognition, Technology & Work*, *21*(1), 99–111. https://doi.org/10.1007/s10111-018-0527-6

Deacon, A. (2020). Team Adaptation in High Reliability Teams. *University of Calgary Digital Repository*. https://prism.ucalgary.ca/bitstreams/d911aa57-6c56-491f-a924-9bb897b4a37d/content

Debnath, D., Vanegas, F., Boiteau, S., & Gonzalez, F. (2024). An Integrated Geometric Obstacle Avoidance and Genetic Algorithm TSP Model for UAV Path Planning. *Drones*, *8*(7), 302. https://doi.org/10.3390/drones8070302

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The Cognitive Underpinnings of Effective Teamwork: A Meta-Analysis. *Journal of Applied Psychology*, *95*(1), 32.

Devlin, S. P., Flynn, J. R., & Riggs, S. L. (2019). Examining the Visual Attention of Pairs of Operators during a Low to High Workload Change. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 201–205. https://doi.org/10.1177/1071181319631168

Devlin, S. P., Moacdieh, N. M., Wickens, C. D., & Riggs, S. L. (2020). Transitions Between Low and High Levels of Mental Workload can Improve Multitasking Performance. *IISE Transactions on Occupational Ergonomics and Human Factors*, *8*(2), 72–87. https://doi.org/10.1080/24725838.2020.1770898

Dietz, A. S., Driskell, J. E., Sierra, M. J., Weaver, S. J., Driskell, T., & Salas, E. (2017). Teamwork Under Stress. *The Wiley Handbook of Teamwork*, 345–365. https://doi.org/10.1002/9781118909997.ch13

Donath, D., Rauschert, A., & Schulte, A. (2010). Cognitive Assistant System Concept for Multi-UAV Guidance Using Human Operator Behaviour Models. *HUMOUS'10*. https://www.researchgate.net/publication/265756393_Cognitive_assistant_system_concept_for_multi-UAV_guidance_using_human_operator_behaviour_models

Durso, F. T., Truitt, T. R., Hackworth, C. A., Crutchfield, J. M., & Manning, C. A. (1998). En Route Operational Errors and Situational Awareness. *The International Journal of Aviation Psychology*, *8*, 177–194. https://doi.org/10.1207/s15327108ijap0802_6

Edgar, L., Jones Jr, M. D., Harsy, B., Passiment, M., & Hauer, K. E. (2021). Better Decision-Making: Shared Mental Models and the Clinical Competency Committee. *Journal of Graduate Medical Education*, *13*(2s), 51–58. https://doi.org/10.4300/JGME-D-20-00850.1

Edmondson, A. C. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, *44*(2), 350–383. https://doi.org/10.2307/2666999

Endsley, M. R. (1995). Direct Measurement of Situation Awareness in Simulations of Dynamic Systems: Validity and Use of SAGAT. *Experimental analysis and measurement of situation awareness*, 107–113.

Endsley, M. R., & Rodgers, M. (1994, October). *Situation Awareness Information Requirements Analysis for En Route Air Traffic Control* (Vol. 38). https://doi.org/10.1177/154193129403800113

Endsley, M. R. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting*, *32*, 97–101. https://doi.org/10.1177/154193128803200221

Endsley, M. R. (2021). A Systematic Review and Meta-Analysis of Direct Objective Measures of Situation Awareness: A Comparison of SAGAT and SPAM. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *63*, 124–150. https://doi.org/10.1177/0018720819875376

Entin, E. E., & Serfaty, D. (1999). Adaptive Team Coordination. *Human Factors*, *41*, 312–325. https://doi.org/10.1518/001872099779591196

Findlay, J. M. (2004). Eye Scanning and Visual Search. Psychology Press.

Flach, J. M. (1995). Situation Awareness: Proceed with Caution. *Human Factors*, *37*(1), 149–157. https://doi.org/10.1518/001872095779049480

Floren, L. C., Donesky, D., Whitaker, E., Irby, D. M., ten Cate, O., & O'Brien, B. C. (2018). Are We on the Same Page? Shared Mental Models to Support Clinical Teamwork Among Health Professions Learners: A Scoping Review. *Academic Medicine*, *93*(3), 498–509. https://doi.org/10.1097/ACM.0000000000002019

Galdikiene, N., Asikainen, P., Rostila, I., Green, P., Balciunas, S., & Suominen, T. (2016). Organizational Social Context in Primary Health Care. *Nordic Journal of Nursing Research*, *36*(2), 103–111. https://doi.org/10.1177/2057158516628728

Garcia Martinez, M., Zouaghi, F., & Garcia Marco, T. (2017). Diversity is Strategy: The Effect of R&D Team Diversity on Innovative Performance. *R&D Management*, *47*(2), 311–329. https://doi.org/10.1111/radm.12244

Gardner, A. K., Kosemund, M., & Martinez, J. (2017). Examining the Feasibility and Predictive Validity of the SAGAT Tool to Assess Situation Awareness Among Medical Trainees. *Simulation in Healthcare*, *12*(1), 17–21. https://doi.org/10.1097/SIH.0000000000000181

Gisick, L. M., Webster, K. L., Keebler, J. R., Lazzara, E. H., Fouquet, S., Fletcher, K., Fagerlund, A., Lew, V., & Chan, R. (2018). Measuring Shared Mental Models in Healthcare. *Journal of Patient Safety and Risk Management*, *23*(5), 207–219. https://doi.org/10.1177/2516043518796442

Golden, S. J., Chang, C. H., & Kozlowski, S. W. (2018). Teams in Isolated, Confined, and Extreme Environments: A Review. *Journal of Organizational Behavior*, *39*, 750–772. https://doi.org/10.1002/job.2288

Golightly, D., Gamble, C., Palacin, R., & Pierce, K. (2020). Applying Ergonomics Within the Multi-Modelling Paradigm with an Example from Multiple UAV Control. *Ergonomics*, *63*(8), 1027–1043. https://doi.org/10.1080/00140139.2020.1723683

Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2020). Measuring Real-Time Team Cognition During Team Training. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *62*, 825–860. https://doi.org/10.1177/0018720819852791

Goswami, A., Walia, G., Mccourt, M., & Padmanabhan, G. (2016). Using Eye Tracking to Investigate Reading Patterns and Learning Styles of Software Requirement Inspectors to Enhance Inspection Team Outcome. *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. https://doi.org/10.1145/2961111

Gyanchandani, R. (2017). The Effect of Transformational Leadership Style on Team Performance in IT Sector. *IUP Journal of Soft Skills*, *11*(3), 29–44.

Hackman, J. R. (2002). *Leading Teams: Setting the Stage for Great Performances*. Harvard Business School Press.

Hagemann, V., Kluge, A., & Ritzmann, S. (2012). Flexibility Under Complexity: Work Contexts and Team Processes of High-Responsibility Teams. *Employee Relations*, *34*, 475–491. https://doi.org/10.1108/01425451211217734

Hamilton, K., Mancuso, V., Mohammed, S., Tesler, R., & McNeese, M. (2017). Skilled and Unaware: The Interactive Effects of Team Cognition, Team Metacognition, and Task Confidence on Team Performance. *Journal of Cognitive Engineering and Decision Making*, *11*(4), 382–395. https://doi.org/10.1177/1555343417731429

Harezlak, K., & Kasprowski, P. (2018). Application of Eye Tracking in Medicine: A Survey, Research Issues and Challenges. *Computerized Medical Imaging and Graphics*, *65*, 176–190. https://doi.org/10.1016/j.compmedimag.2017.04.006

Harinarayana, T., Krishnan, S. V., & Hota, S. (2024). Lyapunov Guidance Vector Field-Based Waypoint Following by Unmanned Aerial Vehicles. *Journal of Guidance, Control, and Dynamics*, 1–11. https://doi.org/10.2514/1.G008056

Hildmann, H., & Kovacs, E. (2019). Review: Using Unmanned Aerial Vehicles (UAVs) as Mobile Sensing Platforms (MSPs) for Disaster Response, Civil Security and Public Safety. *Drones*, *3*(3), 59. https://doi.org/10.3390/drones3030059

Hogan, M. P., Pace, D. E., Hapgood, J., & Boone, D. C. (2006). Use of Human Patient Simulation and the Situation Awareness Global Assessment Technique in Practical Trauma Skills Assessment. *Journal of Trauma and Acute Care Surgery*, *61*(5), 1047–1052. https://doi.org/10.1097/01.ta.0000238687.23622.89

Hoogeboom, M. A., & Wilderom, C. P. (2019). A Complex Adaptive Systems Approach to Real-Life Team Interaction Patterns, Task Context, Information Sharing, and Effectiveness. *Group; Organization Management*, *45*(1), 3–42. https://doi.org/10.1177/1059601119854927

Hultin, M., Jonsson, K., Härgestam, M., Lindkvist, M., & Brulin, C. (2019). Reliability of Instruments That Measure Situation Awareness, Team Performance and Task Performance in a Simulation Setting With Medical Students. *BMJ Open*, *9*(9), e029412. https://doi.org/10.1136/bmjopen-2019-029412

Hutchins, S. G., & Kendall, T. (2011). The Role of Cognition in Team Collaboration During Complex Problem Solving. In *Informed by Knowledge* (pp. 83–104). Psychology Press. https://doi.org/10.4324/9780203847985

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

Jamshed, S., Nazri, M., Abu Bakar, R., & Majeed, N. (2018). The Effect of Knowledge Sharing on Team Performance through Lens of Team Culture. *Oman Chapter of*

Arabian Journal of Business and Management Review, *34*(6112), 1–17. https://doi.org/10.12816/0049504

Kapitaniak, B., Walczak, M., Kosobudzki, M., Jóźwiak, Z., & Bortkiewicz, A. (2015). Application of Eye Tracking in Drivers Testing: A Review of Research. *International Journal of Occupational Medicine and Environmental Health*, *28*, 941–954. https://doi.org/10.13075/ijomeh.1896.00317

Karam, K., Mansour, A., Khaldi, M., Clement, B., & Ammad, M. (2024). A Survey for Unmanned Aerial Vehicles in Smart Agriculture: Types and Modelling Perspectives. *2024 IEEE 7th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 807–818. https://doi.org/10.1109/IAEAC59436.2024.10503803

Katna, S., Singh, S. K., Kumar, S., Manro, D., Chhabra, A., & Sharma, S. K. (2025). Communication Systems for Drone Swarms and Remote Operations. In *AI Developments for Industrial Robotics and Intelligent Drones* (pp. 115–136). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-2707-4.ch006

Kim, Y. J., Engel, D., Woolley, A. W., Lin, J., McArthur, N., & Malone, T. W. (2017). What Makes a Strong Team? Collective Intelligence Predicts Team Performance in League of Legends. *Academy of Management Proceedings*, *2016*(1), 13564. https://doi.org/10.5465/ambpp.2016.13564abstract

Klimoski, R., & Mohammed, S. (1994). Team Mental Model: Construct or Metaphor? *Journal of Management*, *20*(2), 403–437. https://doi.org/10.1016/0149-2063(94)90021-3

Kozlowski, S. W., & Bell, B. S. (2003). Work Groups and Teams in Organizations. *Handbook of Psychology: Industrial and Organizational Psychology*, *12*, 333–375. https://doi.org/10.1002/0471264385.wei1214

Krauzlis, R. (2013). Eye Movements. In *Fundamental neuroscience for basic and clinical applications* (pp. 724–739). Elsevier. https://doi.org/10.1016/B978-0-12-385870-2.00032-9

Kwak, J., & Sung, Y. (2018). Autonomous UAV Flight Control for GPS-Based Navigation. *IEEE Access*, *6*, 37947–37955. https://doi.org/10.1109/ACCESS.2018.2854712

Larson, L. E., Harris-Watson, A. M., Carter, D. R., Asencio, R., DeChurch, L. A., Kanfer, R., & Zaccaro, S. J. (2023). Staying Apart To Work Better Together: Team Structure In Cross-Functional Teams. *Academy Of Management Discoveries*, *9*(3), 320–338. https://doi.org/10.5465/amd.2020.0238

Lee, C., & Wong, C. S. (2017). The Effect of Team Emotional Intelligence on Team Process and Effectiveness. *Journal of Management and Organization*, *25*(6), 844–859. https://doi.org/10.1017/jmo.2017.43

Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., & van Merriënboer, J. J. G. (2020). Different Effects of Pausing on Cognitive Load in a Medical Simulation Game. *Computers in Human Behavior*, *110*, 106385. https://doi.org/10.1016/j.chb.2020.106385

Leifer, L. (1998). Design-Team Performance: Metrics And The Impact Of Technology. In S. M. Brown & C. J. Seidner (Eds.), *Evaluating corporate training: Models and issues* (pp. 121–136, Vol. 46). Springer, Dordrecht. https://doi.org/10.1007/978-94-011-4850-4_14

Leshed, G., Perez, D., Hancock, J. T., Cosley, D., Birnholtz, J., Lee, S., McLeod, P. L., & Gay, G. (2009). Visualizing Real-Time Language-Based Feedback On Teamwork Behavior In Computer-Mediated Groups. *Proceedings Of The SIGCHI Conference On Human Factors In Computing Systems*, 537–546. https://doi.org/10.1145/1518701.1518784

Lim, Y., Gardi, A., Sabatini, R., Ramasamy, S., Kistan, T., Ezer, N., Vince, J., & Bolia, R. (2018). Avionics Human-Machine Interfaces and Interactions for Manned and Unmanned Aircraft. *Progress in Aerospace Sciences*, *102*, 1–46. https://doi.org/10.1016/j.paerosci.2018.05.002

Lin, C. P., He, H., Baruch, Y., & Ashforth, B. E. (2017). The Effect of Team Affective Tone on Team Performance: The Roles of Team Identification and Team Cooperation. *Human Resource Management*, *56*(6), 931–952. https://doi.org/10.1002/hrm.21810

Lin, Y., Zhang, W. J., & Koubek, R. J. (2004). Effective Attention Allocation Behavior and Its Measurement: A Preliminary Study. *Interacting with Computers*, *16*(6), 1195–1210. https://doi.org/10.1016/j.intcom.2004.08.006

Mao, R., Li, G., Hildre, H. P., & Zhang, H. (2021). A Survey of Eye Tracking in Automobile and Aviation Studies: Implications for Eye-Tracking Studies in Marine Operations. *IEEE Transactions on Human-Machine Systems*, *51*(2), 87–98. https://doi.org/10.1109/THMS.2021.3053196

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A Temporally Based Framework and Taxonomy of Team Processes. *Academy of Management Review*, *26*(3), 356–376. https://doi.org/10.5465/amr.2001.4845785

Martin Jr, M. C. (1997). Opportunities And Challenges For Boards In Times Of Change. *New Directions For Community Colleges*, *98*, 55–62. https://eric.ed.gov/?id=EJ549471

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

Matthews, M. D., & Beal, S. A. (2002). Assessing Situation Awareness in Field Training Exercises. *Military Academy West Point NY Office of Military Psychology and Leadership*, 31.

McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2017). Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *60*(2), 262–273. https://doi.org/10.1177/0018720817743223

Meneghel, I., Salanova, M., & Martínez, I. M. (2016). Feeling Good Makes Us Stronger: How Team Resilience Mediates the Effect of Positive Emotions on Team Performance. *Journal of Happiness Studies*, *17*(1), 239–255. https://doi.org/10.1007/s10902-014-9592-6

Meyer, B., Burtscher, M. J., Jonas, K., Feese, S., Arnrich, B., Tröster, G., & Schermuly, C. C. (2016). What Good Leaders Actually Do: Micro-Level Leadership Behaviour, Leader Evaluations, And Team Decision Quality. *European Journal Of Work And Organizational Psychology*, *25*(6), 773–789. https://doi.org/10.1080/1359432X.2016.1189903

Meyer, B., Shemla, M., Li, J., & Wegge, J. (2015). On The Same Side Of The Faultline: Inclusion In The Leader's Subgroup And Employee Performance. *Journal Of Management Studies*, *52*(3), 354–380. https://doi.org/10.1111/joms.12118

Michel, N., Guillaume, M., & Peter, S. (2021). The Tougher the Environment, the Harder the Adaptation: A Psychological View. *Acta Astronautica*, *192*, 88–102. https://doi.org/10.1016/j.actaastro.2021.03.015

Michinov, E., & Michinov, N. (2020). When Emotional Intelligence Predicts Team Performance: Further Validation of the Short Version of the Workgroup Emotional Intelligence Profile. *Current Psychology*, *41*(3), 1323–1336. https://doi.org/10.1007/s12144-020-00659-7

Mickan, S., & Rodger, S. (2000). Characteristics of Effective Teams: A Literature Review. *Australian Health Review*, *23*(3), 201–208. https://doi.org/10.1071/AH000201

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor No More: A 15-Year Review of the Team Mental Model Construct. *Journal of Management*, *36*(4), 876–910. https://doi.org/10.1177/0149206309356804

Momeny, L. S., & Gourgues, M. (2019). Communication That Develops: Clarity Of Process On Transformational Leadership Through Study Of Effective Communication Of Emotional Intelligence. *Christian Education Journal*, *16*(2), 226–240. https://doi.org/10.1177/0739891319829484

Mu, B., Guo, F. B., & Chu, X. (2024). Assisting Human-centred Ship Bridge Design through Virtual Reality. *Advances in Human Factors and Ergonomics*, *28*, 340–356. https://doi.org/10.54941/ahfe1005265

Noton, D., & Stark, L. (1971). Scanpaths in Saccadic Eye Movements While Viewing and Recognizing Patterns. *Vision Research*, *11*, 929–IN8. https://doi.org/10.1016/0042-6989(71)90213-6

Olaisen, J., & Revang, O. (2017). Working Smarter and Greener: Collaborative Knowledge Sharing in Virtual Global Project Teams. *International Journal of Information Management*, *37*(1), 1441–1448. https://doi.org/10.1016/j.ijinfomgt.2016.10.002

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator. *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications*. https://doi.org/10.1145/1743666.1743701

Poole, A., & Ball, L. J. (2006). Eye Tracking in HCI and Usability Research. IGI Global. https://doi.org/10.4018/978-1-59140-562-7.ch034

Richardson, D. C., & Dale, R. (2005). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science*, *29*(6), 1045–1060. https://doi.org/10.1207/s15516709cog0000_29

Salas, E., Reyes, D. L., & Woods, A. L. (2017). The Assessment of Team Performance: Observations and Needs. In *Innovative assessment of collaboration* (pp. 21–36). Springer. https://doi.org/10.1007/978-3-319-33261-1_2

Sassenus, S., Van den Bossche, P., & Schreurs, B. (2024). Team Stress: Grasping Physiological Stress Dynamics in Small Teams Through State Space Grids. *International Journal of Stress Management*, *31*(2), 112–127. https://doi.org/10.1037/str0000322

Shree, J., D.V., Murthy, L. R. D., Saluja, K. S., & Biswas, P. (2018). Operating Different Displays in Military Fast Jets Using Eye Gaze Tracker. *Journal of Aviation Technology and Engineering*, *8*(1), 31. https://doi.org/10.7771/2159-6670.1184

Silva, A., Emmanuel, G., McClain, J. T., Matzen, L., & Forsythe, C. (2015). Measuring Expert and Novice Performance Within Computer Security Incident Response Teams. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial*

Intelligence and Lecture Notes in Bioinformatics), *9183*, 144–152. https://doi.org/10.1007/978-3-319-20816-9_15

Singh, L. (2024). Impact of Induced Stress on Individual Performance, Team Coordination, and Human-AI Collaboration. *University of Southampton Thesis*. https://eprints.soton.ac.uk/487744/

Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N. S., Tripoliti, E., Marias, K., Fotiadis, D., & Tsiknakis, M. (2021). Review of Eye Tracking Metrics Involved in Emotional and Cognitive Processes. *IEEE Reviews in Biomedical Engineering*, *16*, 260–277. https://doi.org/10.1109/RBME.2021.3066072

Škec, S., Cash, P., & Štorga, M. (2017). A Dynamic Approach To Real-Time Performance Measurement In Design Projects. *Journal Of Engineering Design*, *28*(4), 255–286. https://doi.org/10.1080/09544828.2017.1303665

Škvareková, I., Pecho, P., Ažaltovič, V., & Kandera, B. (2020). Number of Saccades and Fixation Duration as Indicators of Pilot Workload. *Transportation Research Procedia*, *51*, 67–74. https://doi.org/10.1016/j.trpro.2020.11.009

Sonoda, Y., Onozuka, D., & Hagihara, A. (2018). Factors Related to Teamwork Performance and Stress of Operating Room Nurses. *Journal of Nursing Management*, *26*(1). https://doi.org/10.1111/jonm.12522

Sorensen, D., Cristancho, S., Soh, M., & Varpio, L. (2024). Team Stress and Its Impact on Interprofessional Teams. *Teaching and Learning in Medicine*, *36*, 45–60. https://doi.org/10.1080/10401334.2022.2163400

Steitz, B. D., Unertl, M., K., & Levy, M. A. (2020). Characterizing Communication Patterns Among Members Of The Clinical Care Team To Deliver Breast Cancer Treatment. *Journal Of The American Medical Informatics Association*, *27*(2), 236–243. https://doi.org/10.1093/jamia/ocz151

Sulistyawati, K., Chui, Y. P., & Wickens, C. D. (2008). Multi-Method Approach to Team Situation Awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(4), 463–467. https://doi.org/10.1177/154193120805200455

Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do Team and Individual Debriefs Enhance Performance? A Meta-Analysis. *Human Factors*, *55*(1), 231–245. https://doi.org/10.1177/0018720812448394

Taylor, R. M. (1990, July). Situational Awareness Rating Technique (Sart): The Development of a Tool for Aircrew Systems Design. https://doi.org/10.4324/9781315087924-8

Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task Performance and Eye Activity: Predicting Behavior Relating to Cognitive Workload. *Aviation, Space, and Environmental Medicine*, *78*(5), B176–B185.

Tuncal, A. (2024). Air Traffic Controllers' Perspectives on Unmanned Aerial Vehicles Integration into Non-Segregated Airspace. *Journal of Aviation*, *8*(2), 153–165. https://doi.org/10.30518/jav.1475735

Van der Hoek, M., Groeneveld, S., & Kuipers, B. (2018). Goal Setting in Teams: Goal Clarity and Team Performance in the Public Sector. *Review of Public Personnel Administration*, *38*(4), 472–493. https://doi.org/10.4324/9781315250601

Vessey, W. B., & Landon, L. B. (2017). Team Performance in Extreme Environments. In E. Salas, R. Rico, & J. Passmore (Eds.), *The wiley blackwell handbook of the psychology of team working and collaborative processes* (pp. 329–352). Wiley Blackwell. https://doi.org/10.1002/9781118909997.ch23

Villamor, M., & Rodrigo, M. M. (2018). Predicting Successful Collaboration in a Pair Programming Eye Tracking Experiment. *UMAP 2018 - Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, 263–268. https://doi.org/10.1145/3213586.3225234

Waller, M. J., Gupta, N., & Giambatista, R. C. (2004). Effects of Adaptive Behaviors and Shared Mental Models on Control Crew Performance. *Management Science*, *50*(11), 1534–1544. https://doi.org/10.1287/mnsc.1040.0210

Walton, J., & Gilbert, S. B. (2022). Evaluating the Effect of Displaying Team vs. Individual Metrics on Team Performance. *International Journal of Human-Computer Studies*, *160*, 102759. https://doi.org/10.1016/j.ijhcs.2021.102759

Wang, Q., Yang, Y., Wang, Q., & Ma, Q. (2014). The Effect of Human Image in B2C Website Design: An Eye-Tracking Study. *Enterprise Information Systems*, *8*(5), 582–605. https://doi.org/10.1080/17517575.2014.925585

Wickens, C. D., & Huey, B. M. (1993). *Workload Transition: Implications for Individual and Team Performance*. National Academy Press. https://books.google.com

Wildman, J. L., Salas, E., & Scott, C. P. R. (2014). Measuring Cognition in Teams: A Cross-Domain Review. *Human Factors*, *56*(5), 911–941. https://doi.org/10.1177/0018720813515907

Wiltshire, T. J., Van Eijndhoven, K., Halgas, E., & Gevers, J. M. (2024). Prospects For Augmenting Team Interactions With Real-Time Coordination-Based Measures In Human-Autonomy Teams. *Topics In Cognitive Science*, *16*(3), 391–429. https://doi.org/10.1111/tops.12606

Woodcock, M. (2005). *Team Metrics: Resources For Measuring And Improving Team Performance* (1st Edition). Routledge. https://doi.org/10.4324/9781315087429

Wu, X. J., Jiang, R., Tsai, J. C.-A., & Klein, G. (2023). Shared Mental Models in Multi-Team Systems: Improving Enterprise System Implementation. *International Journal of Managing Projects in Business*, *16*(2), 185–208. https://doi.org/10.1108/IJMPB-05-2022-0119

Yang, Q., Wang, T., Su, N., Xiao, S., & Kapoula, Z. (2013). Specific Saccade Deficits in Patients with Alzheimer's Disease at Mild to Moderate Stage and in Patients with

Amnestic Mild Cognitive Impairment. *Age*, *35*, 1287–1298. https://doi.org/10.1007/s11357-012-9420-z

Yarbus, A. L. (1967). Eye Movements During Perception of Complex Objects. Springer US. https://doi.org/10.1007/978-1-4899-5379-7_8

Yusoff, N. M., & Salim, S. S. (2020). Shared Mental Model Processing in Visualization Technologies: A Review of Fundamental Concepts and a Guide to Future Research in Human-Computer Interaction. In D. Harris & W. C. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics. Mental Workload, Human Physiology, and Human Energy* (pp. 255–265, Vol. 12186). Springer, Cham. https://doi.org/10.1007/978-3-030-49044-7_20

Zhang, W., Liu, Y., & Kaber, D. B. (2024). Effect of Interface Design on Cognitive Workload in Unmanned Aerial Vehicle Control. *International Journal of Human-Computer Studies*, *189*, 103287. https://doi.org/10.1016/j.ijhcs.2024.103287

Zhao, Z., Fu, H., Tang, R., Zhang, B., Chen, Y., & Jiang, J. (2023). Failure Mechanisms in Flexible Electronics. *International Journal of Smart and Nano Materials*, *14*(4), 510–565. https://doi.org/10.1080/19475411.2023.2261775

Chapter 3

Quantifying Team Collaboration and Identifying Performance Breakdowns Using Scanpath Similarity Analysis

3.1 Introduction

In high-stakes environments such as unmanned aerial vehicle (UAV) command and control (C2) operations, effective performance often relies on the ability of multiple operators to coordinate their efforts and work together towards a shared goal. Data-rich domains require operators to deal with large quantities of data, often in a short period of time. However, in these domains, the increasing complexity of technology and automation is leading to a growing demand for operators to complete more tasks at varying levels of cognitive workload (Prytz & Scerbo, 2015). High workload and workload variations can negatively affect performance; however, it is not clear how the transition from low to high workload influences collaboration in a dual-task setting (Atweh et al., 2022). This knowledge is important as the change from low to high workload conditions can pose challenges in complex, data-rich domains. From a design standpoint, it is critical to help multiple operators manage varying task loads over time (Dekker & Nyce, 2004).

Over the years, researchers have developed various methods to assess and enhance our understanding of team collaboration. These efforts have spanned multiple disciplines, employing diverse approaches such as behavioral analysis, communication pattern tracking, and physiological measurements. While these methodologies have substantially advanced our knowledge, their application and replicability in complex, real-world scenarios like UAV C2 operations remain limited. Moreover, researchers are still in continuous search for real-time measures to quantify team collaboration.

One promising avenue for addressing this challenge lies in the use of eye tracking technology, which provides detailed real-time insights into visual attention and coordination between team members. This technology provides a trace of people's eye movements, which enables researchers to monitor visual attention in real-time (Lin et al., 2004). Eye tracking technology is increasingly being used to study team performance in complex domains (J. Y. Lee et al., 2020). By tracking a pair's eye movements simultaneously, researchers can gain insights into joint attention and collaboration in complex domains. Eye tracking technology allows researchers to collect objective, quantitative data on how teammates collaborate in real-time. It provides valuable insights into the cognitive processes involved in collaborative problem-solving, including the distribution of attentional resources and the coordination of attention between teammates. This information can be used to identify areas for improvement in team performance and develop strategies for optimizing collaboration in high-stress, data-rich environments.

Eye tracking has been successfully used to quantify workload and stress in individuals, revealing how visual attention shifts under varying demands (e.g., fixation duration; Fan et al., 2023; Yang et al., 2013). However, extending this approach to understand when teammates collectively experience high workload as a team remains unclear, especially in dynamic systems. Studies that use eye tracking to study pair's performance and attention allocation often use gaze coupling or overlap which refers to moments when teammates are looking at the same Area of Interest (AOI), a specific area or region on the screen that researchers have identified to be relevant for analysis purposes. Previous work has shown that the coupling of gaze between collaborating partners may improve the quality of interaction and comprehension (Richardson & Dale, 2005), but this is not always the case (Villamor & Rodrigo, 2022).

To date, the focus has been on the percentage of gaze overlap (i.e., when both teammates look at the same point; Devlin et al., 2019). While these analyses are needed, it is also important to explore the percentage of identical scanpath segments between two people over time (i.e., portions of eye movement paths that match each other between teammates) and the average duration the teammates are synchronized or aligned, especially during changes in workload (Silva et al., 2015). In complex domains, scanning large amounts of data is a crucial component of team collaboration, and quantitative measures of scanning can provide insights into how teams process and analyze information. These measures can indicate how much attention is being paid to specific information and identify potential challenges in the scanning process. By analyzing quantitative data, researchers can identify bottlenecks, inefficiencies, or areas where teams excel.

Understanding the link between scanpath similarity and performance can guide display designs and training instructions to ensure effective attention allocation among teammates (Kang & Landry, 2014). By capturing where and how long individuals look at specific elements in their environment, eye tracking data can then serve as a proxy for cognitive processes and interaction patterns. However, the sheer volume and complexity of this data necessitate sophisticated analytical techniques to extract meaningful insights. Scanpath similarity techniques emerge as a powerful tool in this regard. Scanpath similarity refers to the comparison of the sequences of eye movements (fixations and saccades) made by individuals while they view a visual stimulus. By analyzing how similar or different these sequences are, researchers can gain insights into visual attention patterns, cognitive processes, and decision-making behaviors.

This chapter of this dissertation investigates the three scanpath similarity approaches. Specifically, we examine two well-established metrics, ScanMatch and MultiMatch, and introduce a novel technique originally developed for time series analysis, Multidimensional Cross-Recurrence Quantification Analysis (MdCRQA). MdCRQA allows for the examination of temporal coordination between multiple time series, making it particularly well-suited for studying the intricate dynamics of team interactions. By applying MdCRQA alongside ScanMatch and MultiMatch, we aim to explore its potential in capturing the nuances of visual attention patterns and cognitive processes. The goal is to apply these metrics to eye tracking data of pairs working on UAV C2 tasks as workload increases to assess whether these metrics (1) are sensitive to workload changes and (2) can quantify team collaboration and adaptation.

3.2 Background

3.2.1 ScanMatch

ScanMatch is a notable example of scanpath similarity (Cristino et al., 2010). ScanMatch has been used to compare the scanpaths of physics problem solvers (Madsen et al., 2012), discover the preferences of individuals with autism (Król & Król, 2019), and study complex visual search patterns (Frame et al., 2019). By analyzing the similarity of scanpaths in tasks requiring visual search, researchers were able to identify common strategies or patterns used by participants. This method is based on the Needleman-Wunsch algorithm (Needleman &

Wunsch, 1970) that was created to compare biological sequences. The ScanMatch algorithm includes two main steps: (1) creating sequences of letters that denote the sequence of AOIs fixated by the user and (2) calculating similarity scores between these sequences.

Eye movement data collected through any eye tracker can be utilized to construct a sequence of eye movements amenable to the application of the ScanMatch method, as outlined by Cristino et al. (2010). Initially, the data requires filtering to discern saccades and fixations. Subsequently, in order to encode string sequences, images must be partitioned into Regions of Interest (ROIs), which may represent specific features within an image (such as eyes, mouth, and nose in a facial image; or doors, windows, and individuals in an indoor scene), or be delineated by discretely binning the image. Each region is then designated a letter, with every eye fixation within that region being tagged accordingly.

A limitation observed in previous string-based methodologies was the absence of fixation duration encoding. To address this concern, the authors introduced temporal binning into the string, by repetitively assigning the letter corresponding to the ROI based on the fixation duration. This method ensures that the resulting string encompasses spatial location, sequential information, and temporal durations.

As depicted in Figure 3.1, each fixation is denoted by the letter corresponding to the ROI it landed on (e.g., ACB). To incorporate fixation durations, dwell times can be sampled (as demonstrated in the example with 50-millisecond bins), with each sample assigned the ROI letter of the fixation. Imagine as if we were progressing through the eye movement sequence and periodically revisiting it every 50 milliseconds. During each revisit, we observe where the eye is fixating at that particular moment. This results in a sequence such as "AACCCCBBBBB".

The similarity score is a value ranging between 0 and 1. The higher the score, the more similar the scanpaths are. In other words, two identical sequences of AOIs would then result

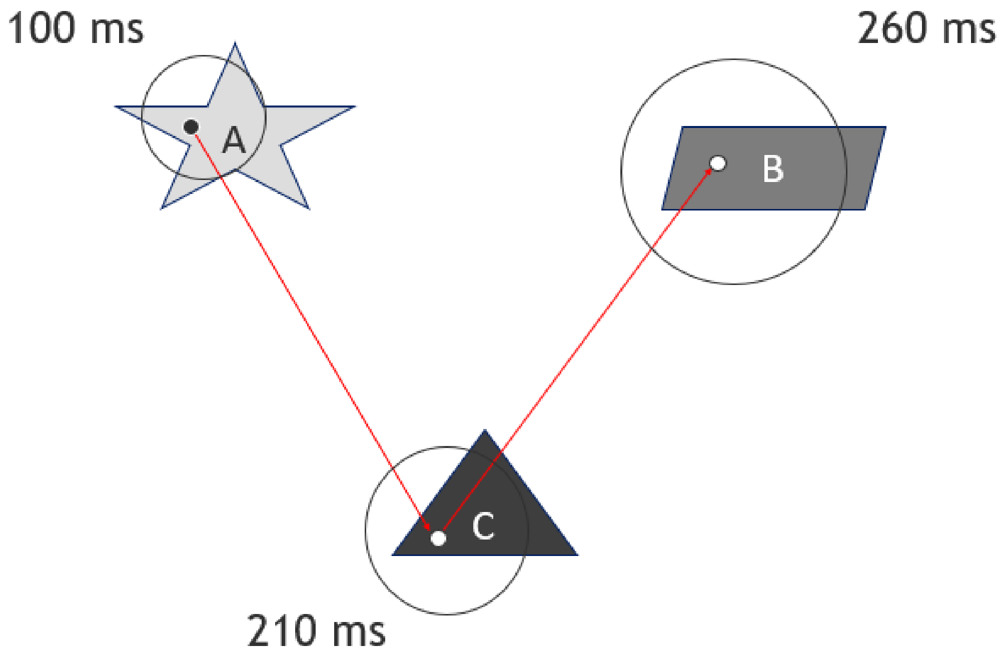


Figure 3.1: From fixations to eye movement sequences. Each fixation is given the letter of the region of interest (ROI) where it landed (in this example, ACB). To take into account fixation durations, dwell times can be sampled (in this example, with 50-msec bins), with each sample given the ROI letter of the fixation, resulting in the following sequence: AACCCCBBBBB, (Cristino et al., 2010).

in a ScanMatch score of 1. ScanMatch thus provides a single quantitative measure of the similarity of two scanpaths. However, ScanMatch's dependence on AOIs means that AOI's size and order can greatly affect the output (Anderson et al., 2015). In addition, condensing scanpath similarity to just one measure does not paint a complete or detailed picture of what is going on; it neither provides insight into the duration of team member's scanpaths and how they are related nor in what aspects the scanpaths are similar. This illustrates the need for a multi-dimensional measure which provides information about the temporal aspects of scanpaths, relationships between fixations, and specific areas of similarity or divergence.

3.2.2 MultiMatch

MultiMatch is one such scanpath comparison method that attempts to address some of the limitations of ScanMatch. It has been used in the literature in experiments to test memory performance (Foulsham et al., 2012), assess student cognitive processes (Stranc & Muldner, 2020), and study weather forecasters' decision-making processes (Wilson et al., 2018). MultiMatch is also notable for its robustness, as it manages spatial noise and perturbed scanpaths well (Dewhurst et al., 2012). In addition, the code for calculating MultiMatch is freely available online.

The MultiMatch algorithm involves several key steps. Initially, the scanpath is transformed into a sequence of vectors, with each vector representing a saccade (Anderson et al., 2015). This transformation is followed by a series of simplifications (see Figure 3.2). The first simplification merges vectors with similar directions, while another simplification clusters consecutive saccades with amplitudes below a predetermined threshold into a single vector. Subsequently, the scanpaths are aligned temporally (Anderson et al., 2015). Finally, the corresponding vectors are compared (Foulsham et al., 2012).

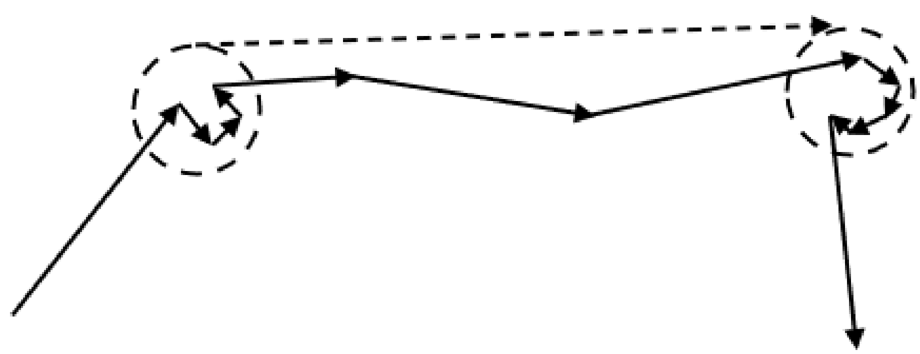


Figure 3.2: Illustration of the MultiMatch simplification steps. Amplitude-based (dashed circles) and direction-based (dashed arrow) clustering (Dewhurst et al., 2012).

Five separate comparisons are performed, resulting in five measures (shape, length, direction, position, duration). The results are averaged across the number of vectors and normalized to yield a value ranging between 0 and 1, where 1 represents perfect similarity. Each measure has its own significance and represents a certain spatial or temporal aspect of similarity as seen in table 3.1 (Anderson et al., 2015).

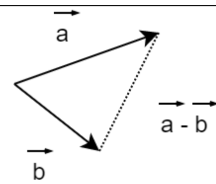
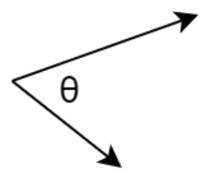
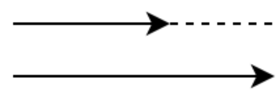
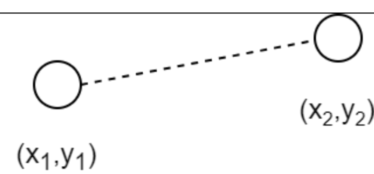
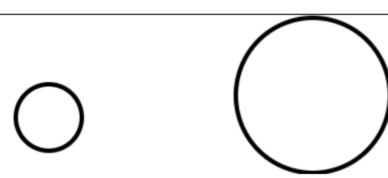
MultiMatch's different measures allow for assessing scanpath similarity at a more fine-grained level than ScanMatch, and also allows for the comparison of scanpaths that have different lengths (Dewhurst et al., 2012). By providing five distinct measures, MultiMatch allows for a detailed analysis of different aspects of scanpath similarity. Researchers can gain insights into not just overall similarity but also specific characteristics like length, direction, and duration of visual exploration. It is important to note that absolute scores of each MultiMatch measure cannot be compared against each other as each measure is calculated and normalized differently (Dewhurst et al., 2012; Wilson et al., 2018). Even though one downside of MultiMatch is that the threshold needs to be carefully selected, its present advantages were the reason it was selected as one of the metrics for this study.

3.2.3 Multidimensional Cross-Recurrence Quantification Analysis

Overview of MdCRQA

Recurrence is a fundamental characteristic of many dynamic systems, based on the assumption that, over time, a system will revisit conditions or states that are arbitrarily close to its prior conditions and follow a similar evolution (Marghitu et al., 1997; Poincaré, 1890). Cross-Recurrence Quantification Analysis (CRQA) is an extension of Recurrence Quantification Analysis (RQA; Marwan & Kurths, 2002) to a bivariate analysis technique, allowing for the quantification of temporal coupling or similarity between two time series. CRQA examines the recurrence of values in one time series with those in another, enabling the analysis of the co-evolution of two signals. This method is robust against outliers and does

Table 3.1: Summary of the MultiMatch measures and their indications. In all cases, a measure of 1 would indicate perfect similarity between the two scanpaths being compared (Atweh & Riggs, 2024; Dewhurst et al., 2012).

Measure	Definition
Shape Similarity	 The vector difference between aligned saccade pairs.
Direction Similarity	 The angular difference between aligned saccades.
Length Similarity	 The endpoint difference in length of the pair.
Position Similarity	 The Euclidean distance between aligned fixations.
Duration Similarity	 The difference in duration between aligned fixations.

not assume linearity or specific distribution types, making it suitable for complex dynamic systems (Marwan & Kurths, 2002). CRQA has been applied in various fields, such as investigating joint actions, physiological arousal during rituals, facial movements, eye movements, postural sway during conversations, and joint gaze behavior. However, traditional CRQA is limited to coupling unidimensional time series, restricting its application to inherently multivariate behaviors.

Multidimensional Cross-Recurrence Quantification Analysis (MdCRQA) extends the principles of CRQA to multidimensional time series, allowing for the analysis of complex dynamic systems involving multiple interacting variables. MdCRQA can be used for two main purposes. Similar to univariate RQA, MdCRQA can quantify the dynamics of a multidimensional construct by simultaneously analyzing its various measurable dimensions (e.g., heart rate, breathing, and body temperature; Byrd et al., 2020). Second, MdCRQA can analyze the shared dynamics of multiple individual time series, such as electrodermal signals from team members during a collaborative task (Coco et al., 2021; Wallot et al., 2016). This approach captures higher-order inter-correlative properties between the signals at the group level.

MdCRQA generates a recurrence matrix that represents the repetition of patterns in two multivariate time series datasets such as the eye movements of two individuals. This matrix is a two-dimensional recurrence plot of binary elements, where each time series is represented by each of the dimensions. This allows for the detection and quantification of similar patterns repeated in both datasets and the recording of their temporal recurrence. This process forms a cross-recurrence plot (CRP; see Figure 3.3), illustrating the recurring patterns of coordination between the two time series. Figure 3.3 presents two CRPs, a simplified CRP with a 10x10 matrix, illustrating basic recurrence patterns (Figure 3.3a) and a more detailed CRP derived from actual eye tracking data, showcasing complex patterns including diagonal and vertical structures (Figure 3.3b). Figure 3.3a shows a simple CRP with a 10x10 matrix, resulting in 100 possible squares. Each black square in the matrix indicates a point of recurrence, where a value in the sequence on the x-axis matches a value in the sequence on the y-axis. The CRP is not just a tool for visualizing sequential properties but also for quantifying them. For example, counting the recurrent points (the black squares in Figure 3.3a) in a CRP provides information about how many individual elements between the two sequences are shared, a measure known as percent recurrence (%REC) or Recurrence Rate (RR). In Figure 3.3a, we see 21 instances of recurrence, meaning that there are 21 points where an occurrence in the sequence on the x-axis is repeated in the sequence on the y-axis. Given that there are 100 possible points (the size of the CRP matrix), RR or %REC can be calculated as $21/100 = 0.21 = 21\%$. This indicates that individual values cross-recur 21% of the time. RR metric measures the overall similarity between time series.

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j} \tag{3.1}$$

In general, RR can be computed using Equation 3.1 below where N is the number of points on the phase space trajectory (Marwan et al., 2007) and $R_{i,j}$ is the recurrence plot as defined by Equation 3.2 below (Marwan et al., 2007):

$$R_{i,j} = \Theta(\epsilon_i - (\|\vec{x}_i - \vec{x}_j\|)) \tag{3.2}$$

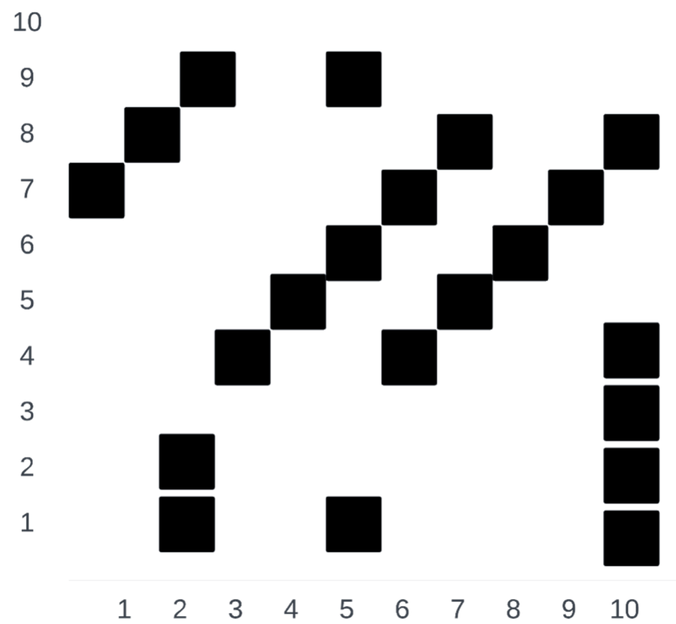
where x_i and x_j are the phase space trajectories of time series i and time series j respectively. $\Theta(x)$ is the Heaviside function and ϵ is the threshold. The states of a natural or engineering dynamic system usually change over time. The state of a system x can be described by its d state variables, $x_1(t), x_2(t), \ldots, x_d(t)$ (Abarbanel et al., 1993). The vector $(\vec{x}(t))$ in a d-dimensional space is called phase space. The system's evolving state over time traces a path, which is called the phase space trajectory of the system.

A high RR indicates frequent visits to similar states, suggesting regularity or periodicity in the systems. In UAV environments and eye tracking studies, a high RR signifies that team members frequently focus on similar AOIs, such as the same UAV controls or areas on the display, indicating similar visual exploration patterns. Low RR, on the other hand, suggests that team members have more divergent visual attention, possibly due to different task allocations or areas of responsibility.

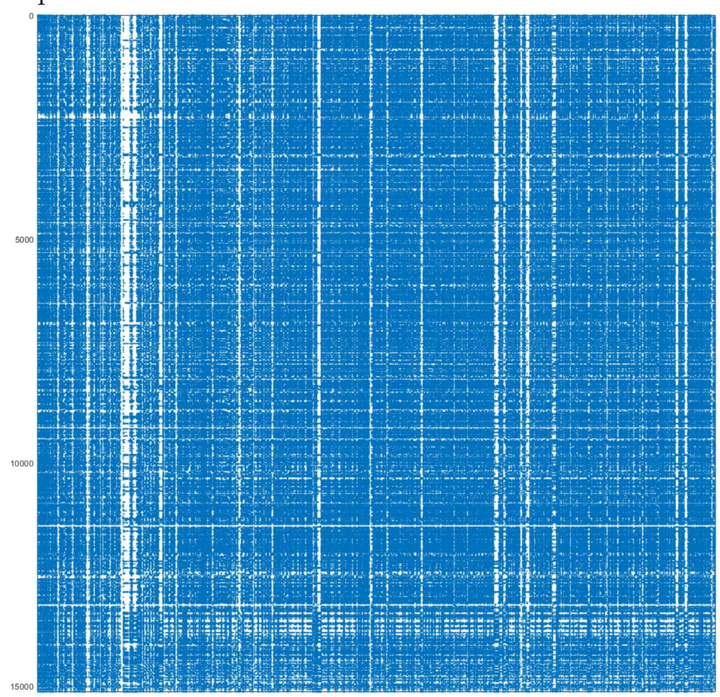
Lines in the CRP can be formed along the main diagonal and vertical structures, each providing unique insights into the dynamics of the time series. The recurrent points in a CRP can form vertical and diagonal lines, which provide further insights into the systems' dynamics. Diagonal lines indicate periods of synchronous or coupled behavior between the systems, while vertical lines suggest periods of stability in one system while the other changes.

Four MdCRQA metrics provide nuanced insights into the diagonal structures of a CRP:

1. **Determinism (DET).** DET is the proportion of recurrent points forming diagonal lines in the CRP. It indicates the predictability of the system. High DET suggests that the system's behavior is deterministic and predictable, while low DET indicates more random or stochastic behavior. DET is calculated by dividing the number of recurrent points forming diagonal lines by the total number of recurrent points. In Figure 3.3a, we identified 21 recurrent points in the CRP. Among these, 13 form diagonally adjacent patterns—meaning that these points are part of longer sequences where the values repeat in a structured manner across both time series. For example, we observed two diagonal lines of length 5 and one diagonal line of length 3. Thus, DET = 13/21=0.6667=66.67%. This result indicates that 66.67% of the recurrences between the two sequences are organized into longer, connected patterns. This suggests that a substantial proportion of the cross-recurrences are not just isolated points but are part of extended trajectories of values that repeat in both sequences, reflecting more



(a) Simplified CRP with a 10x10 matrix as an example. The black squares in the matrices indicate recurrence, and white spaces indicate the absence of recurrence.



(b) A more detailed CRP (15135x15135; with embedding = 1, delay = 1, and threshold or radius r = 0.239). The blue dots in the matrices indicate recurrence, and white spaces indicate the absence of recurrence.

Figure 3.3: Two Cross-Recurrence Plots (CRPs) are presented: (a) a simplified CRP with a 10x10 matrix, illustrating basic recurrence patterns and (b) a more detailed CRP derived from actual eye tracking data, showcasing complex patterns including diagonal and vertical structures.

structured and deterministic dynamics between the two time series. In UAV operations and eye tracking analysis, high DET implies that team members' gaze patterns are predictable and follow a structured sequence when monitoring UAV activities, reflecting coordinated teamwork. Low DET would suggest more erratic gaze patterns, potentially indicating a less coordinated or more adaptive response to dynamic situations.

The diagonal measures are based on the histogram $P(l)$ of diagonal lines of length l given by Equation 3.3 below (Marwan et al., 2007):

$$P(l) = \sum_{i,j=1}^{N} (1 - R_{i-1,j-1})(1 - R_{i+1,j+1}) \prod_{k=0}^{l-1} R_{i+k,j+k} \tag{3.3}$$

The ratio of recurrence points that form diagonal structures (of at least length $l_{\min}$) to all recurrence points is introduced as a measure for determinism of the system. DET can be computed using Equation 3.4 below (Marwan et al., 2007):

$$DET = \frac{\sum_{l=l_{\min}}^{N} l \cdot P(l)}{\sum_{l=1}^{N} l \cdot P(l)} \tag{3.4}$$

2. **Average Diagonal Line Length (ADL or L).** The Average Diagonal Line Length (ADL or L) measures the mean length of all diagonal lines of recurrent points in the CRP. L represents the average duration of similar behavior between systems. Longer diagonal lines suggest more sustained periods of synchronization or coupling between the systems. In Figure 3.3a, we identified 3 diagonal lines of lengths 5, 5, and 3, so L = (5+5+3)/3 = 13/3 = 4.3. In our analysis, a high L value indicates that team members maintain their gaze on similar AOIs for extended periods, showing stable and coordinated attention to critical tasks. Low L values imply brief and intermittent periods of recurrent states or shared attention, signifying more rapid transitions and frequent shifts in focus between different AOIs.

Diagonal lines of length l indicate that segments of the trajectories are in proximity to one another during l time steps, which suggests a connection to the divergence of trajectory segments. L can be computed using Equation 3.5 below (Marwan et al., 2007):

$$L = \frac{\sum_{l=l_{\min}}^{N} l \cdot P(l)}{\sum_{l=l_{\min}}^{N} P(l)} \tag{3.5}$$

3. **Maximum Diagonal Length (MaxL).** MaxL represents the length of the longest diagonal line of recurrent points in the CRP. MaxL indicates the longest uninterrupted period of similar behavior between the systems. Higher values suggest longer intervals of synchronization or stable interaction. In UAV operations, a high MaxL value reflects that team members share prolonged attention to the same AOIs, indicating strong and sustained team coordination. Low MaxL values imply shorter intervals of shared focus, potentially reflecting rapid changes in task demands or individual task focus. In Figure 3.3a, MaxL is then 5.

   MaxL can be computed using Equation 3.6 below (Marwan et al., 2007):

$$MaxL = \max\{l_i\}_{i=1}^{N_l} \tag{3.6}$$

   where

$$N_l = \sum_{l > l_{\min}} P(l)$$

4. **Shannon Entropy of Distribution of Diagonal Lines (EntrL).** EntrL measures the complexity or unpredictability of diagonal line lengths in the CRP. High entropy indicates a more complex and varied interaction, while low entropy suggests simpler and more uniform behavior. In UAV environments, high EntrL suggests that the team members' gaze patterns are complex and varied, indicating flexible and adaptive visual strategies in response to changing UAV operational demands. Low EntrL indicates more uniform gaze patterns, which may reflect routine tasks or highly structured pro-

cedures.

In general terms, the metric entropy refers to the Shannon entropy of the probability, i.e., $p(l) = P(l)/N_l$, to find a diagonal line of exactly length l in the CRP which reflects the complexity of the CRP in respect of the diagonal lines. EntrL can be computed using Equation 3.7 (Marwan et al., 2007):

$$EntrL = -\sum_{l=l_{\min}}^{N} p(l) \ln p(l) \tag{3.7}$$

Four MdCRQA metrics provide nuanced insights into the vertical structures of a CRP:

1. **Laminarity (LAM)**. LAM is analogous to DET but for the vertical structure. LAM is the percentage of recurrent points forming vertical structures in the cross-recurrence plot (i.e., whether a recurrence point has at least one vertical neighbor). It indicates the extent of laminar (smooth and continuous) behavior in the system. High LAM suggests more extended periods of stable states, while low LAM indicates more turbulent or abrupt changes. For example, in Figure 3.3a, we can see that we have 6 recurrent points forming two vertical lines. LAM is then 6/21 which is 28.57%. In our analysis, high LAM means that one team member's gaze remains stable on certain AOIs while the other team member's gaze changes, showing periods of visual stability in one member while the other is more dynamic. Low LAM suggests that both team members are frequently shifting their gaze, possibly indicating high task dynamics or frequent updates.

   To calculate LAM and other vertical metrics, a vertical line (with v the length of the vertical line) marks a time interval in which a state does not change or changes very slowly: $\vec{x}_i \approx \vec{x}_j, \vec{x}_i \approx \vec{x}_{j+1}, \ldots, \vec{x}_i \approx \vec{x}_{j+v-1}$. The total number of vertical lines $P(v)$ of the length v in the plot is calculated using Equation 3.8 (Marwan et al., 2007):

$$P(v) = \sum_{i,j=1}^{N} (1 - R_{i,j})(1 - R_{i,j+v}) \prod_{k=0}^{v-1} R_{i,j+k} \tag{3.8}$$

The computation of LAM is realized for those v that exceed a minimal length $v_{\min}$ to decrease the influence of the tangential motion which can be computed using Equation 3.9 below:

$$LAM = \frac{\sum_{v=v_{\min}}^{N} v \cdot P(v)}{\sum_{v=1}^{N} v \cdot P(v)} \tag{3.9}$$

2. **Mean Vertical Line Length (MeanV).** MeanV quantifies the average length of vertical lines in the CRP. MeanV indicates the average duration of periods where one system remains in a similar state while the other changes. Longer vertical lines suggest extended periods of stability in one system. Similarly, in Figure 3.3a, MeanV is then (4+2)/2=3. In UAV operations ana eye tracking studies, a high MeanV value reflects that one team member maintains their gaze on specific AOIs for longer durations while the other team member's gaze shifts, indicating sustained visual attention by one member. Low MeanV values suggest shorter periods of stability, indicating more dynamic interaction where both team members are frequently adjusting their focus.

   MeanV is calculated using Equation 3.10:

$$MeanV = \frac{\sum_{v=v_{\min}}^{N} v \cdot P(v)}{\sum_{v=v_{\min}}^{N} P(v)} \tag{3.10}$$

3. **Maximum Vertical Line Length (MaxV).** MaxV measures the length of the longest vertical line in the CRP. MaxV is thus 4 in figure 3.3a. This metric indicates the longest period where one system remains in a similar state. Higher values suggest extended periods of stability or fixation in one system. In UAV operations, a high MaxV value shows that one team member fixates on a particular AOI for a long period, demonstrating intense visual focus. Low MaxV values indicate that such

fixations are shorter, suggesting more rapid shifts in visual attention.

MaxV can be computed using Equation 3.11 below:

$$MaxV = \max\{v_i\}_{i=1}^{N_v} \tag{3.11}$$

where

$$N_v = \sum_{v > v_{\min}} P(v)$$

4. **Shannon Entropy of Distribution of Vertical Lines (EntrV).** EntrV measures the complexity or unpredictability of vertical line lengths in the CRP. It provides insight into the variability and disorder of the recurrent vertical structures. High entropy indicates more varied and complex periods of stability, while low entropy suggests more uniform and predictable behavior. In UAV environments, high EntrV signifies that the team members' gaze patterns exhibit diverse periods of stability, indicating varied visual attention strategies. Low EntrV indicates more uniform periods of stability, suggesting routine or less varied visual behavior.

   EntrV can be computed using Equation 3.12 (Marwan et al., 2007):

$$EntrV = -\sum_{v=v_{\min}}^{N} p(v) \ln p(v) \tag{3.12}$$

Application Recurrence Analysis on Teams

Recurrence analysis has emerged as a powerful method for examining the dynamics of complex systems by identifying recurrent patterns within time series data. This approach has been widely applied to understand coordination and communication within teams. Richardson and Dale (2005) first used CRP to analyze gaze similarity between two people. They studied the relationship between a speaker and a listener based on their eye movements and found that the coupling between a speaker's and a listener's eye movements was an indicator of listener engagement. Early studies used RQA to explore syntactic coordination between children and caregivers in conversation. For instance, Dale and Spivey (2006) applied recurrence analysis to uncover patterns of syntactic coordination, showing how the structure of language interactions could be quantified and analyzed over time. RR was employed to measure the degree of similarity in syntactic structures used by conversational partners, providing a foundation for understanding how recurrence metrics can reveal underlying coordination dynamics. Later, CRQA was used to quantify team collaboration (Pietinen et al., 2010). It was found that a high rate of overlapping fixations could possibly be an indicator of collaboration and problems in comprehension (Zheng et al., 2016). Another study used gaze cross-recurrence analysis to measure the coupling of the programmers' focus of attention (Villamor & Rodrigo, 2017). Their findings also showed that pairs who used text selection to perform collaborative references have high levels of gaze cross-recurrence.

Building on these foundational studies, researchers began exploring team communication in more complex settings. Marwan et al. (2007) used Joint Recurrence Quantification Analysis (JRQA) to estimate determinism from communication data. This served as an index of flexible behavior, indicating how teams adapt and coordinate in dynamic environments. The use of metrics like DET helped illustrate the degree of structured, predictable communication within teams, reflecting their ability to maintain consistent interaction patterns.

In aviation, CRQA has been used to analyze visual attention patterns during takeoff and landing procedures, as well as in air traffic control tasks. For example, Zheng (2022) used CRQA to examine visual attention patterns during takeoff and landing procedures and found that pilots exhibited low cross-recurrence rates (CRRs), indicating that they may benefit from additional training in visual attention coordination. In a study of air traffic control tasks, Godavarthi et al. (2018) found that controllers who worked together frequently exhibited higher CRRs compared to controllers who did not work together frequently. This finding suggests that frequent collaboration between team members might improve coordination during workload variations, but more work is needed.

Within healthcare, CRQA has been used to analyze visual attention during surgery and nursing tasks. For example, Hajari et al. (2016) used CRQA to analyze visual attention patterns during surgery and found that surgical teams exhibited higher CRRs during critical events, indicating that visual attention coordination improved during high workload periods. In military operations, CRQA has been used to study visual attention patterns during reconnaissance and surveillance tasks. For example, CRQA has been used to analyze the visual attention patterns of soldiers performing marksmanship tasks and to identify key visual attention behaviors that are associated with marksmanship performance (Saeed et al., 2023).

Recurrence analysis has also been applied to measure real-time team cognition during team training. For instance, studies involving medical teams and submarine crews used communication turn-taking data to detect significant reorganizations in response to training events (Gorman et al., 2020). By analyzing these reorganizations, researchers could infer the emergence of leadership and role restructuring within teams. DET was also used here again to track communication reorganization, revealing how teams transition from stable to new communication patterns in response to unexpected challenges.

When it comes to UAV applications, Gorman et al. (2012) used recurrence-based measures to validate team coordination dynamics. In their study, they focused on communication determinism and pattern information, hypothesizing that intact teams, teams with a consistent set of members who have worked together over time, would exhibit greater DET, and higher pattern information compared to mixed teams. Their results showed that intact teams had significantly higher DET and maximum pattern information, demonstrating that consistent team membership fosters more structured and predictable communication patterns over time.

In a recent UAV study, RQA was used to evaluate operator training performance through eye tracking data (Veerabhadrappa et al., 2024). This research aimed to investigate the relationship between gaze behavior dynamics and operator performance during a computer-based simulation task. By transforming gaze data using autocorrelation and applying RQA, metrics such as RR, DET, overall entropy, and LAM were analyzed. The study found that superior performers demonstrated higher DET, entropy, and LAM, indicating that systematic gaze activity aligned with the task structure was a key factor in effective performance. More recent studies have used CRQA in the analysis of environmental factors that affect team performance such as prior knowledge (Villamor & Rodrigo, 2018), speech and communication strategies (Russell et al., 2012), and leadership techniques (Dindar et al., 2022).

Despite these advances, significant gaps remain in our understanding of recurrence metrics in UAV operations. While studies have extensively used RR and DET in analyzing team dynamics, there has been a noticeable lack of focus on other important RQA metrics such as L, MaxL, MeanV, and MaxV. All these metrics have not been validated together in a single study, particularly in the context of collaborative eye tracking and UAV tasks. Incorporating these additional metrics can provide a more comprehensive understanding of team coordination and performance. By utilizing a MdCRQA approach, we can capture these nuanced dynamics and their impact on team performance.

Moreover, the sensitivity of these metrics to workload changes in complex environments and their relationship with performance remains underexplored. Existing research, such as the study on evaluating the relationship between team performance and joint attention using longitudinal multivariate mixed models, has primarily focused on RR alone (Tolston et al., 2016). This study highlighted the importance of joint attention in team performance improvement but did not extend its analysis to UAV tasks or include other metrics. Understanding how a broader range of recurrence metrics behaves under different workload conditions and how they correlate with performance is crucial for advancing our knowledge of team dynamics in complex environments. By addressing these gaps, we can develop a more robust framework for assessing team performance in UAV operations, ultimately leading to more effective training and operational strategies.

By applying MdCRQA, we can further contribute to the quantification of the adaptation processes UAV teams go through in response to changes in workload in more complex domains. Prior research indicates that as workload increases, teams often engage in more explicit division of labor (Cooke et al., 2013), which would likely result in decreased L and MaxL metrics due to less joint attention on the same tasks (Luo et al., 2023; Ozturk, 2022). Similarly, MeanV and MaxV are expected to decrease as workload rises, reflecting more specialized and focused task engagement (Salas et al., 2008). We also expect DET and LAM to decrease and both Entropy numbers to increase. This generally reflects less predictability and more complexity in the system, suggesting diverse and complex patterns of shared attention and more random, unpredictable gaze patterns.

3.3 Methods

3.3.1 Participants

Twenty-six pairs of undergraduate and graduate students at the University of Virginia (52 participants) were recruited for the study (M = 24.6 years, SE = 1.13 years). Each pair consisted of one male and one female. The experiment took 75-90 minutes and participants were compensated $10/hour for their time. This study was approved by the University of Virginia's Institutional Review Board (protocol #3480).

3.3.2 Experimental Design

There were two workload conditions, low and high, that were manipulated by varying the number of active UAVs for the primary (target detection) task. For the low workload condition, 3-5 UAVs were always active, while 13-16 UAVs were active at all times for the high workload condition. These numbers were validated using NASA-TLX and performance measurements. The NASA-TLX questionnaire can be found in Appendix A. NASA-TLX dimensions of interest for the purposes of this study included mental demand, temporal demand, and performance. In each experimental condition, pairs completed two 15-minute sessions, one with each of the two workload conditions. Pairs always completed the low workload condition before the high workload condition. The design of the simulation was based on the "Vigilant Spirit Control Station" the U.S. Air Force uses to develop interfaces to control multiple UAVs (Feitshans et al., 2008; Figure 3.4).

The simulation was developed using the Unity gaming engine and ran on a desktop computer (28" monitor, 2560×1440 screen resolution; Figure 3.4). Participants sat 26-28 inches from the monitor and used a standard mouse to input responses. Pairs were collocated, but each participant viewed separate monitors and used separate mice to input responses. The simulation was networked so participants could see inputs from their partner in real-time (e.g., when Participant 1 responded to a chat message, Participant 2 could see his/her response in real-time). Two desktop-mounted FOVIO eye trackers with a sampling rate of 60 Hz were used to collect point of gaze data. The average degree of error for this eye tracker is 0.78° (SD = 0.59°).

3.3.3 UAV Tasks and Point Values

Each pair was responsible for completing a primary task and three secondary tasks—i.e., four tasks total—for up to 16 UAVs. Although all tasks were the pair's responsibility, only

Figure 3.4: The experimental setup with the simulation shown on the two networked computers.

one participant from each pair had to complete each task. Response times and accuracy for each task for each pair were recorded. The four tasks were as follows:

1. *Target detection task (primary task).* Pairs monitored each UAV's video feed and indicated whether a target – a semi-transparent cube – was present (Figure 3.6). When a UAV was approaching a waypoint (predetermined area of interest denoted on the Map panel), its video feed could become "active" (i.e., video feed became highlighted; Figure 3.6). If a semi-transparent cube appeared while the video feed was active, the pair was instructed to press the target button to indicate a target was present; if no target was present, then no response was necessary. UAV video feeds were active for 10 seconds and a target could appear with 4-7 seconds left in this time interval. Pairs were instructed that the target detection task had the highest priority among the four

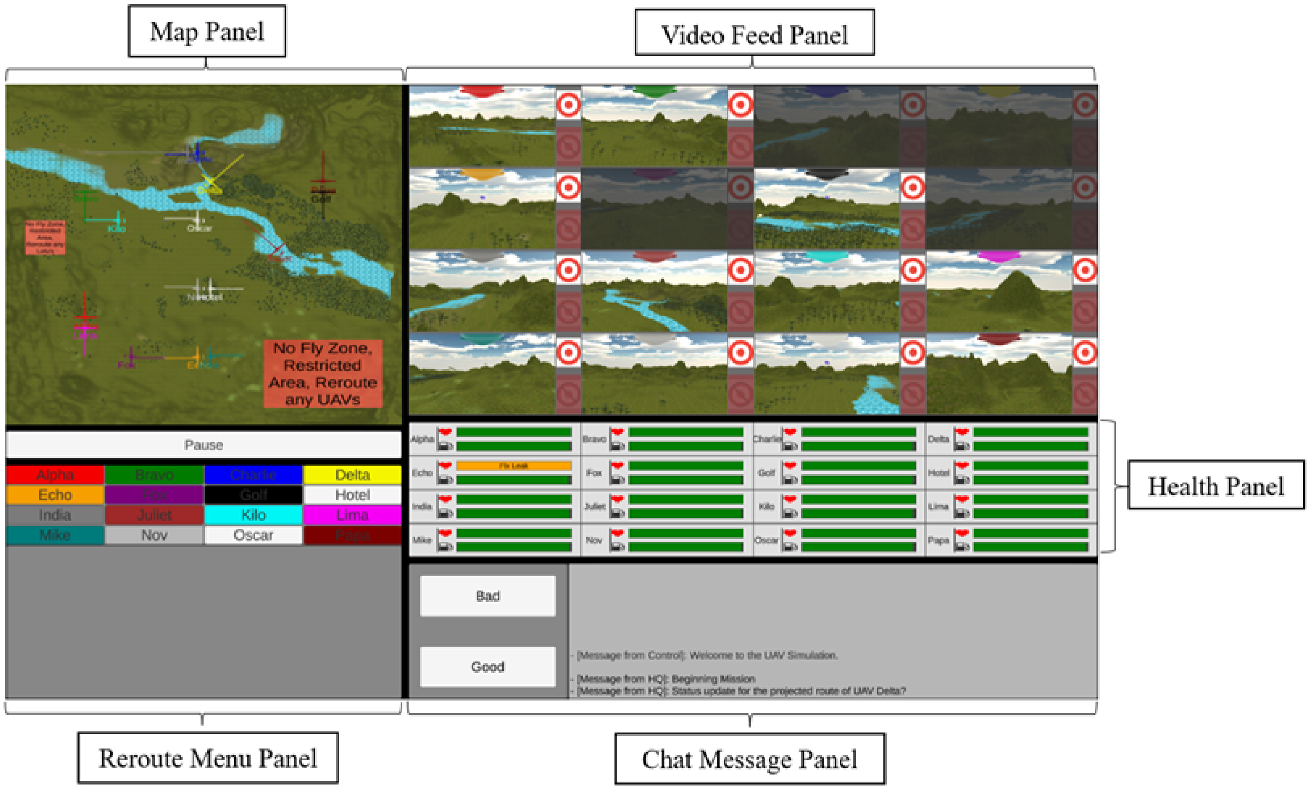


Figure 3.5: A screenshot of the UAV simulation with panels labeled.

tasks. In the low workload condition, one target appeared on one of the active UAV video feeds every 10 seconds. For the high workload condition, three targets appeared on three different active UAV video feeds every 10 seconds.

2. *Reroute task (secondary task 1)* Pairs were tasked to reroute a UAV when it was projected to enter a no-fly-zone, denoted by a red square on the Map panel (Figure 3.7). To reroute a UAV, a participant clicked on the respective UAV's numbered square in the Reroute Menu panel to activate the reroute menu that listed three alternative route options. Participants could click 'Preview' to see a specific alternative's suggested route. When the UAV was not rerouted in time (i.e., entered a no-fly-zone) it would no longer be able to complete the remainder of the mission. The rerouting task occurred 17 times in each condition.

3. *Fuel Leak task (secondary task 2)* Pairs were also tasked with monitoring and maintaining the overall health of each UAV. Participants used the General Health panel,

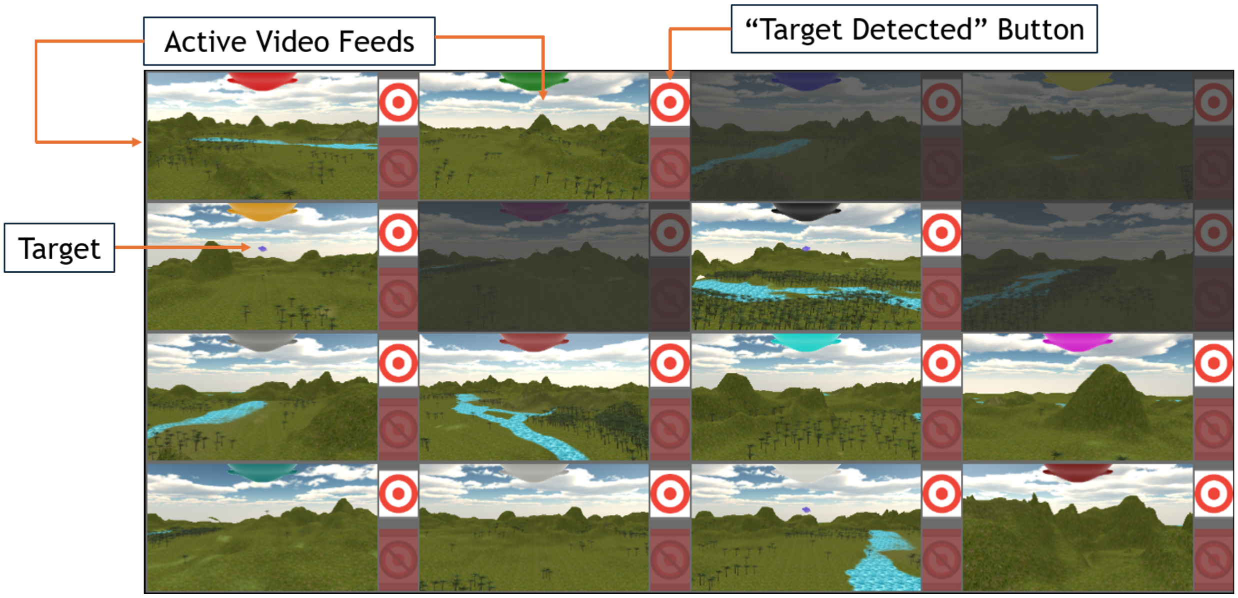


Figure 3.6: Example of active and inactive UAVs on the Video Feed panel and the primary task

which consisted of a health status bar and fuel level bar for each UAV (Figure 3.8). One instance where a UAV would need assistance is if it experienced a fuel leak, which consisted of the UAV's fuel level bar rapidly decreasing in fuel, the color of its health status bar changing from green to yellow, and the message "FIX LEAK" appearing in the health status bar. To stop a fuel leak, the participant clicked on the health status bar. This would change it back to green and stop the fuel from decreasing rapidly. If the leak was not stopped in time, the UAV would reach the "FATAL FUEL LEAK" condition and the task could no longer be completed. A fuel leak occurred a total of 14 times for each condition.

4. *Chat message task (secondary task 3).* Pairs were tasked with responding to messages from headquarters by selecting one of two options on the left-hand side of the chat message panel (Figure 3.9). They were told to respond to as quickly and accurately as possible. There were 19 messages in each condition.

Table 3.2 shows the point value associated with each task. Points were assigned to emphasize the priority of the primary task (i.e., target detection) as well as to convey the

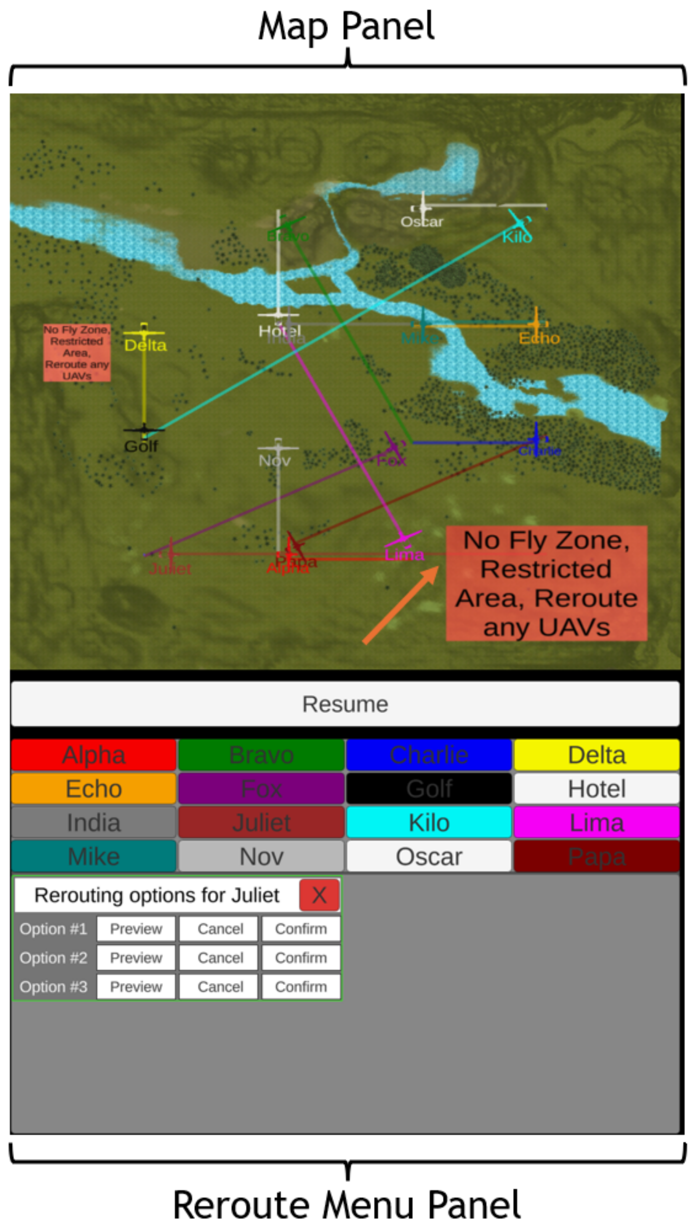


Figure 3.7: Map panel (top half) shows projected routes for each UAV and Reroute Menu panel (bottom half) is where participants could choose a new route for a given UAV to reroute. After clicking on the UAV's name from the buttons in the top four rows in Reroute Menu panel, a menu of route options was presented. The "Preview" button allowed participants to see if the alternative route avoided the No-Fly-Zone, the "Confirm" button rerouted the UAV to that alternative route, and the "Cancel" button removed the overlaid alternative route from the Map panel.

severity of incorrectly or not attending to a task (e.g., UAV flies through no-fly-zone).

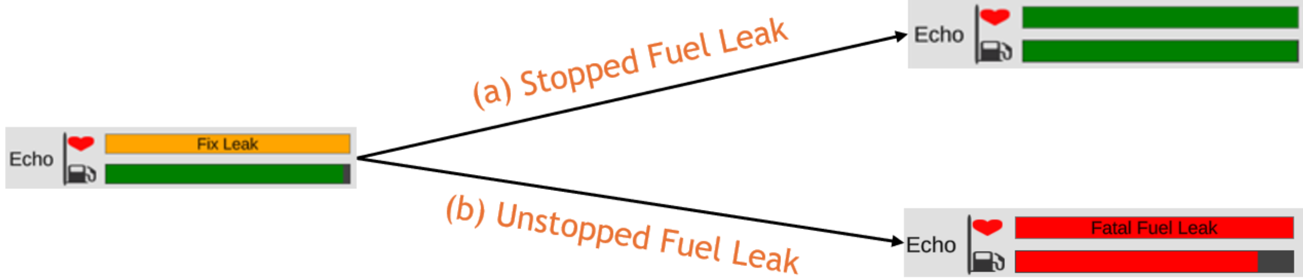


Figure 3.8: Example of how the status of a UAV could change in the General Health panel. Participants were tasked to press the health status bar when a fuel leak occurred: (a) when fuel leak was fixed in time, the health status bar changed from yellow to green and the "FIX LEAK" warning disappeared; (b) when a fuel leak was not fixed in time, the health status bar changed to red and read "FATAL FUEL LEAK" meaning the fuel leak could no longer be stopped.

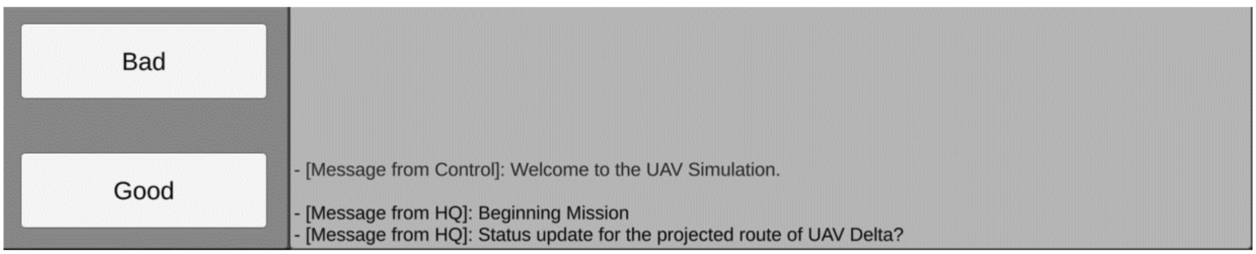


Figure 3.9: Example of an incoming message from the Headquarters prompting a response in the Chat Message panel. Participants could select from one of two options for every message.

If a team fails to reroute a UAV in time, leading it to enter a No-Fly-Zone, or if a UAV experienced a fatal fuel condition due to not fixing a fuel leak, the UAV will be lost, as illustrated in Figure 3.10. Consequently, it will no longer be in mission and participants will (1) lose points for this lost UAV and (2) no longer be able to accumulate points from any tasks assigned to this UAV.

3.3.4 Experimental Procedure

Participants read and signed the consent form, completed a pre-experiment questionnaire (which can be found in Appendix B), and were then briefed about the experiment's goals and task expectations as a pair. Participants then independently completed a five-minute

Table 3.2: Point system for scoring total team performance.

Response to Task	Points per Response
Correctly recognizing a target	+100
Correctly recognizing a non-target	+50
All secondary tasks (reroute, fuel leak, and chat message)	+30
Any incorrect or lack of response (false positive or negative to target detection task, UAV flies through no-fly-zone, or “FATAL FUEL LEAK” condition)	-100

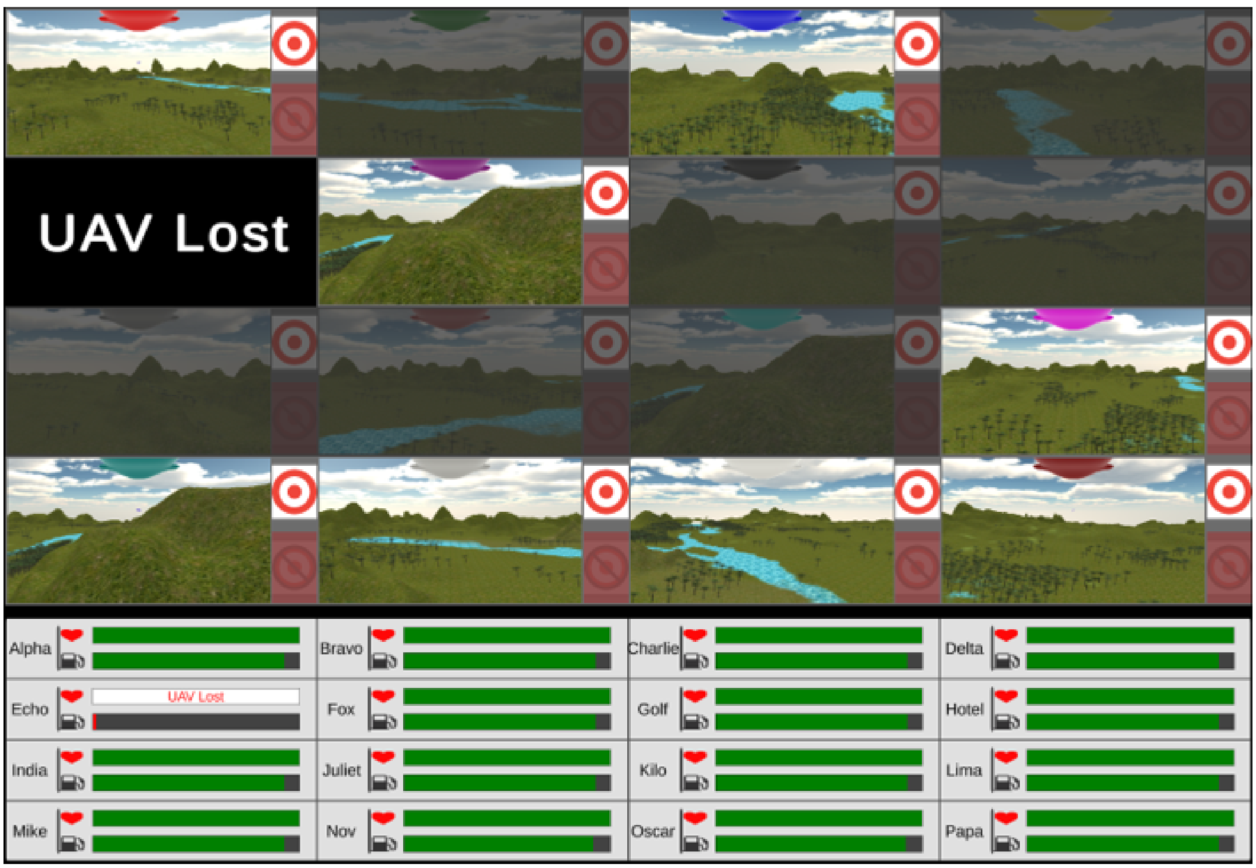


Figure 3.10: Example of a lost UAV (e.g., Echo) on the display.

training trial. By the end of training, participants had to demonstrate they could achieve 70% accuracy for all tasks. We then informed the pairs about how the simulation was networked and provided them with three minutes to introduce themselves to one another and discuss anything they deemed necessary. There were no restrictions on how the participants could interact during these three minutes, i.e., the experimenter gave no guidance on what should be discussed, so discussing team strategies before the experimental portion was completely participant-driven. Afterwards, the audio recording started, and participants completed the low workload condition, were provided a short break, and then completed the high workload condition. We opted for a sequential order in which all pairs completed the low workload condition first, followed by the high workload condition. This deliberate design choice was motivated by the nature of workload variations in complex domains, where the shift from low to high workload often presents a critical period associated with increased cognitive demands and potential human errors. The rationale behind this order prioritization aligns with the operational context of complex domains, enhancing the ecological validity of our findings and offering a more nuanced understanding of workload management in real-world scenarios.

Participants could communicate verbally with each other during the experiment. The same tasks appeared at both stations and the actions of each team member were reflected on both stations, but a participant could not see the cursor movements of their teammate. At the conclusion of the experiment, participants completed a debriefing questionnaire (Table C.1 in Appendix C) and were compensated for their time.

3.3.5 Data Analysis

After we gathered the eye tracking data from the FOVIO eye tracker, we filtered the datasets and removed invalid entries (negative, empty, missing, and out-of-bound values). The gaze data was screened to meet data quality requirements as outlined in ISO/TS 15007-2:2014-09, which states that at most 15% data loss is acceptable. The data loss across all participants and trials was on average 9.8% (SD = 2.5%). We detected fixations and saccades using the code we developed (Atweh et al., 2024). This code is used to analyze eye tracking data collected from experimental studies with participants and it serves two main purposes: (1) filtering the eye tracking dataset and (2) detecting fixations and saccades based on Nyström and Holmqvist (2010)'s velocity-based and data-driven adaptive algorithm. The code, implemented in Python, first takes the raw eye tracking files as input, and filters out empty or invalid recordings. Then, it passes the data through a Savitzky-Golay smoothing filter and calculates the angular velocities in preparation for the data-driven iterative algorithm, which keeps iterating until the absolute difference between the newly calculated velocity threshold and the previous one converges to less than 1°. The event detection code contains five main steps: peak velocities detection, saccade onset detection, saccade offset detection, fixation detection, and saccades detection based on velocity constraints for saccade detection and spatial and duration constraints for fixation detection. See Atweh et al. (2024) for more details on the preprocessing and event detection process for the FOVIO eye tracker.

We used MATLAB to calculate the ScanMatch, MultiMatch, and MdCRQA scores for each pair of participants (one set of scores for low workload and another for high workload):

ScanMatch

Since the monitor has a resolution of 2560×1440 with 16 video feeds (203x381 pixels each), eye movements were binned with a grid size of 7×8 pixels. The screen was thus divided into 56 grids, each representing an AOI which was denoted by a letter. Each fixation was then represented by a letter depending on which AOI the participant fixated on. Every time we ran the algorithm to compare two scanpaths, we calculated the saccade amplitude standard deviation (visual angle in degrees) for each set of eye tracking data and then the mean saccade amplitude standard deviation for the two data sets and converted it to pixels. After importing the eye tracking datasets and inputting the parameters illustrated by Cristino et al. (2010), we converted all sequences of fixations into sequences of letters. This sequence was then used to calculate the ScanMatch score, one each for low and high workload per pair.

MultiMatch

The five measurements were extracted for each condition using the *doComparison* function, the main algorithm of MultiMatch (Dewhurst et al., 2012). The eye tracking data were then divided into one-minute segments and the algorithm was run for each segment in turn. Each pair thus had between 10 and 14 segments to run, and additional code was written to perform the *doComparison* function in batches. This process had to be done due to the large size of the data files that exceeded the RAM limit available. Note that, for our experiment, the *SimplifyExcel* function in the toolbox that pre-processes the eye tracking data was not used, as all the necessary pre-processing had been done beforehand by the event detection software. Finally, we calculated the Pearson correlation coefficients between each MultiMatch measure (i.e., shape, length, position, direction, and duration) and each of the performance measures (points earned by each team, response time, and accuracy). This was done for low workload and high workload separately. The assumptions of normality (assessed using Shapiro-Wilks tests) were met for all variables, and homoscedasticity was checked using plots. In addition, Welch paired t-tests were used to compare the performance results in low and high workload. These tests were used since the variances of the performance results at each workload condition were unequal. In all cases, significance was considered at $p < .05$.

Multidimensional Cross-Recurrence Quantification Analysis Input Parameters

The data that we gathered from the FOVIO eye tracker is 2D, which are the x- and y-coordinates of the eye tracking fixation point. This dataset initially consisted of approximately 54,000 rows of data points (15 minutes per trial or 900,000 ms divided by the eye tracker's refresh rate of 16.667 ms). However, after accounting for the eye tracking data loss, the remaining dataset comprised of around 47,000 rows. Given the memory constraints of MATLAB, we downsampled the data by a factor of 3, preserving the nature of the time series. This resulted in approximately around 16,500 rows per file. Due to varying participant behaviors (e.g., differing blink times), each file had a different number of rows. To address this, we thinned the data to match the size of the smallest time series, ensuring that the time series were of equal size to input into the MdCRQA. Therefore, CRP sizes ranged from 11,243 to 18,152 samples on each axis.

Following preprocessing, the first step was to normalize the data to ensure all parameters could be accurately estimated. The MdCRQA function by Wallot (2019) was employed for the analysis. The function takes the following inputs:

1. **TS1 and TS2:** The two normalized time series (eye tracking files of the two participants in each pair for each condition).
2. **EMB:** Data is two-dimensional; therefore, no embedding is necessary (Iwanski & Bradley, 1998). Embedding dimension is thus set to 1.
3. **DEL:** Delay parameter is thus also set to 1.
4. **NORM:** Type of norm used for phase-space normalization. We used the Euclidean distance norm.
5. **DLINE** Minimum length of diagonal lines, set to the default number of 2.
6. **VLINE:** Minimum length of vertical lines, set to the default number of 2.
7. **RAD:** Radius size within which points in phase-space are counted as recurrent. This was adjusted to maintain a %REC of 5% with a tolerance of 0.1% (Wallot & Leonardi, 2018; Webber & Zbilut, 2005). We started with an initial radius of 10% of the maximum

pairwise distance and developed a function to iteratively run MdCRQA and adjust the radius until the desired %REC was achieved. Radii ranged from 0.229 to 0.292.

8. **ZSCORE:** Indicator of whether the data should be z-scored (normalized) before performing MdCRQA. This was set to 0, as the data was normalized beforehand.

With all inputs prepared for the 52 analyses (26 pairs under 2 conditions), we proceeded to run the MdCRQA for each pair in each condition and calculated the 9 metrics each time. Specifically, for each pair of participants and for both the low and high workload conditions, we applied the MdCRQA function with the parameters as described earlier. As mentioned, the recurrence rate (RR) was set to 5%, which means it cannot be directly compared across conditions. Interestingly, we observed that both determinism (DET) and laminarity (LAM) were consistently at 100% across all conditions. This means that each recurrence point had at least one diagonal and one vertical neighbors. This consistent finding of 100% for both DET and LAM could be influenced by the parameters set for diagonal lines (DLINE) and vertical lines (VLINE), which were both set to 2. We noticed that even when these parameters were increased, DET and LAM remained at 100% and then dropped to zero when reached a certain number. Additionally, this outcome could also be influenced by the application of the Savitzky-Golay smoothing filter, as the filter may have reduced variability in the time series, potentially leading to more uniform recurrence structures (Chelidze & Matcharashvili, 2015). Consequently, we decided to exclude DET and LAM from the analysis of the results. However, other metrics for both diagonal and vertical structures can still provide valuable information about the data.

3.4 Results

3.4.1 Performance Metrics

The mean score in the low workload condition was 43,500 points (SD = 2,200) and for high workload it was 38,000 points (SD = 4,500). The mean response time in the low workload condition was 3.99 s (SD = 0.65) and for high workload it was 4.56 s (SD = 0.74). The mean accuracy in the low workload condition was 91% (SD = 4%) and for high workload it was 87% (SD = 7%). Welch paired t-tests revealed significant differences in scores ($t(25)$ = 2.12, p = .042), response time ($t(25)$ = -3.12, p = .005) and accuracy ($t(25)$ = 2.41, p = .024) between low and high workload.

3.4.2 ScanMatch

Paired t-tests were used to determine whether there is a significant difference in the ScanMatch value between low and high workload. ScanMatch scores were higher in the high workload condition (0.229 $\pm$ 0.05) than the low workload condition (0.226 $\pm$ .014). However, there was not a statistical difference in ScanMatch scores between the two workload conditions ($t(9)$ = -0.061, p = 0.877). Table 3.3 shows the Pearson correlation values between the ScanMatch output and three different measures of performance (points, response time, and accuracy). The correlations with the ScanMatch scores and performance were not significant.

Table 3.3: Summary of the correlation analysis between ScanMatch scores and performance measures between workload conditions.

Variable Correlated with ScanMatch	**Low Workload**		**High Workload**	
	Correlation Coefficient	p	Correlation Coefficient	p
Points	0.23	.53	-0.27	.45
Response Time	-0.38	.28	0.47	.17
Accuracy	0.29	.42	-0.29	.42

3.4.3 MultiMatch

Table 3.4 shows the descriptive as well as the paired t-test results of the MultiMatch metrics between low and high workload scenarios. Four out of the five MultiMatch metrics (shape, direction, length, and duration) significantly changed as workload changed. Shape, length, and duration significantly increased while direction significantly decreased as workload increased.

Table 3.4: Descriptive and paired t-test results of the MultiMatch metrics between low and high workload conditions (N=26).

MultiMatch Metric	**Low Workload**		**High Workload**		***t*-value**	***p***
	Mean	SD	Mean	SD		
Shape	0.68	0.013	0.82	0.032	-4.27	**.012***
Direction	0.79	0.016	0.77	0.027	2.33	**.034***
Length	0.76	0.033	0.82	0.052	-4.3	**.002***
Position	0.35	0.068	0.37	0.093	-0.66	.53
Duration	0.69	0.035	0.72	0.024	-5.088	**<.001****

Note: *p<0.05, **p<0.001

Table 3.5 shows the correlation values between MultiMatch values and the three measures of performance (points, response time, and accuracy). Shape, length, and duration exhibited strong and significant correlation with various performance measures.

3.4.4 Multidimensional Cross-Recurrence Quantification Analysis

Sensitivity of MdCRQA Metrics to Workload Changes

Six paired t-tests were conducted to analyze the MdCRQA metrics between low and high workload conditions. To account for the multiple comparisons and control the family-wise error rate, we applied a Bonferroni correction, adjusting the significance threshold to $\alpha/6 = 0.05/6 = 0.0083$. The analysis of the MdCRQA metrics revealed significant changes between low and high workload conditions. All six metrics showed a decrease from low to high workload, with five out of the six metrics (L, MaxL, EntrL, MeanV, and EntrV) exhibiting statistically significant reductions.

Figure 3.11 visually represents the changes in MdCRQA metrics between low and high workload conditions. Table 3.6 provides detailed statistical analysis including mean differences, standard deviations, t-values, p-values, and Cohen's d effect sizes for each metric, illustrating the change of these measures as workload increased.

Table 3.6: Descriptive and paired t-test results of the MdCRQA metrics between low and high workload scenarios (N=26).

MdCRQA Metric	**Low Workload**		**High Workload**		***t*-value**	***p*-value**	**Cohen's *d***
	Mean	**SD**	**Mean**	**SD**			
L	5.73	2.22	5.23	1.87	4.39	**<.001****	0.86
MaxL	64.58	41.76	53.58	27.88	2.95	**.0069***	0.58

Continued on next page

Table 3.6: Descriptive and paired t-test results of the MdCRQA metrics between low and high workload scenarios (N=26) (continued)

MdCRQA Metric	**Low Workload**		**High Workload**		**t-value**	**p-value**	**Cohen's d**
	Mean	**SD**	**Mean**	**SD**			
EntrL	3.33	0.6	3.18	0.58	4.46	**<.001****	0.87
MeanV	9.93	3.94	8.89	3.77	4.18	**<.001****	0.82
MaxV	140.19	102.62	109.85	63.27	1.86	.074	0.37
EntrV	4.33	0.51	4.13	0.6	4.92	**<.0001*****	0.97

Note: *p<0.05, **p<0.001, ***p<0.0001

Correlation Between MdCRQA Metrics and Performance Measures

Table 3.7 presents the correlation results between the six MdCRQA measures and the three performance measures for low and high workload, points, response time and accuracy. Notably, no strong correlations (defined as $r > 0.6$) were observed in the low workload condition. However, strong and significant correlations were found in the high workload condition. The results show that five MdCRQA metrics have significant negative correlations with response time under high workload (L, MaxL, EntrL, MeanV, EntrV), indicating that as these metrics increase under high workload, response time decreases, reflecting more efficient performance.

Surrogate Analysis

A surrogate analysis was performed to ensure that any observed patterns and results in the original data are not due to random chance (Kantz & Schreiber, 2003). By comparing the original data to its surrogate counterpart, we can confirm that any significant results we derive are statistically significant and not artifacts of the data structure. To generate the surrogate data, we employed a Fourier-based method, which is commonly used for time

Table 3.5: Correlation analysis between the MultiMatch measures and the performance measures. Highlighted values in bold represent strong significant correlations (i.e., an absolute correlation value above 0.6; Hinkle et al., 2003).

MultiMatch Metric	**Performance Measure**	**Low Workload**		**High Workload**	
		Correlation Coefficient	*p*	Correlation Coefficient	*p*
Shape	Points	0.23	.42	**0.78**	<.001
	Response Time	0.16	.65	-0.37	.29
	Accuracy	0.24	.51	**0.82**	.009
Direction	Points	0.43	.19	0.41	.25
	Response Time	-0.21	.56	-0.39	.25
	Accuracy	0.44	.2	0.41	.23
Length	Points	**0.64**	.048	**0.81**	<.001
	Response Time	0.008	.98	-0.46	.18
	Accuracy	0.52	.12	**0.83**	.003
Position	Points	0.36	.3	0.16	.67
	Response Time	-0.48	.16	-0.28	.43
	Accuracy	0.33	.36	0.12	.75
Duration	Points	0.41	.2	0.52	.073
	Response Time	**-0.63**	.005	**-0.65**	<.001
	Accuracy	0.5	.15	**0.67**	.048

Table 3.7: Correlation analysis between the six MdCRQA measures and the three performance measures for low and high workload. Highlighted values in bold represent strong significant correlations (i.e., an absolute correlation value above 0.6; Hinkle et al., 2003).

MdCRQA Metric	**Performance Measure**	**Low Workload**		**High Workload**	
		Correlation Coefficient	p	Correlation Coefficient	p
L	Points	0.45	.2	0.53	.08
	Response Time	-0.09	.66	**-0.63**	**<.001**
	Accuracy	-0.4	.043	0.45	.023
MaxL	Points	0.22	.2	0.4	.22
	Response Time	-0.13	.54	**-0.66**	**<.001**
	Accuracy	-0.35	.081	0.52	.007
EntrL	Points	-0.11	.5	0.35	.18
	Response Time	0.11	.6	**0.65**	**<.001**
	Accuracy	-0.41	.036	-0.43	.029
MeanV	Points	0.46	.16	0.3	.12
	Response Time	-0.018	.93	**-0.62**	**<.001**
	Accuracy	-0.31	.12	0.54	.004
MaxV	Points	0.3	.2	0.52	.073
	Response Time	0.25	.22	-0.3	.13
	Accuracy	0.043	.83	0.41	.038
EntrV	Points	-0.51	.15	0.44	.18
	Response Time	0.03	.88	**0.62**	**<.001**
	Accuracy	-0.35	.079	-0.53	.006

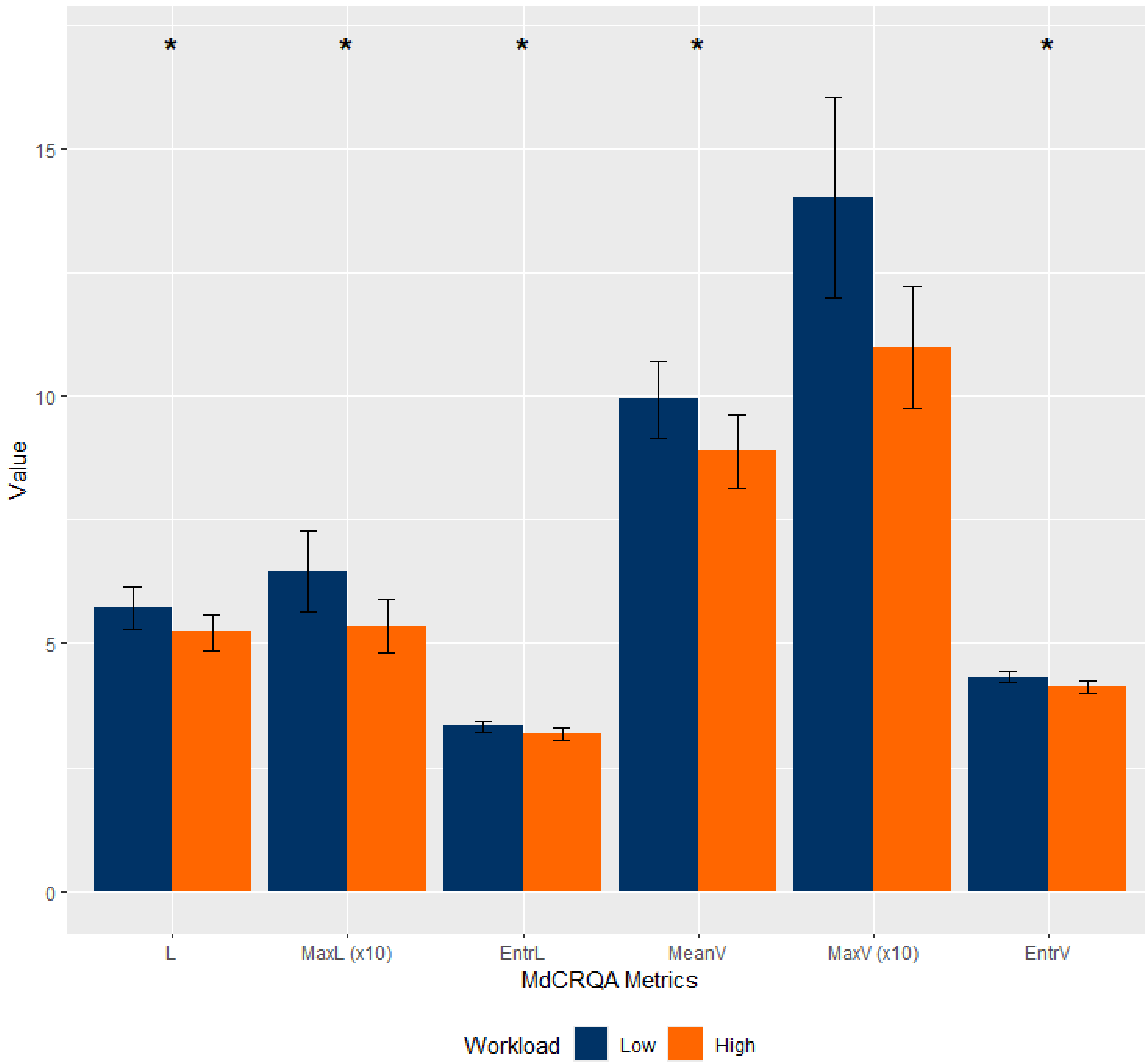


Figure 3.11: MdCRQA metrics for each condition. An asterisk (*) indicates significant main effects for a metric.

series data to preserve the original data's amplitude spectrum while randomizing its phase information. We first applied a Fourier transform to the original data to convert it to the frequency domain. Next, we randomized the phases of the Fourier coefficients, except for the DC component, to ensure that the surrogate data retains the same power spectrum as the original data (e.g., Lopes et al., 2023). Finally, an inverse Fourier transform was applied to convert the data back to the time domain, yielding the surrogate data. For each of the 104 datasets (corresponding to 26 pairs, each with two conditions: low and high workload, and each condition for both participants in the pair), we generated a surrogate dataset using the method described above. The surrogate datasets were then processed in the same manner as the original data to calculate the MdCRQA metrics.

To statistically compare the original and surrogate datasets, we conducted 12 paired t-tests—six for each metric under the low workload condition and six for each metric under the high workload condition. Given the number of comparisons, we controlled the family-wise error rate by applying a Bonferroni correction, setting the significance threshold to $\alpha/12 = 0.05/12 = 0.0042$. The results of the paired t-tests showed significant differences between the original and surrogate datasets across all six metrics for each condition (all $p < 0.000001$). This indicates that any observed patterns in the original data, if available, are highly unlikely to be due to random chance. Consequently, we can proceed with analyzing and testing for differences between conditions, confident that any observed differences are statistically significant and not merely artifacts of the data.

3.4.5 Debriefing Questionnaire Analysis

To complement the quantitative findings from the MdCRQA and performance metrics, we employed open coding to the debriefing questionnaire answers to gain qualitative insights into how participants adjusted their task strategies under different workload conditions. The questionnaire focused on understanding the participants' approaches to task management and coordination as workload increased. Participants were asked whether or not they adjusted their tasks when workload increased. The responses varied but highlighted a common theme of improved task division and focus during high workload conditions. The debriefing questionnaire analysis showed the best five performing pairs strategically adapted both their visual attention allocation and task completion strategies when workload increased, whereas this did not occur for the lowest five performing pairs. The results found each of the best performing pairs changed their visual attention specifically for the primary target detection task, i.e. the task that modulated workload.

One participant from the top five best performing pairs mentioned that *"when the workload was low, we were both multitasking and helping each other out with different aspects of the task. However, as the workload increased, we decided to stick strictly to our predefined roles. This allowed us to avoid confusion and work more efficiently."* Similarly, another participant from the same high-performing pair noted, *"We communicated more explicitly about what each of us was doing. Instead of both trying to do everything, we focused on our specific tasks. This made our actions more predictable to each other, reducing the time we spent coordinating."*

3.5 Discussion

The goal of this chapter was to apply scanpath similarity metrics such as ScanMatch, MultiMatch, and MdCRQA to eye tracking data of pairs of operators in the context of command and control of military UAVs while they are subject to workload changes. The aim is to assess whether these metrics are sensitive to workload changes and whether these metrics significantly correlate with performance. It is interesting that not every scanpath comparison metric we assessed is sensitive to workload changes or has strong significant correlation with performance metrics. Now we will discuss each technique in turn.

3.5.1 ScanMatch

Despite observing slightly higher ScanMatch scores in the high workload condition compared to the low workload condition, the paired t-test did not reveal a statistically significant difference. This suggests that the ScanMatch metric may not be sensitive enough to distinguish variations in workload levels in dynamic settings (Day et al., 2018; Jraidi et al., 2022). The correlation analysis between ScanMatch scores and performance measures further enriches our understanding. In both low and high workload conditions, the correlations were generally weak, and most did not reach statistical significance. This implies that, at least in the context of this study, ScanMatch scores may not strongly relate to performance outcomes. The nuanced correlations with points, response time, and accuracy highlight the complexity of the relationship between eye movement patterns (as captured by ScanMatch) and task performance. These findings contribute to the ongoing discourse on the applicability and limitations of ScanMatch in assessing cognitive workload and performance, emphasizing the need for a more nuanced understanding of these relationships in different task environments. Something noteworthy to mention is that ScanMatch doesn't yield relatively high values on real data throughout the literature, where ScanMatch scores in one study ranged between 0.3 and 0.41 (Madsen et al., 2012), 0.3 to 0.37 in another experiment (Król & Król, 2019), and 0.5 to 0.59 in a third experiment (Crowe et al., 2018). This is intuitive, for ScanMatch is a grid-based algorithm, so changing our grid setup and gap penalty in our experiment may have given different scores. Despite that, our ScanMatch scores were still generally lower than that from previous literature.

3.5.2 MultiMatch

We had predicted that a higher MultiMatch similarity between participants' scanpaths would result in better performance. In other words, if the participants' attention allocation strategies were similar in terms of location, shape, sequence, etc., they would be assumed to be more synchronized and more aware of each other's actions. This would, in turn, enable better team performance. This would then translate to a positive correlation between scanpath similarity and total points, and a negative correlation between scanpath similarity and response time (faster response times meant better performance). This would be in line with previous work that showed pairs with similar attention allocation performed better as a team (e.g., Cherubini et al., 2010; D'Angelo & Begel, 2017), albeit without the level of detail provided by MultiMatch. The findings here could also extend to build on the literature on gaze sharing, i.e., allowing teams to view each other's gaze points on their respective displays while simultaneously completing their tasks, which has been shown to improve performance (G. Lee et al., 2017; Špakov et al., 2019).

Our hypothesis held true for two dimensions of the MultiMatch algorithm: length similarity and duration similarity. For length similarity, there was a strong (> 0.6) and significant positive correlation with total points both in low and high workload, while for duration similarity there was a strong negative correlation with response time in both low and high workload. This suggests that similarities in teammates' saccade lengths and fixation durations matter more than similarities in their fixation positions. It appears that how teammates scan makes more of an impact than where exactly the pair was looking, as evidenced by the low and non-significant correlation for position similarity. It is important to emphasize that what matters here is not necessarily the saccade length or fixation duration of each team member per se, but rather that these are similar for both teammates. Similarly, the high correlation coefficients for length similarity and total points suggest that similarity in saccade length indicates better team performance as well.

The significant difference in performance measures between the low and high workload conditions confirms that performance decrements did occur due to the workload manipulation. We expected there would be a stronger link between scanpath similarity and performance during high workload, i.e., more positive correlation coefficients with total points scored and more negative ones with response time. This was observed for three of the five measures: shape, length, and duration. These measures showed higher correlation coefficients (in absolute value) than their low workload counterpart. The effect of workload was also evident in the significant Fisher z-test results and the slopes of the best fitted lines, where the high workload slopes were greater than their low workload counterparts for all five MultiMatch measures (shape, direction, position, length, and duration similarity). We posit several explanations for these results. First, in the more challenging high workload condition, there is a stronger correlation between scanpath similarity and total points scored. This may be due to the teammates becoming more focused on the task that modulated workload and resulted in the team narrowing their attention allocation to the respective AOI (as evident in Devlin et al., 2019). This could explain why the teams had more similar scanpaths (Wickens & Alexander, 2009). This was true of the best performing pairs as they had a change in their attention allocation strategy, i.e., having a more focused strategy that resulted in more similar scanpaths during high workload compared to a more open-ended/free-gaze strategy during low workload. Second, a notably high correlation coefficient was between shape similarity and total points scored in the high workload condition, that indicates that team members who had more similar scanpath shapes performed better. Dewhurst et al. (2012) noted that shape similarity has been found to be important in fields such as visual imagery research, where fixation order and position are not as crucial as in interface s that have very clear-cut and well-structured AOIs, such as a website. For example, Gbadamosi and Zangemeister (2001) used scanpath shape to compare scanpaths when participants were viewing an image. Given the present testbed consisted of a complex interface with a lot of imagery (e.g., the video feeds), this may be the reason for the observed relation. It seems that shape similarity is capturing a unique and specific aspect of teammates scanpaths and therefore it may be a valid indicator of team performance in a visually data-rich environment.

Thus, it appears that shape, length, and duration similarity are the aspects of MultiMatch that are best suited to assess the performance of teams experiencing high workload in complex domains, much like this experiment's simulation (Fahimi & Bruce, 2021; Newport et al., 2021). It could be that position and direction similarity will be more strongly linked to performance in the context of a simpler/more directed task with fewer areas and targets that can be carefully defined using AOIs.

Overall, MultiMatch appears to be a useful and very promising tool for assessing team attention allocation strategies and how they related to performance especially during high workload periods. The strong correlation of performance with the three MultiMatch measures (shape, length, and duration) can help provide suggestions for interface design and teamwork strategies in complex, multitasking domains. The findings provide support for developing training programs that teach teammates how to coordinate their scanpaths as a means to optimize team performance. This could be done by showing novice teams the scanning approach of expert teams. For example, novices who were trained to mimic expert's visual patterns while reading medical images of lungs (Dempere-Marco et al., 2002) or chest X-rays (Litchfield et al., 2010) showed improved performance. The findings also provide support for design solutions that encourage teammates to scan a display in a similar fashion. For example, the system could highlight what a team member is looking at/scanning (e.g., a box changes color to highlight the shared area if both users are looking at the same lines of code; D'Angelo & Begel, 2017). These developments would be especially beneficial in high-workload and data-rich settings, such as emergency dispatching or process control.

3.5.3 Multidimensional Cross-Recurrence Quantification Analysis

Diagonal Structure MdCRQA Metrics

As workload increased from low to high, Average Diagonal Line Length (L) decreased. This indicates the duration of consistent gaze patterns, showed a reduction, suggesting that teams became more selective and focused their attention on fewer Areas of Interest (AOIs) under higher demands. Similarly, Maximum Diagonal Line Length (MaxL), which measures the longest period of sustained shared visual attention, also decreased with increasing workload. This reduction indicates shorter maximum intervals of similar gaze patterns, suggesting that teams experienced brief moments of shared focus. Nonetheless, higher values of MaxL in high workload conditions were associated with better response times. This suggests that even though the overall duration of sustained shared attention decreased, maintaining extended periods of focus on specific AOIs was beneficial for performance. Teams that effectively managed their visual attention, even intermittently, performed better, emphasizing the role of strategic focus in managing high workloads. This quantitative finding is evidently supported by the qualitative analysis of the debriefing questionnaire, illustrating how clear strategies and effective division of labor contributed to better performance under high workload conditions.

Interestingly, even with a clear division of tasks, high-performing teams demonstrated better response times when they maintained some level of shared attention to specific AOIs, such as their partners' (Ma et al., 2025). This suggests that while L decreased under high workload, higher values of L within this condition still contributed to improved performance, highlighting the importance of retaining some degree of coordinated focus even as task demands escalate. Participants often maintained some level of visual monitoring of their partner's activities, which could explain occasional increases in L and MaxL. This adaptive behavior—balancing focused attention with situation awareness—illustrates how teams strive to maintain both efficiency and coordination, resulting in improved performance despite the challenges of high workload. For instance, when pairs tend to share more identical scanpaths under increased workload, interface designers can prioritize the placement of critical information or key task-related elements in those shared scanpath regions or help direct their attention outside of the shared scanpath regions if needed. This strategic placement ensures that important information is prominently displayed in areas where both users are more likely to focus their attention during high workload (Eraslan et al., 2017; Starke & Baber, 2018).

Moreover, our findings show that although teams experienced reduced periods of shared attention (lower L) and shorter intervals of sustained focus (lower MaxL) in the condition, this shared attention was more ordered focusing on specific AOIs more consistently and predictably. This is shown by the Diagonal Line Entropy (EntrL), which, contrary to our expectations, decreased in high workload conditions. We anticipated more complex and chaotic shared gaze patterns under high workload conditions in UAV operations, reflecting a more erratic scanning behavior. However, the results indicated more regular and predictable gaze patterns, indicating that teams' attention to specific AOIs became more consistent and predictable. This shift towards a more structured and less complex interaction aligns with improved performance. As workload increased, teams demonstrated a more efficient approach by focusing their attention on specific AOIs in a regular and predictable manner, which facilitated better coordination and faster response times. This finding aligns with recent research by Devlin et al. (2020), which suggests that additional workload in UAV operations is not always detrimental. Their study found that workload changes improved response time and accuracy compared to when workload was held constant at low or high levels. Although their study was based on individual UAV operators and not teams, our work extends their result to indicate that higher workload may allow UAV operators to better regulate their mental resources together. This insight can inform the design of operations and technology to assist operators in managing cognitive resources, mitigating the adverse effects of vigilance decrements during low workload periods and data overload during high workload periods (McKendrick et al., 2014; Wohleber et al., 2019).

Furthermore, the insights from L and EntrL can inform the development of training programs by teaching individuals to adopt more efficient and predictable scanpaths during high workload scenarios, thereby improving attentional allocation and decision-making skills. For example, training modules can include exercises and simulations that require individuals to practice scanning techniques that align with the identified patterns of decreased EntrL. By providing feedback and guidance, trainers can help individuals develop a heightened awareness of their scan patterns and encourage the adoption of more optimized strategies. Additionally, decision support systems can benefit from this understanding by incorporating algorithms that consider the predictability of scan patterns, enabling targeted assistance and cues to guide operators' attention toward critical areas or potential sources of information outside of teammates' scans that might be overlooked. By utilizing DET to optimize attentional allocation and scanpath patterns, these applications have the potential to enhance human performance in complex environments (Verdiere et al., 2020).

Vertical Structure Metrics

In addition to the diagonal structure metrics, we also examined the sensitivity of vertical structure metrics. Mean Vertical Line Length (MeanV) showed a significant decrease as workload increased. As workload increased, teams exhibited more frequent gaze shifts, indicating a less stable but more responsive approach to managing multiple AOIs. The observed decrease in MeanV suggests that, as workload intensifies, participants' attention becomes more fragmented, with less time spent on any specific AOI. This increase in gaze shifts might be a strategy to manage the complexity and volume of information presented under higher workloads. Despite this general trend, the relationship between MeanV and performance is nuanced. Higher values of MeanV in high workload conditions were associated with lower response times, suggesting that when teams manage to sustain their focus on specific AOIs for longer periods, even amidst a high workload, it leads to more efficient performance. This is indicative of an effective division of labor, where one participant might maintain attention on a particular AOI while the other engages in broader scanning. The negative correlation between MeanV and response times highlights that well-managed focus on specific AOIs can enhance performance, even under high workload conditions. This implies that while frequent gaze shifts may become necessary due to the increased demands, effectively maintaining periods of focused attention on critical AOIs can enhance performance.

Vertical Line Entropy (EntrV) also decreased significantly as workload increased. Lower EntrV values indicate that when participants focused on specific AOIs as mentioned before, they exhibited more uniform and stable gaze patterns, reflecting a predictable and focused attention on specific AOIs. This shift towards a more ordered visual strategy aligns with improved performance outcomes. Lower EntrV values, which were linked to lower response times in high workload conditions, underscore the advantage of having a more structured and predictable approach to visual attention. Teams that managed to keep their gaze patterns more consistent and organized under high workload conditions were able to perform tasks more efficiently. Maximum Vertical Line Length (MaxV), which measures the longest period of stable gaze by one participant while the other's gaze changes, showed a decrease but did not reach statistical significance. This suggests that MaxV might not be as sensitive to workload variations as MeanV and EntrV. The duration of stable gaze periods was less indicative of performance changes, indicating that MaxV alone might not fully capture the dynamics of how workload impacts team performance.

3.6 Limitations and Future Work

Overall, while our analysis provides valuable insights into the temporal and spatial characteristics of collaborative cognitive processing, it is important to use it in conjunction with other methods to gain a more comprehensive understanding of cognitive processes and recognize its limitations that warrant further investigation. A larger sample size could have allowed us to perform MultiMatch and MdCRQA analysis on the best and worst performing pairs (e.g., top 20 and bottom 20 pairs) to determine to what extent their metrics are predictors of performance. Moreover, with a larger sample size, we may find that there is a significant difference between workload levels for MaxV as it neared significance with this initial analysis ($p = .074$). If significant, MaxL analysis would provide insights into how sustained periods of shared visual attention are impacted by increased workload.

In our experimental design, we deliberately chose to present the different workload conditions in the same order to (a) maintain consistency between participants and (b) capture the change in behavior that occurs when workload increases from low to high. While our approach allowed for a controlled examination of workload, it may not fully capture the nature of all workload considerations observed in operational settings. Future work could consider adding more workload levels (e.g., medium), direction of workload change (i.e., high to low), and rate of change (e.g., gradual). Additionally, examining how teams can maintain certain levels of performance amidst changing cognitive demands could be another future research question (Hill et al., 2020).

More work should further assess whether there is a relationship between the identified metrics and performance outcomes in other data-rich domains. While we found that a subset of MultiMatch and MdCRQA metrics was correlated with performance under low and high workload, it is important for future research to investigate the applicability of these metrics in a variety of domains and tasks, and to identify the specific conditions under which they are most useful.

Additionally, future studies could explore the interaction between workload, these metrics, and other relevant factors, such as team composition, expertise, personality traits, and communication patterns (Atweh et al., 2022). Investigating how these variables interact and influence performance can provide a more comprehensive understanding of the complex dynamics involved in team collaboration and workload management.

Moreover, advanced analytical techniques such as machine learning or predictive modeling could be considered in future research (Atweh & Riggs, 2025a; Atweh et al., 2023; Atweh & Riggs, 2025b). These techniques could also help uncover subtle patterns and predictive relationships between the identified metrics and performance outcomes. For instance, this work provides more future research in terms of team dynamics and integration of experts with novices by showing novices the scanning approach of expert teammates as the workload increases. Moreover, extending the findings to different domains and contexts, such as emergency response, aviation, or healthcare, would enhance the generalizability and applicability of the findings to real-world decision-making settings. Future research could also explore potential interventions or strategies to mitigate the impact of workload variations on team performance, such as training interventions or adaptive decision support systems. These directions could contribute to the development of more effective cognitive engineering interventions aimed at improving team performance in complex and dynamic environments and hopefully gets us one step closer towards quantifying the collaboration process in complex domains.

3.7 Conclusion

In this chapter, we present significant findings that contribute to the field of cognitive engineering and decision making. We have successfully applied a comprehensive collection of scanpath similarity and time series tools and identified a subset of these metrics that are sensitive to workload increases. These metrics offer insights into the occurrence of joint and recurrent patterns, the continuity of similarity, and may be particularly useful in tasks where scanning patterns are critical for task performance. These findings highlight the importance of considering workload dynamics in understanding team scanpath patterns and information gathering behavior in complex tasks, with potential implications for cognitive engineering in real-world settings. The findings highlight the dynamic nature of workload changes in complex systems. To illustrate, the observed decrease in the MultiMatch direction metric and L and MaxL MdCRQ metrics indicate a shift towards effective division of labor and coordinated scan patterns, suggesting a potential adaptive response to manage information processing efficiently. This adaptation may reflect the cognitive strategies employed by individuals and teams to optimize attentional resources and maintain performance in complex tasks. Understanding the specific mechanisms underlying these adaptive strategies, such as the coordination of attentional focus or the allocation of cognitive resources, can provide valuable insights for the design of interventions and training programs aimed at enhancing human performance in dynamic work environments. Future research can delve deeper into the examination of these adaptive strategies, exploring their cognitive underpinnings and their impact on decision-making, task performance, and overall team effectiveness in UAV C2 operations.

We also identified metrics that demonstrated strong and significant correlations with performance only in high workload conditions. In high workload scenarios, maintaining shared and focused attention (reflected by higher L, MaxL, MeanV, and EntrV) seems to enhance the ability to complete tasks more rapidly, which may be crucial when managing complex and demanding tasks. However, these metrics do not directly influence the accuracy of responses, indicating that while quicker responses are achievable with effective attention management, they do not necessarily translate to improved accuracy. If MultiMatch and/or MdCRQA these metrics are calculated in real-time or close to real-time, it can serve as a barometer for identifying: (a) the workload level (b) and predict whether performance is deteriorating or not. Designers might consider establishing all or a subset of these metrics' thresholds as part of adaptive interfaces that could provide users with information about their cognitive states and performance or adjust the task load or the way the information is presented accordingly.

While this finding opens possibilities for real-time applications, it is essential to acknowledge that not all MultiMatch and MdCRQA metrics exhibited strong significant correlations with performance measures. The findings here challenge the notion that only long simultaneous coordination is needed for effective collaboration (i.e., measured by Position, L, MaxL, and not by MaxV). Instead, the findings as a whole here show the importance of the temporal dimension of shared attention that is considered in the MeanV metric. The findings show that teams spending more time looking at the same AOIs may result in tasks completed faster, even if they are not looking at the AOI in the same way. From an applied context, designers may watch to find ways to encourage teams to collectively look at the same AOIs over time, especially during high workload periods.

Overall, this research validates MultiMatch metrics such as Direction and Duration and MdCRQA metrics such as L, MaxL, and MeanV and highlights the need for research to inform the design of displays by integrating methods where teammates can know where their teammates are looking such as gaze sharing (Neider et al., 2010; Siirtola et al., 2019). For example, D'Angelo and Begel (2017) developed a system where a pair of programmers were shown what the other was looking at while they worked, and they found providing this shared gaze information aids in coordination and effective communication. Moreover, Akkil et al. (2016) developed a shared gaze interface called GazeTorch which facilitated the collaboration in physical tasks. Several other studies found that shared gaze improved performance and remote collaboration in several domains (G. Lee et al., 2017; Schneider et al., 2013; Trösterer et al., 2015).

In conclusion, our research findings provide a significant contribution to the field by offering a comprehensive set of metrics (1) which is sensitive to workload changes and (2) correlates with performance measures in high workload conditions. Additionally, our quantitative and qualitative analyses reveal patterns of redundant overlaps in visual attention between teammates, especially in high workload conditions, suggesting inefficiencies in coordination that could impact performance. Understanding these overlaps and their implications further emphasizes the need for interventions that can enhance team coordination. This work then paves the way for research on gaze sharing which is the focus of chapters 4, 5, and 6 of this dissertation.

3.8 References

Abarbanel, H. D. I., Brown, R., Sidorowich, J. J., & Tsimring, L. S. (1993). The Analysis of Observed Chaotic Data in Physical Systems. *Reviews of modern physics*, *65*(4), 1331.

Akkil, D., James, J. M., Isokoski, P., & Kangas, J. (2016). GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. *Conference on Human Factors in Computing Systems - Proceedings*, 1151–1158. https://doi.org/10.1145/2851581.2892459

Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A Comparison of Scanpath Comparison Methods. *Behavior Research*, *47*, 1377–1392. https://doi.org/10.3758/s13428-014-0550-3

Atweh, J. A., Hayek, J. A., & Riggs, S. L. (2023). Quantifying Visual Attention of Teams During Workload Transitions Using AOI-Based Cross-Recurrence Metrics. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *67*(1), 1882–1887. https://doi.org/10.1177/21695067231193683

Atweh, J. A., Moacdieh, N. M., & Riggs, S. L. (2022). Identifying Individual-, Team-, and Organizational-Level Factors that Affect Team Performance in Complex Domains Based on Recent Literature. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1795–1799. https://doi.org/10.1177/1071181322661213

Atweh, J. A., & Riggs, S. L. (2024). ScanMatch Versus MultiMatch: A Comparison of the Sensitivity of Scanpath Similarity Metrics to Changes in Workload. *2024 IEEE Systems and Information Engineering Design Symposium (SIEDS)*, 57–62. https://doi.org/10.1109/SIEDS61124.2024.10534720

Atweh, J. A., & Riggs, S. L. (2025a). Display Design Shapes How Eye Tracking and Workload Predict Team Performance. *2025 Systems and Information Engineering Design Symposium (SIEDS)*.

Atweh, J. A., Tabbara, I., Nasrallah, G., & Riggs, S. L. (2024). Reducing Barriers in Analyzing Eye Tracking Data: The Development of a GUI to Preprocess Eye Tracking

Data. *2024 IEEE Systems and Information Engineering Design Symposium (SIEDS)*, 286–291. https://doi.org/10.1109/SIEDS61124.2024.10534671

Atweh, J. A., & Riggs, S. L. (2025b). Predicting Mental Demand of Teammates Using Eye Tracking Metrics: A Machine Learning Approach. *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*. https://doi.org/10.1145/3715669.3726802

Byrd, C. J., Johnson, J. S., Radcliffe, J. S., Craig, B. A., Eicher, S. D., & Lay, D. C. (2020). Nonlinear Analysis of Heart Rate Variability for Evaluating the Growing Pig Stress Response to an Acute Heat Episode. *Animal*, *14*(2), 379–387. https://doi.org/10.1017/S1751731119001630

Chelidze, T., & Matcharashvili, T. (2015). Dynamical patterns in seismology. In C. L. J. Webber & N. Marwan (Eds.), *Recurrence quantification analysis* (pp. 225–248). Springer. https://doi.org/10.1007/978-3-319-07155-8_10

Cherubini, M., Nüssli, M.-A., & Dillenbourg, P. (2010). This Is It! : Indicating and Looking in Collaborative Work at Distance. *Journal of Eye Movement Research*, *3*. https://doi.org/10.16910/jemr.3.5.3

Coco, M. I., Mønster, D., Leonardi, G., Dale, R., & Wallot, S. (2021). Unidimensional and Multidimensional Methods for Recurrence Quantification Analysis with CRQA. *The R Journal*, *13*(1). https://arxiv.org/abs/2006.01954

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive Team Cognition. *Cognitive Science*, *37*, 255–285. https://doi.org/10.1111/cogs.12009

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A Novel Method for Comparing Fixation Sequences. *Behavior Research Methods*, *42*(3), 692–700.

Crowe, E. M., Gilchrist, I. D., & Kent, C. (2018). New Approaches to the Analysis of Eye Movement Behaviour Across Expertise While Viewing Brain MRIs. *Cognitive Research: Principles and Implications*, *3*, 12. https://doi.org/10.1186/s41235-018-0097-4

Dale, R., & Spivey, M. J. (2006). Unraveling the Dyad: Using Recurrence Analysis to Explore Patterns of Syntactic Coordination Between Children and Caregivers in Conversation. *Language Learning*, *56*(3), 391–430. https://doi.org/10.1111/j.1467-9922.2006.00372.x

D'Angelo, S., & Begel, A. (2017). Improving Communication Between Pair Programmers Using Shared Gaze Awareness. *Conference on Human Factors in Computing Systems - Proceedings*, 6245–6255. https://doi.org/10.1145/3025453.3025573

Day, E. L., Tang, H., Kendhammer, L. K., & Pienta, N. J. (2018). Sequence Analysis: Use of Scanpath Patterns for Analysis of Students' Problem-Solving Strategies. In *Eye tracking for the chemistry education researcher* (pp. 73–97). American Chemical Society. https://doi.org/10.1021/bk-2018-1292.ch005

Dekker, S., & Nyce, J. (2004). How Can Ergonomics Influence Design? Moving from Research Findings to Future Systems. *Ergonomics*, *47*(15), 1624–1639.

Dempere-Marco, L., Hu, X., MacDonald, S. L. S., Ellis, S. M., Hansell, D. M., & Yang, G. Z. (2002). The Use of Visual Search for Knowledge Gathering in Image Decision Support. *IEEE Transactions on Medical Imaging*, *21*(7), 741–754. https://doi.org/10.1109/TMI.2002.801153

Devlin, S. P., Flynn, J. R., & Riggs, S. L. (2019). Examining the Visual Attention of Pairs of Operators during a Low to High Workload Change. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 201–205. https://doi.org/10.1177/1071181319631168

Devlin, S. P., Moacdieh, N. M., Wickens, C. D., & Riggs, S. L. (2020). Transitions Between Low and High Levels of Mental Workload can Improve Multitasking Performance. *IISE Transactions on Occupational Ergonomics and Human Factors*, *8*(2), 72–87. https://doi.org/10.1080/24725838.2020.1770898

Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It Depends On How you Look At It: Scanpath Comparison in Multiple Dimensions With MultiMatch, a Vector-Based Approach. *Behavior Research Methods*, *44*(4), 1079–1100. https://doi.org/10.3758/s13428-012-0212-2

Dindar, M., Jarvela, S., Ahola, S., Huang, X., & Zhao, G. (2022). Leaders and Followers Identified by Emotional Mimicry During Collaborative Learning: A Facial Expression Recognition Study on Emotional Valence. *IEEE Transactions on Affective Computing*, *13*(3), 1390–1400. https://doi.org/10.1109/TAFFC.2020.3003243

Eraslan, S., Yesilada, Y., & Harper, S. (2017). Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis. *Journal of Eye Movement Research*, *10*(4), 10–16910. https://doi.org/10.16910/jemr.10.4.6

Fahimi, R., & Bruce, N. D. (2021). On Metrics for Measuring Scanpath Similarity. *Behavior Research Methods*, *53*, 609–628. https://doi.org/10.3758/s13428-020-01441-0

Fan, L., Wang, J., Li, Q., Song, Z., Dong, J., Bao, F., & Wang, X. (2023). Eye Movement Characteristics and Visual Fatigue Assessment of Virtual Reality Games with Different Interaction Modes. *Frontiers in Neuroscience*, *17*, 1173127. https://doi.org/10.3389/fnins.2023.1173127

Feitshans, G., Rowe, A., Davis, J., Holland, M., & Berger, L. (2008). Vigilant Spirit Control Station (VSCS): The Face of COUNTER. *AIAA Guidance, Navigation and Control Conference and Exhibit*. https://doi.org/10.2514/6.2008-6309

Foulsham, T., Dewhurst, R., Nyström, M., Jarodzka, H., Johansson, R., Underwood, G., & Holmqvist, K. (2012). Comparing Scanpaths During Scene Encoding and Recognition: A Multi-Dimensional Approach. *Journal of Eye Movement Research*, *5*. https://doi.org/10.16910/jemr.5.4.3

Frame, M. E., Warren, R., & Maresca, A. M. (2019). Scanpath Comparisons for Complex Visual Search in a Naturalistic Environment. *Behavior Research*, *51*, 1454–1470. https://doi.org/10.3758/s13428-018-1154-0

Gbadamosi, J., & Zangemeister, W. H. (2001). Visual Imagery in Hemianopic Patients. *Journal of Cognitive Neuroscience*, *13*(7), 855–866. https://doi.org/10.1162/089892901753165782

Godavarthi, V., Pawar, S. A., Unni, V. R., Sujith, R. I., Marwan, N., & Kurths, J. (2018). Coupled Interaction Between Unsteady Flame Dynamics and Acoustic Field in a Turbulent Combustor. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(11), 113111. https://doi.org/10.1063/1.5052210

Gorman, J. C., Cooke, N. J., Amazeen, P. G., & Fouse, S. (2012). Measuring Patterns in Team Interaction Sequences Using a Discrete Recurrence Approach. *Human Factors*, *54*(4), 503–517. https://doi.org/10.1177/0018720811426140

Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2020). Measuring Real-Time Team Cognition During Team Training. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *62*, 825–860. https://doi.org/10.1177/0018720819852791

Hajari, N., Cheng, I., Zheng, B., & Basu, A. (2016). Determining Team Cognition from Delay Analysis Using Cross Recurrence Plot. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3482–3485. https://doi.org/10.1109/EMBC.2016.7591478

Hill, Y., Den Hartigh, R. J., Cox, R. F., De Jonge, P., & Van Yperen, N. W. (2020). Predicting Resilience Losses in Dyadic Team Performance. *Nonlinear dynamics, psychology, and life sciences*, *24*(3), 327–351.

Hinkle, D., Wiersma, W., & Jurs, S. (2003). *Applied Statistics for the Behavioral Sciences* (5th ed.).

Iwanski, J. S., & Bradley, E. (1998). Recurrence Plots of Experimental Data: To Embed or Not to Embed? *Chaos*, *8*(4), 861–871. https://doi.org/10.1063/1.166372

Jraidi, I., Chaouachi, M., Ben Khedher, A., Lajoie, S. P., & Frasson, C. (2022). Understanding Clinical Reasoning through Visual Scanpath and Brain Activity Analysis. *Computation*, *10*(8), 130. https://doi.org/10.3390/computation10080130

Kang, Z., & Landry, S. J. (2014). Using Scanpaths as a Learning Method for a Conflict Detection Task of Multiple Target Tracking [Original work published 2014]. *Human Factors*, *56*(6), 1150–1162. https://doi.org/10.1177/0018720814523066

Kantz, H., & Schreiber, T. (2003). *Nonlinear Time Series Analysis*. Cambridge University Press.

Król, M., & Król, M. (2019). Scanpath Similarity Measure Reveals Not Only a Decreased Social Preference, But Also an Increased Nonsocial Preference in Individuals With Autism. *Autism*. https://doi.org/10.1177/1362361319865809

Lee, G., Kim, S., Lee, Y., Dey, A., Piumsomboon, T., Norman, M., & Billinghurst, M. (2017). Improving Collaboration in Augmented Video Conference Using Mutually Shared Gaze. *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, ICAT-EGVE 2017*, 197–204. https://doi.org/10.2312/egve.20171359

Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., & van Merriënboer, J. J. G. (2020). Different Effects of Pausing on Cognitive Load in a Medical Simulation Game. *Computers in Human Behavior*, *110*, 106385. https://doi.org/10.1016/j.chb.2020.106385

Lin, Y., Zhang, W. J., & Koubek, R. J. (2004). Effective Attention Allocation Behavior and Its Measurement: A Preliminary Study. *Interacting with Computers*, *16*(6), 1195–1210. https://doi.org/10.1016/j.intcom.2004.08.006

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing Another Person's Eye Movements Improves Identification of Pulmonary Nodules in Chest X-ray Inspection. *Journal of Experimental Psychology. Applied*, *16*(3), 251–262. https://doi.org/10.1037/a0020082

Lopes, F., Leal, A., Pinto, M. F., Dourado, A., Schulze-Bonhage, A., Dümpelmann, M., & Teixeira, C. (2023). Removing Artefacts and Periodically Retraining Improve Performance of Neural Network-Based Seizure Prediction Models. *Scientific Reports*, *13*(1), 5918. https://doi.org/10.1038/s41598-023-30864-w

Luo, H., Bo, L., Peng, C., & Hou, D. (2023). Detection and Quantification of Oil Whirl Instability in a Rotor-Journal Bearing System Using a Novel Dynamic Recurrence Index. *Nonlinear Dynamics*, *111*(3), 2229–2261. https://doi.org/10.1007/s11071-022-07932-3

Ma, E., Dixit, P., Han, A., Marsano, N., Sparks, B., Sun, A., & Papoutsaki, A. (2025). EyeDraw: Investigating the Perceived Effects of Shared Gaze on Remote Collaborative Drawing. *Proceedings of the ACM on Human-Computer Interaction*, *9*(1), 1–24. https://doi.org/10.1145/3701196

Madsen, A., Larson, A., Loschky, L., & Rebello, N. S. (2012). Using ScanMatch Scores to Understand Differences in Eye Movements Between Correct and Incorrect Solvers on Physics Problems. *Eye Tracking Research and Applications Symposium (ETRA)*, 193–196. https://doi.org/10.1145/2168556.2168591

Marghitu, D. B., Beale, D. G., & Sinha, S. C. (1997). Dynamic Stability of Vehicle Systems Via Poincaré Maps. *Vehicle System Dynamics*, *28*(1), 41–55.

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence Plots for the Analysis of Complex Systems. *Physics Reports*, *438*(5–6), 237–329. https://doi.org/10.1016/j.physrep.2006.11.001

Marwan, N., & Kurths, J. (2002). Nonlinear Analysis of Bivariate Data with Cross Recurrence Plots. *Physics Letters A*, *302*, 299–307. https://doi.org/10.1016/S0375-9601(02)01170-2

McKendrick, R., Shaw, T., de Visser, E., Saqer, H., Kidwell, B., & Parasuraman, R. (2014). Team Performance in Networked Supervisory Control of Unmanned Air Vehicles: Effects of Automation, Working Memory, and Communication Content. *Human Factors*, *56*(3), 463–475. https://doi.org/10.1177/0018720813496269

Needleman, S. B., & Wunsch, C. D. (1970). A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *Journal of Molecular Biology*, *48*, 443–453.

Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. G. J. (2010). Coordinating Spatial Referencing using Shared Gaze. *Psychonomic Bulletin and Review*, *17*(5), 718–724. https://doi.org/10.3758/PBR.17.5.718

Newport, R. A., Russo, C., Al Suman, A., & Di Ieva, A. (2021). Assessment of Eye-Tracking Scanpath Outliers Using Fractal Geometry. *Heliyon*, *7*(7), e07616. https://doi.org/10.1016/j.heliyon.2021.e07616

Nyström, M., & Holmqvist, K. (2010). An Adaptive Algorithm for Fixation, Saccade, and Glissade Detection in Eyetracking Data. *Behavior Research Methods*, *42*(1), 188–204. https://doi.org/10.3758/BRM.42.1.188

Ozturk, U. (2022). Prediction of Actual Passing Distance for Path Planning of Unmanned Surface Ships. *OCEANS 2022, Hampton Roads*, 1–7. https://doi.org/10.1109/OCEANS47191.2022.9977368

Pietinen, S., Bednarik, R., & Tukiainen, M. (2010). Shared Visual Attention in Collaborative Programming: A Descriptive Analysis. *Proceedings of the 2010 ICSE Workshop on Cooperative and Human Aspects of Software Engineering*.

Poincaré, H. (1890). Sur Le Problème Des Trois Corps et Les Équations De La Dynamique. *Acta mathematica*, *13*(1), A3–A270.

Prytz, E. G., & Scerbo, M. W. (2015). Changes in Stress and Subjective Workload Over Time Following a Workload Transition. *Theoretical Issues in Ergonomics Science*, *16*(6), 586–605. https://doi.org/10.1080/1463922X.2015.1084397

Richardson, D. C., & Dale, R. (2005). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science*, *29*(6), 1045–1060. https://doi.org/10.1207/s15516709cog0000_29

Russell, S. M., Funke, G. J., Knott, B. A., & Strang, A. J. (2012). Recurrence Quantification Analysis Used to Assess Team Communication in Simulated Air Battle Management. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 468–472. https://doi.org/10.1177/1071181312561046

Saeed, B., Ullah, J. S., Aziz, M., Shahid, M., Siraj, A. S., & Tariq, S. R. (2023). Comparison of Recurrence Rate of Wrist Ganglion Between Seton and Open Surgical Excision in Military Hospitals of Pakistan. *Pakistan Armed Forces Medical Journal*, *73*(1), 50–53. https://doi.org/10.51253/pafmj.v73i1.7045

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*, 540–547. https://doi.org/10.1518/001872008X288457

Schneider, B., Abu-El-Haija, S., Reesman, J., & Pea, R. (2013). Toward Collaboration Sensing: Applying Network Analysis Techniques to Collaborative Eye-tracking Data. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*.

Siirtola, H., Špakov, O., Istance, H., & Räihä, K. J. (2019). Shared Gaze in Collaborative Visual Search.

Silva, A., Emmanuel, G., McClain, J. T., Matzen, L., & Forsythe, C. (2015). Measuring Expert and Novice Performance Within Computer Security Incident Response Teams. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9183*, 144–152. https://doi.org/10.1007/978-3-319-20816-9_15

Špakov, O., Istance, H., Räihä, K. J., Viitanen, T., & Siirtola, H. (2019). Eye Gaze and Head Gaze in Collaborative Games. *Eye Tracking Research and Applications Symposium (ETRA)*. https://doi.org/10.1145/3317959.3321489

Starke, S. D., & Baber, C. (2018). The Effect of Four User Interface Concepts on Visual Scan Pattern Similarity and Information Foraging in a Complex Decision Making Task. *Applied Ergonomics*, *70*, 6–17. https://doi.org/10.1016/j.apergo.2018.01.010

Stranc, S., & Muldner, K. (2020). Scanpath Analysis of Student Attention During Problem Solving with Worked Examples. https://doi.org/10.1007/978-3-030-52240-7_56

Tolston, M. T., Strang, A. J., Funke, G. J., Miller, B., Brown, R., & Menke, L. (2016). Evaluating the Relationship between Team Performance and Joint Attention with Longitudinal Multivariate Mixed Models. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 740–744. https://doi.org/10.1177/1541931213601170

Trösterer, S., Gärtner, M., Wuchse, M., Maurer, B., Baumgartner, A., Meschtscherjakov, A., & Tscheligi, M. (2015). Four Eyes See More Than Two: Shared Gaze in the Car. *IFIP Conference on Human-Computer Interaction*, 331–348. https://doi.org/10.1007/978-3-319-22668-2_26

Veerabhadrappa, R., Hettiarachchi, I. T., Hanoun, S., Jia, D., Hosking, S. G., & Bhatti, A. (2024). Evaluating Operator Training Performance Using Recurrence Quantification Analysis of Autocorrelation Transformed Eye Gaze Data. *Human Factors*, *66*(3), 818–838. https://doi.org/10.1177/00187208221116953

Verdiere, K. J., Albert, M., Dehais, F., & Roy, R. N. (2020). Physiological Synchrony Revealed by Delayed Coincidence Count: Application to a Cooperative Complex Environment. *IEEE Transactions on Human-Machine Systems*, *50*(5), 395–404. https://doi.org/10.1109/THMS.2020.2986417

Villamor, M., & Rodrigo, M. M. (2017). Impact of Both Prior Knowledge and Acquaintanceship on Collaboration and Performance: A Pair Program Tracing and Debugging Eye-Tracking Experiment. *Eye Tracking and Pair Programming View project Impact* *of Both Prior Knowledge and Acquaintanceship on Collaboration and Performance.* https://www.researchgate.net/publication/318907797

Villamor, M., & Rodrigo, M. M. (2018). Predicting Successful Collaboration in a Pair Programming Eye Tracking Experiment. *UMAP 2018 - Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, 263–268. https://doi.org/10.1145/3213586.3225234

Villamor, M., & Rodrigo, M. M. (2022). Predicting Pair Success in a Pair Programming Eye Tracking Experiment Using Cross-Recurrence Quantification Analysis. *APSIPA Transactions on Signal and Information Processing*, *11*. https://doi.org/10.1561/116.00000031

Wallot, S. (2019). Multidimensional Cross-Recurrence Quantification Analysis (MdCRQA) – A Method for Quantifying Correlation between Multivariate Time-Series. *Multivariate Behavioral Research*, *54*(2), 173–191. https://doi.org/10.1080/00273171.2018.1512846

Wallot, S., & Leonardi, G. (2018). Analyzing Multivariate Dynamics using Cross-Recurrence Quantification Analysis (CRQA), Diagonal-Cross-Recurrence Profiles (DCRP), and Multidimensional Recurrence Quantification Analysis (MdRQA)–A Tutorial in R. *Frontiers in Psychology*, *9*, 2232. https://doi.org/10.3389/fpsyg.2018.02232

Wallot, S., Roepstorff, A., & Mønster, D. (2016). Multidimensional Recurrence Quantification Analysis (MdRQA) for the Analysis of Multidimensional Time-Series: A Software Implementation in MATLAB and its Application to Group-Level Data in Joint Action. *Frontiers in Psychology*, *7*, 1835. https://doi.org/10.3389/fpsyg.2016.01835

Webber, C. L. J., & Zbilut, J. P. (2005). Recurrence Quantification Analysis of Nonlinear Dynamical Systems. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, 142–177.

Wickens, C. D., & Alexander, A. L. (2009). Attentional Tunneling and Task Management in Synthetic Vision Displays. *The International Journal of Aviation Psychology*, *19*(2), 182–199. https://doi.org/10.1080/10508410902766549

Wilson, K. A., Heinselman, P. L., & Kang, Z. (2018). Comparing Forecaster Eye Movements during the Warning Decision Process. *Weather and Forecasting*, *33*, 501–521. https://doi.org/10.1175/WAF-D-17-0119.1

Wohleber, R. W., Matthews, G., Lin, J., Szalma, J. L., Calhoun, G. L., Funke, G. J., Chiu, C.-Y. P., & Ruff, H. A. (2019). Vigilance and Automation Dependence in Operation of Multiple Unmanned Aerial Systems (UAS): A Simulation Study. *Human Factors*, *61*(3), 488–505. https://doi.org/10.1177/0018720818799468

Yang, Q., Wang, T., Su, N., Xiao, S., & Kapoula, Z. (2013). Specific Saccade Deficits in Patients with Alzheimer's Disease at Mild to Moderate Stage and in Patients with Amnestic Mild Cognitive Impairment. *Age*, *35*, 1287–1298. https://doi.org/10.1007/s11357-012-9420-z

Zheng, B. (2022). Team Cognition Assessment From Concept to Practice. In *Brain and cognitive intelligence control in robotics* (pp. 47–69, Vol. 1).

Zheng, B., Hajari, N., & Atkins, M. (2016). Revealing Team Cognition from Dual Eye-tracking in the Surgical Setting. *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. https://doi.org/10.1145/2857491.2884062

Chapter 4

Effect of Real-Time Gaze Sharing Visualizations on Team Performance, Workload, and Situation Awareness

4.1 Introduction

The ability to work together effectively is essential for success in many dynamic and data-rich domains, such as aviation, military, and healthcare (Gorman et al., 2020). As discussed in Chapter 2, one challenge associated with teamwork is maintaining shared situation awareness (SA)—i.e., the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley, 1988). It is estimated that 88% of accidents and incidents can be attributed to lapses in SA (Endsley, 1995b). Given that poor SA has been a major contributor to human error, it poses a threat to safety-critical systems. Thus, improving SA can potentially enhance teamwork and performance (Endsley, 1995a). To date, research has mainly focused on developing design recommendations to support team performance (Atweh & Riggs, 2024; Berger et al., 2023; El Iskandarani et al., 2023; Walton & Gilbert, 2022) with less attention on supporting team and shared SA.

Knowing where a team member is looking when viewing a display can potentially help teams to understand each other's focus of attention and improve coordination. To date, there has been limited work on gaze sharing—i.e., visualizing in real-time, where team members are looking on a shared display. Real-time gaze sharing can be a valuable tool for teams looking to improve their communication, collaboration, and performance (Atkinson et al., 2023). Successful team coordination relies on synchronized actions and having shared mental models between team members (Cannon-Bowers et al., 1993). Shared mental models represent a shared understanding among team members that emerges from these collective cognitive efforts that involve how team members perceive, interpret, and store information (Fiore & Salas, 2004).

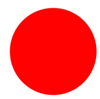
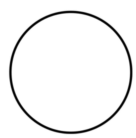
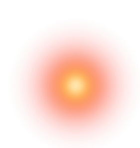
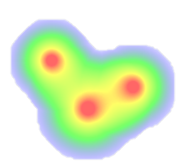
4.2 Background

Gaze sharing is a form of non-verbal communication used to exchange visual information (i.e., eye movements) between teammates as it allows partners to view each other's gaze points on their respective displays while simultaneously completing their tasks (Sung et al., 2021). By knowing where others are looking, operators can potentially better coordinate their activities, identify issues, and respond more effectively to dynamic situations. In UAV operations, where teams may be either co-located or remote, gaze sharing could be an important tool for the safety and efficiency of missions. For instance, in a scenario where a UAV is navigating through a complex environment, knowing that a teammate is monitoring a specific area of the screen can help another operator focus on complementary tasks, thus optimizing overall team performance.

Gaze sharing can be facilitated through various technological and procedural mechanisms. These include eye tracking systems, shared visual displays, and augmented reality interfaces (Jing et al., 2022; Zhang et al., 2017). This overlay of teammates' gaze points onto individual displays employs diverse visualization methods, including dots, spotlights, heatmaps, and fixation trails. Each of these mechanisms could play a role in enhancing the visibility of an operator's gaze direction to their teammates. Eye tracking technology is one of the most direct ways to implement gaze sharing. These systems use cameras and sensors to monitor the eye movements of operators and can display this information on shared screens or individual monitors. Eye tracking systems can indicate where each team member is looking in real-time, allowing for seamless coordination without verbal communication (Bulling & Gellersen, 2010).

Table 4.1 provides a description and image of the most common visualization techniques. Early work in gaze sharing suggests a potentially better way to communicate spatial information that utilizes deictic cues, such as cursor movements (Brennan et al., 2008). A commonly used visualization technique is the coordinate-based dot symbol that has mostly been adopted for its simplicity and minimal added mental load (Y. Li et al., 2019). Jing et al. (2022) studied gaze behavior to detect overlapping gaze between teammates where overlapping gaze points turned green to inform the users of when mutual gazes occurred. This technique allowed players to accordingly modify their strategies based on their partner's current area of interest. Other choices have been adopted by researchers which include heatmaps (Zhang et al., 2017). A heatmap represents the distribution of gaze points across a display and provides a visual representation of the areas of the display that the teammates focus on the most (Špakov & Miniotas, 2007). Another method involves trajectory-based visualizations to maintain the user's past gaze history for a specific time span. For example, the fixation trail presents the trajectory of gaze points over time, allowing the teammates to see the eye movement patterns of their team members.

Table 4.1: The type, description, visualization form, advantages, and disadvantages of existing gaze sharing visualization techniques.

Type	Description	Visualization Form	Advantages	Disadvantages
Dot	Translucent colored dot		- High precision in translating users' gaze. - Allows users to process visual information quickly.	- Can be potentially distracting. - Quick jittery movements. - High visual information: amount of gaze data.
Cursor	Large ring with transparent center		- Equivalent to size of foveal vision. - Provides more visual info (what's enclosed inside).	- Can be potentially distracting. - Quick jittery movements. - Low visibility
Spotlight	A circle with high opacity in the center and decreasing opacity towards the edges		- Less distracting - Informative	- Low visibility (cannot see through it). - Quick jittery movements
Fixation Trail	Uses fixation data to create a trail of fixations		- Users know previous fixation states - Moderate visibility	- Can be potentially distracting
Heatmap	Represents the distribution of gaze points with a color gradient		- Informative	- Occludes parts of the display

Selecting which gaze sharing technique to use depends on practicality, ease of implementation, and being able to convey relevant information without being distracting or overwhelming. Another factor to consider is the precision of the eye tracking data. This is especially important in tasks where fine-grained analysis of eye movements is required, such as in reading or visual search tasks (He et al., 2023). Inaccuracies in the visualization can lead to misinterpretation of the user's gaze behavior. Table 1 also includes the advantages and disadvantages of current gaze sharing visualizations.

In co-located control centers, large shared displays can be used to project the visual attention of each operator. For example, a multi-operator UAV control center might have a central screen showing the UAV's real-time video feed with overlaid indicators representing where each operator is currently looking. This shared visual context helps all team members to stay aware of each other's focus areas, especially in unexpected events (Brennan et al., 2008; Di Gregorio et al., 2021). For remote teams, augmented reality (AR) can provide an immersive means of gaze sharing (Bai et al., 2020). AR headsets can also potentially display the gaze directions of remote team members within the operator's field of view. This approach allows for a more integrated experience, where gaze cues are overlaid on the operator's normal visual environment, enhancing the sense of presence and coordination among team members (Huang et al., 2020).

To date, gaze sharing has been studied using relatively simple tasks (e.g., Lee et al., 2017) and little is known about its attentional costs and effect on performance in more complex, dynamic work environments. Gaze sharing has been shown to improve performance and efficiency within teams, with simple tasks that includes assembling jigsaw puzzles (Lee et al., 2017), finding the letter "O" amongst different letters (Zhang et al., 2017), building Lego models (Gupta et al., 2016), and building a circuit (Akkil et al., 2016). Schneider and Pea (2017) conducted an experiment to investigate the impact of gaze sharing on collaborative learning. They found that gaze sharing helped students achieve a higher quality of collaboration and resulted in better learning outcomes.

Despite its potential benefits, the effectiveness and role of gaze sharing in UAV C2 operations remain underexplored. It is unclear whether gaze sharing can truly complement or replace traditional communication methods and how its utility might vary depending on different communication and task strategies. For instance, in some scenarios, gaze sharing might significantly reduce the need for verbal exchanges, while in others, it may serve only as a supplementary tool that enhances verbal communication. Therefore, this chapter aims to address this research gap by examining whether gaze sharing can improve team SA, decrease workload, and thus enhance team performance in a simulated complex, data-rich domain. By addressing this gap, we can gain a deeper understanding of gaze sharing and provide valuable insights for the design of real-time, adaptive systems in safety-critical domains.

While it is recognized in the literature that gaze sharing can be both helpful and challenging, its effectiveness and application in dynamic and data-rich domains remains underexplored. To bridge this gap, this study assesses team SA in a simulated complex, data-rich domain—i.e., UAV command and control—using two promising gaze sharing visualization techniques: dot and trail. To evaluate the effectiveness of the gaze sharing visualizations, we conducted a study to understand its effect on eye tracking metrics, workload, team SA measures overall and for each level of SA, and performance. We expect that our gaze visualizations will support team SA and decrease workload which in turn will be a catalyst to improve team performance (Jing et al., 2022; Y. Li et al., 2019).

4.3 Methods

4.3.1 Participants

This study was approved by the University of Virginia's Institutional Review Board (protocol number #3480). Thirty-five teams (70 participants) from the University of Virginia were recruited for the study (M = 24.46 years, SD = 4.36 years). Each team included one male and one female who were not acquainted with one another.

4.3.2 Experimental Setup

The setup was the same setup in Chapter 3. The only difference is that these collocated teams were separated by a divider (Figure 4.1) .

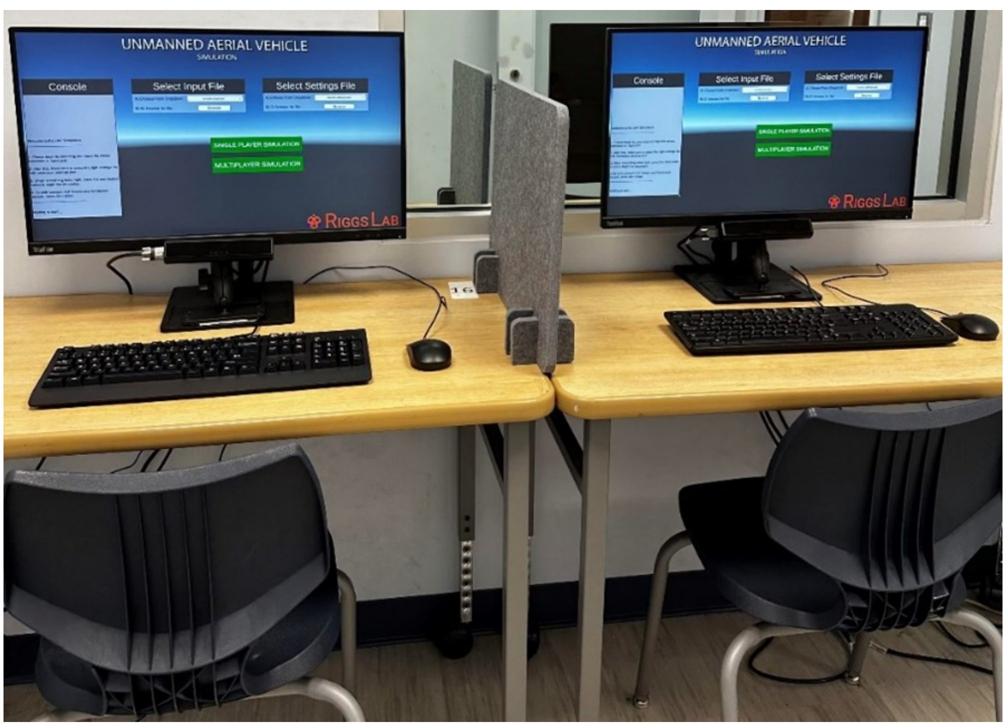


Figure 4.1: Experimental setup with the two networked desktop computers side-by-side with a divider in between.

4.3.3 UAV Tasks

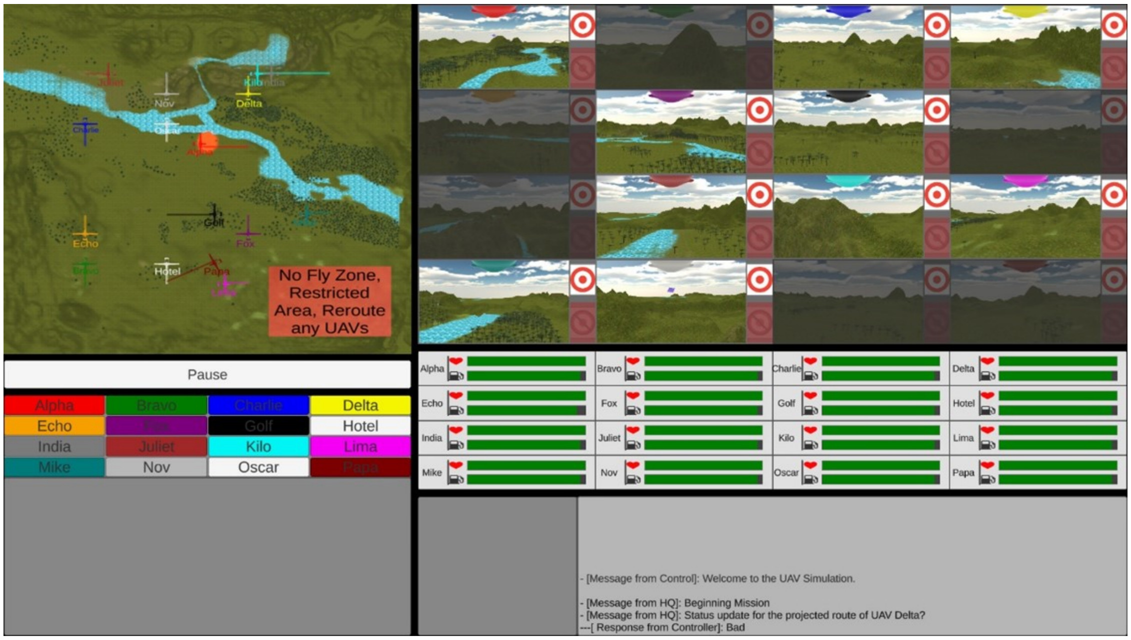
Participants were responsible for the same set of tasks (one primary and three secondary) as Chapter 3.

4.3.4 Experimental Design

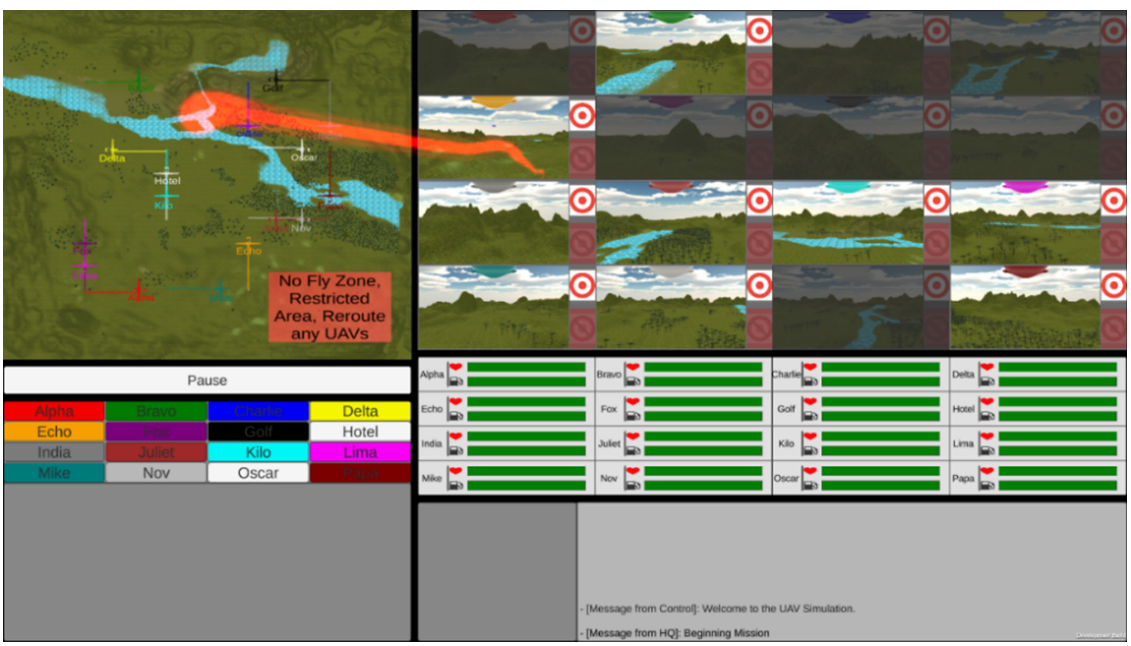
A within-subjects study was conducted, where all teams completed three 15-min scenarios, one for each of the following three visualizations: (a) no gaze sharing, (b) gaze sharing with a real-time fixation dot (Figure 4.2a), and (c) gaze sharing using a real-time fixation trail offering a visual representation of the preceding two seconds (Newn et al., 2017; Figure 4.2b). The order of conditions was counterbalanced across teams to account for order effects. We maintained a consistent number of targets, reroutings, fuel leaks, and chat messages tasks across all conditions. Each instance of a task was randomized within each condition.

Performance & Workload Measures

Performance was measured using both the (a) point system performance metric for overall performance and (b) accuracy for each task. Same point systems for scoring performance was sued here (Table 3.2). The points values were assigned to encourage participants to prioritize certain tasks (i.e., target detection). Each participant provided a subjective workload rating at the end of each trial using the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988).



(a) Dot (see red dot in the map panel)



(b) Fixation Trail

Figure 4.2: The dot (a) and fixation trail (b) gaze sharing visualization techniques.

SA Measures

SAGAT was used to measure the SA of the participants in this study. The 13 query SAGAT used as part of this study included relevant SA queries that had been developed (Endsley, 2021) and others we developed to ensure we assessed all SA levels [level 1 (perception): 4 queries; level 2 (comprehension): 4 queries; level 3 (projection): 5 queries). Table 4.2 shows a sample SAGAT questionnaire for Freeze 2 of the No Gaze Sharing condition. All questions were the same in all freezes across conditions; however, the specific UAVs asked about changed.

SA measures were derived from participants' SAGAT query responses as follows. Each participant's responses were scored as correct (1 point) or incorrect (0 points) based on predefined tolerances (Table 3). The *mean query score* for each team member across multiple freezes was determined by summing the correct responses and dividing by the total number of freezes. *The SAGAT-TSA query score* was then calculated by averaging the mean query scores across team members. The *overall SAGAT-TSA* score for each team and condition was computed by summing the SAGAT-TSA query scores across all queries and dividing by the total number of queries. This approach allowed for a comprehensive evaluation of team SA performance across different freezes and conditions. Notably, the overall SAGAT-TSA scores were computed for each team across all queries and for each SA level, providing a nuanced understanding of TSA dynamics.

In calculating SAGAT-SSA, we focused on the shared tasks identified as the primary task through 13 SAGAT queries. SSA was assessed based on the responses to five specific SAGAT queries related solely to this shared primary task. SSA analysis included instances where both teammates provided identical answers, regardless of whether those answers were correct or incorrect, encompassing scenarios of "Both teammates correct" and "Both teammates incorrect (in the same way)".

Table 4.2: Sample SAGAT questionnaire for Freeze 2 of the "No Gaze Sharing" condition.

Number	Query	Options	SA Level	Scoring Tolerance
1	Indicate a UAV that is active on the map panel	Free response	1	0
2	Indicate a UAV that is inactive on the map panel	Free response	1	0
3	How much fuel is currently remaining for UAV Echo?	Free response	1	$\pm 10\%$
4	Is there a target in UAV Nov?	• Yes • No • Not applicable	1	0
5	Is UAV Oscar currently headed into a NFZ?	• Yes • No • Not applicable	2	0
6	Is UAV Delta currently in the zone of interest?	• Yes • No • Not applicable	2	0
7	How much of the search zone has been completed?	Free response	2	$\pm 10\%$

Continued on next page

Table 4.2 – continued from previous page

Number	Query	Options	SA Level	Scoring Tolerance
8	What is the status of UAV Bravo's health?	• UAV is healthy. • UAV is experiencing a fuel leak, need to refill fuel. • UAV is lost.	2	0
9	Does UAV Fox have enough fuel to complete the assigned route?	• Yes • No • Not applicable	3	0
10	Is UAV Alpha showing signs of UAV loss?	• Yes • No • Not applicable	3	0
11	If no action is performed in the Video Feed Panel, a target will cross:	• UAV Hotel • UAV Juliet • UAV Lima • Not applicable	3	0
12	If no rerouting is performed, UAV Charlie will:	• Go into a NFZ and thus be lost. • Continue its projected route, away from the NFZ. • This UAV is already lost.	3	0

Continued on next page

Table 4.2 – continued from previous page

Number	Query	Options	SA Level	Scoring Tolerance
13	If no action is performed in the Health Panel, UAV Mike will:	• Remain in the mission. • Be lost. • This UAV is already lost.	3	0

Although there are limitations to SAGAT simulation freezes, Endsley (1995a) has found that up to three SAGAT freezes that are less than two minutes during a 20-min experimental trial is reasonable and has no effect on subsequent performance. This notion is supported by Hogg et al. (1993) who showed that there were no significant effects from freezes. Furthermore, studies in air traffic control have also reported that using either SAGAT or SPAM does not affect the workload of control operations (Morgan et al., 2012).

For this study, three SAGAT freezes occurred in each 15-minute scenario. We used a random-number generator to determine the freeze times to occur between 3-13 minutes in each scenario. Any freeze that was closer than 3 minutes to the previous freeze time was eliminated and a new random number was generated. This allowed participants to be engaged in their tasks before the SAGAT was administered, and enough time to get reengaged before any subsequent SAGAT battery is presented (Endsley, 2021). Table 4.3 presents the timing of the SAGAT freezes for each condition.

4.3.5 Eye Tracking Data Analysis

After we gathered the eye tracking data from the FOVIO eye tracker, we filtered the datasets and removed invalid entries (negative, empty, missing, and out-of-bound values). The gaze

Table 4.3: Timing of SAGAT freezes for each condition (minutes:seconds after scenario start).

Condition	SAGAT Freezes		
	Freeze 1	Freeze 2	Freeze 3
No Gaze Sharing (Control)	4:35	8:46	12:49
Fixation Dot	3:22	7:08	10:31
Fixation Trail	3:50	7:42	11:34

data was screened to meet data quality requirements as outlined in ISO/TS 15007-2:2014-09, which states that at most 15% data loss is acceptable. The data loss across all participants and trials was on average 10.5% (SD = 8.1%). We detected fixations and saccades using the code we developed (Atweh, Tabbara, et al., 2024). As mentioned in Chapter 3, this code is used to analyze eye tracking data collected from experimental studies with participants and it serves two main purposes: (1) filtering the eye tracking dataset and (2) detecting fixations and saccades based on Nyström and Holmqvist (2010)'s velocity-based and data-driven adaptive algorithm. The code, implemented in Python, first takes the raw eye tracking files as input, and filters out empty or invalid recordings. Then, it passes the data through the Savitzky-Golay smoothening filter and calculates the angular velocities in preparation for the data-driven iterative algorithm that keeps iterating until the absolute difference between the newly calculated velocity threshold and the previous one converges to less than 1°. The event detection code contains five main steps: peak velocities detection, saccade onset detection, saccade offset detection, fixation detection, and saccades detection based on velocity constraints for saccade detection and spatial and duration constraints for fixation detection. See Atweh, Tabbara, et al. (2024) for more details on the preprocessing and event detection process.

After detecting fixations and saccades, we calculated 6 eye tracking metrics (number of fixations, fixation duration, number of saccades, saccade duration, velocity, and amplitude) for each participant in each trial. We averaged the numbers for each metric per team per condition across the two participants.

4.3.6 Experimental Procedure

Participants read and signed the consent form, completed a pre-experiment questionnaire (Appendix B), and were briefed about the study goals, tasks that needed to be completed as a team, and how the testbed was networked. This meant that the same tasks occurred in real-time on both monitors, but only one teammate would have to complete every instance of a task. Participants were provided with an overview of the NASA-TLX and SAGAT and how it would be administered. They were instructed to answer SAGAT queries as rapidly as possible and to make their best guess even were unsure of their answer (Endsley, 2021). Afterwards, the participants completed a five-minute training session together with a SAGAT freeze 3 minutes and 15 seconds into the session to provide teams an opportunity to practice answering the SAGAT queries before the experimental sessions. Each query was discussed with them, and they were allowed to ask questions if they were unsure of the meaning of any query. By the end of training, participants had to demonstrate they could achieve 70% accuracy across all tasks. The participants were explicitly instructed to collaborate on the primary task. They were also provided with three minutes to discuss anything they deem necessary (e.g., strategies for secondary tasks).

During the experimental sessions, participants could not communicate verbally with each other as we were interested in understanding how gaze sharing techniques would affect performance. Allowing participants to communicate during the study could introduce additional factors such as social dynamics and personal preferences that are unrelated to the research question at hand.

After the training session, the teams completed three 15-minute scenarios. The same tasks appeared at both stations and the actions of each team member were reflected on both workstations, but a participant could not see the cursor movements of their teammate. At the scheduled freeze times, the experimenter would announce “FREEZE” and the participants would then click the pause button to freeze the scenario (Figure 3.5). A black screen then popped up captioned, “The Experiment is Paused. Please wait for the Conductor’s Instructions.” The experimenter gave each participant a copy of the SAGAT questionnaire which they had two minutes to complete (**Note: this was ample time for everyone to complete the questionnaire*). Following the SAGAT freeze, the participants clicked the “Resume” button to resume the scenario. After each scenario, participants individually completed a NASA-TLX questionnaire to assess workload (Appendix A). At the conclusion of the study, participants filled out a debriefing questionnaire (Table C.2 in Appendix C). The experiment session lasted 120-150 minutes and participants were compensated $30 for their time.

4.4 Results

4.4.1 Performance Measures

Figure 4.3 shows the mean and standard error of performance score across the 35 teams for each condition based on the designated scoring convention (Table 4.2). A one-way repeated measures MANOVA was conducted to check for any statistical difference between the performance measures, (1) total score and (2) accuracy measure for each task. A significant multivariate effect was observed for the within-subjects variable (gaze conditions), $F(12{,}126) = 7.053$, $p<.001$; Wilks’ $\Lambda = 0.36$, partial $\eta^2 = 0.4$. Six follow-up repeated measures univariate ANOVAs showed that both the total point score ($F(2{,}68) = 49.87$, $p<.001$, partial $\eta^2 = 0.6$) and accuracy for the rerouting task ($F(2{,}68) = 6.21$, $p = .003$, partial $\eta^2 = 0.15$) were statistically significantly different between the three conditions, using a Bonferroni adjusted

$\alpha = 0.0083$ level by dividing the standard significance of $\alpha = 0.05$ by the number of tests which is in this case six.

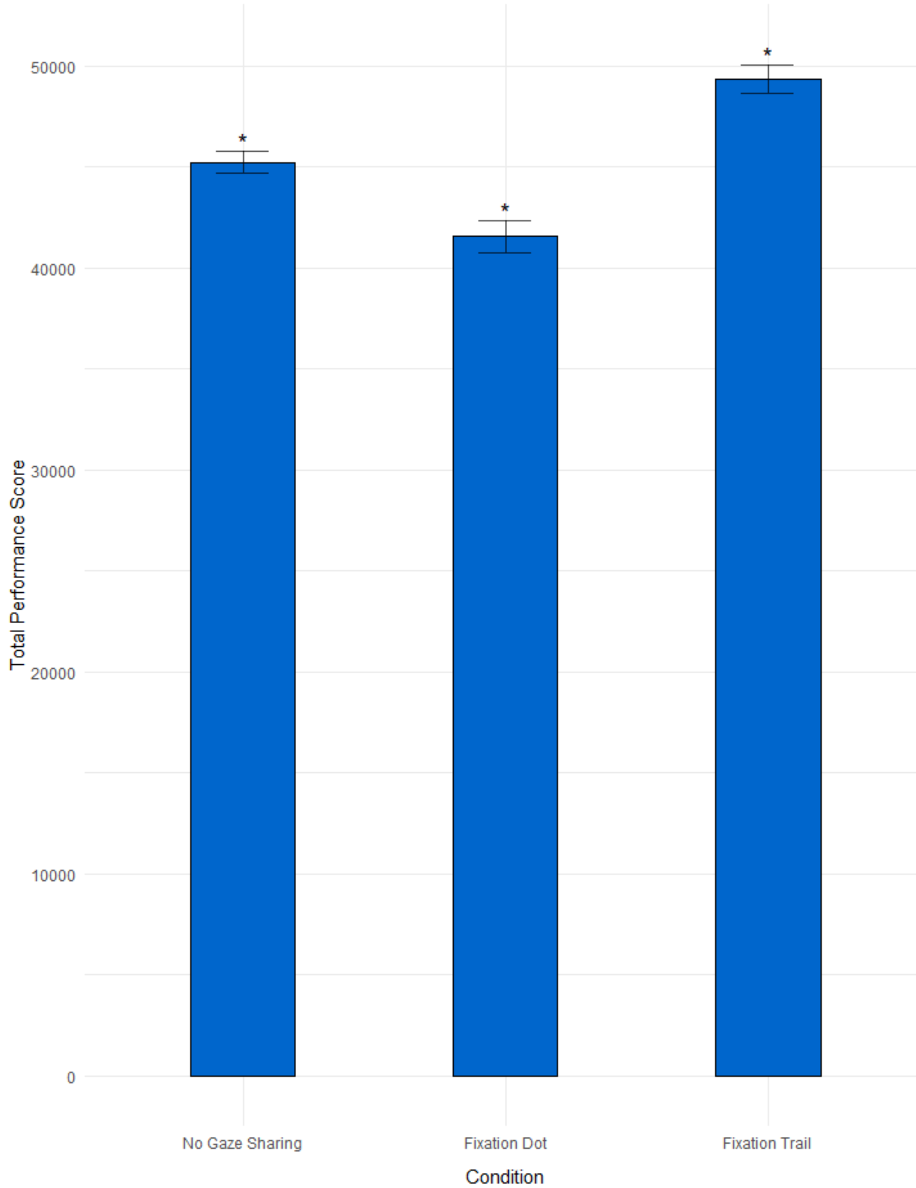


Figure 4.3: Performance scores for each condition. An asterisk (*) indicates significance.

Post hoc tests using Bonferroni correction revealed that all three conditions were significantly different from one another in total score with the fixation trail (mean score = 49,361 points) resulting in the best performance followed by no gaze sharing (mean score = 45,240 points) and the fixation dot (mean score = 41,563 points; Figure 4.3; all p<.001). Moreover, the post hoc tests revealed that the accuracy for the rerouting task was statistically significantly higher using the fixation trail (mean accuracy = 52.1%) compared to the fixation dot (mean accuracy = 28.7%; Figure 4.4; p = .002). There were no statistical differences in terms of score for all other pairwise comparisons (all p>.05).

4.4.2 NASA-TLX

Figure 4.5 shows the mean and standard error of the NASA-TLX scores for each of the six dimensions. We decided to analyze the six dimensions separately based on recent recommendations in the literature (i.e., Bolton et al., 2023). A one-way repeated measures MANOVA was conducted to check for any statistical difference between the NASA-TLX scores of the three conditions across the different dimensions. A significant multivariate effect was observed for the gaze conditions $F(10,268) = 8.32$, $p<.001$; Wilks' $\Lambda = 0.58$, partial $\eta^2 = 0.24$. Six follow-up repeated measures univariate ANOVAs showed that the mental ($F(2,138) = 20.99$, $p<.001$, partial $\eta^2 = 0.23$), performance ($F(2,138) = 7.76$, $p<.001$, partial $\eta^2 = 0.2$), effort ($F(2,138) = 15.92$, $p<.001$, partial $\eta^2 = 0.39$), and frustration ($F(2,138) = 15.92$, $p<.001$, partial $\eta^2 = 0.19$) dimensions were statistically significantly different between the three conditions, using a Bonferroni adjusted α of 0.0083.

For mental demand, effort, and frustration, scores were significantly higher for no gaze sharing compared to the fixation dot and fixation trail (all p<.0083). Moreover, the participants experienced more mental demand and were significantly more frustrated with the fixation dot compared to the fixation trail (all p<.0083). There was no difference between the visualization conditions for the effort dimension.

For the performance dimension, post hoc tests revealed that the participants believed they were significantly more successful in accomplishing the tasks using the fixation trail compared to using the fixation dot (p<.001). There were no statistical differences for in terms of accuracy for all other pairwise comparisons (all p>.05).

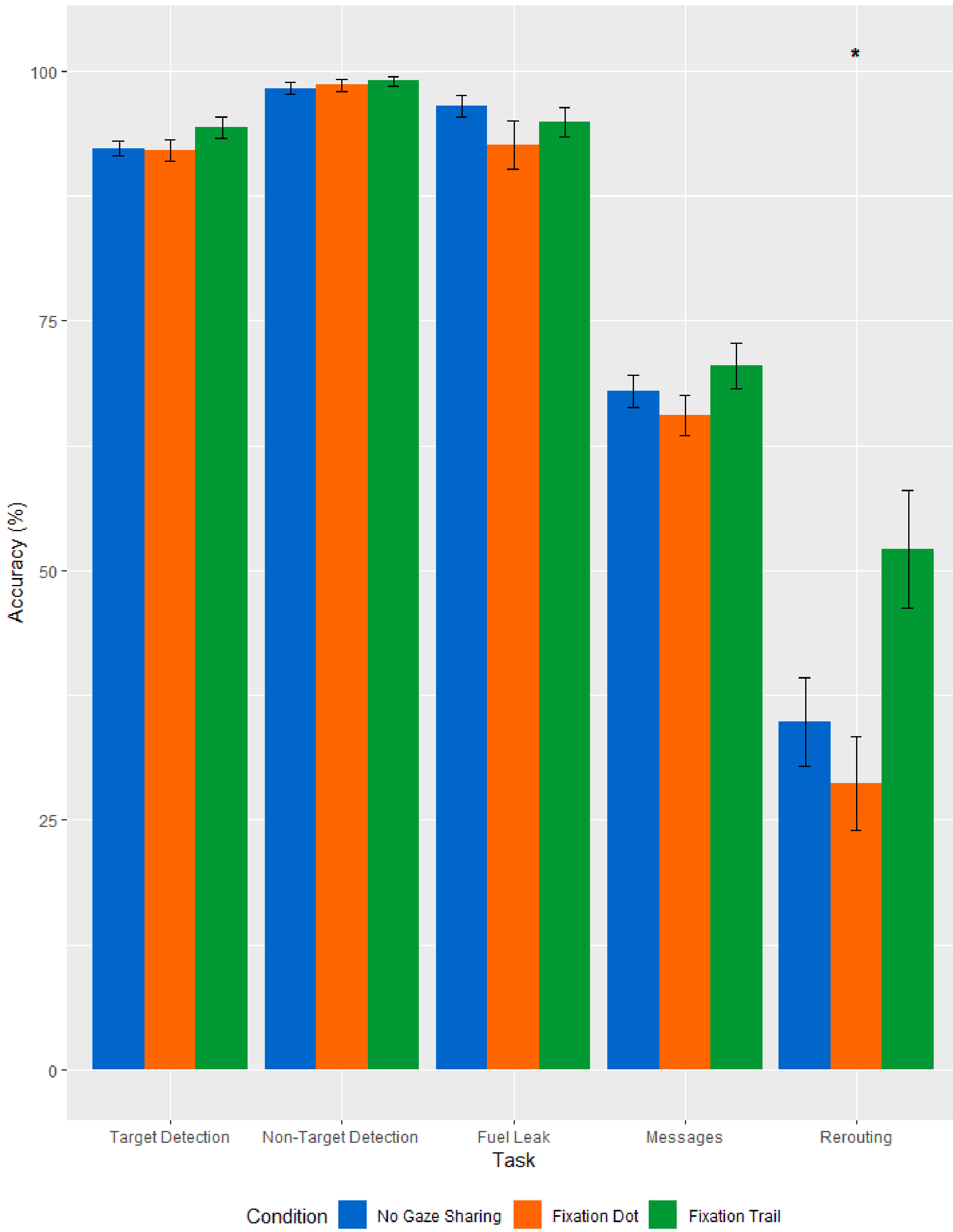


Figure 4.4: Accuracy (%) by task for each condition. An asterisk (*) indicates significant main effects in accuracy for a task.

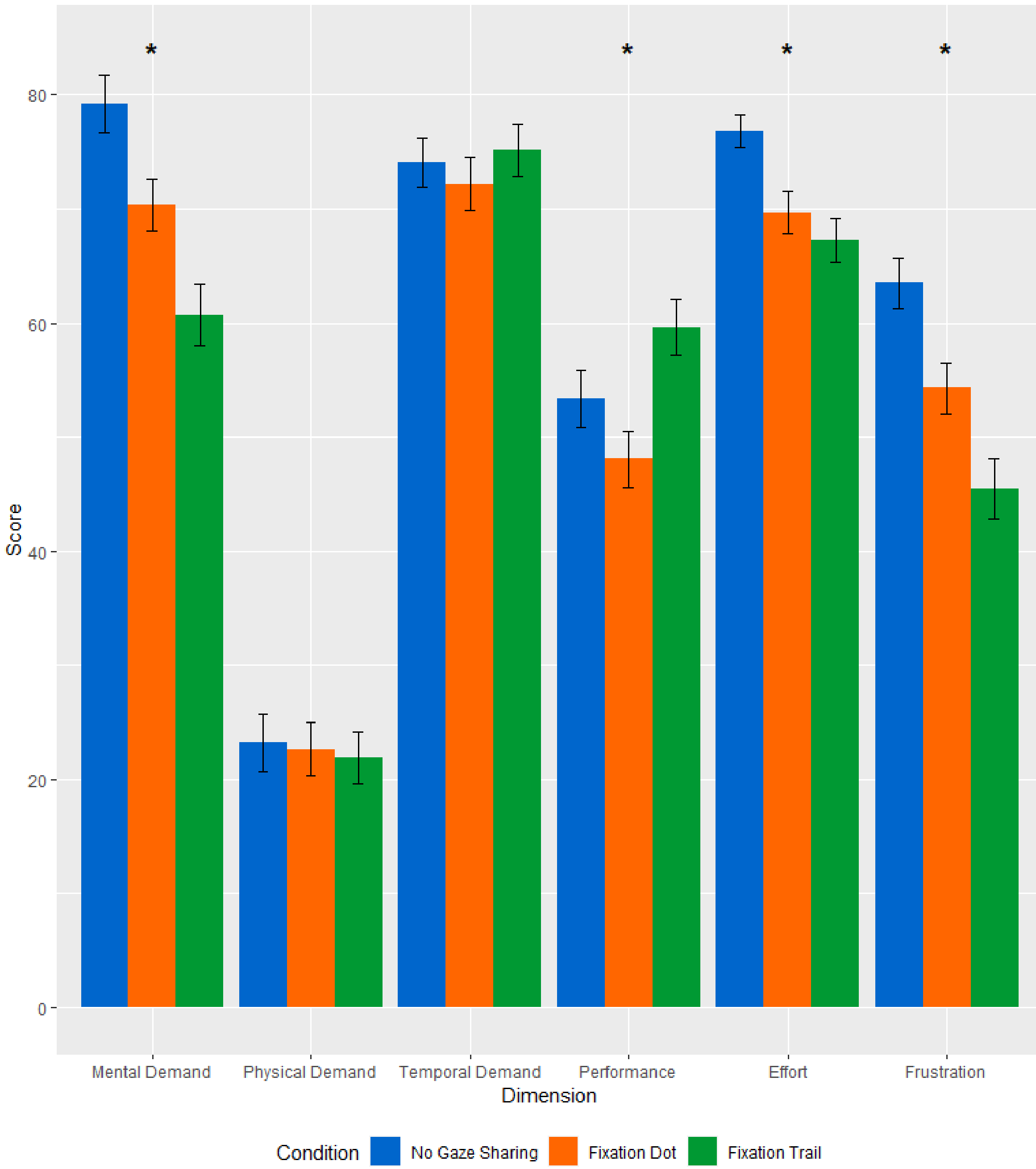


Figure 4.5: NASA-TLX scores for each dimension by condition. An asterisk (*) indicates significant main effects for a dimension.

4.4.3 Situation Awareness Measures

Team Situation Awareness

The first set of bars in Figure 4.6 shows the mean and standard error of the overall SAGAT-TSA scores across the 35 teams for each condition. Since the overall SAGAT score constitutes of the scores from the three levels, we ran an individual repeated measure ANOVA with a Greenhouse-Geisser correction to test for significant difference between the three conditions. The results showed that the overall SAGAT-TSA score ($F(1.59,53.91) = 80.72$, $p<.001$, partial $\eta^2 = 0.7$) was statistically significantly different between the three conditions. The use of fractional degrees of freedom is due to the Greenhouse-Geisser correction. This correction was necessary because the assumption of sphericity was violated, indicating that the variances of the differences between all possible pairs of conditions were not equal. By applying the Greenhouse-Geisser correction, we ensure the validity of the ANOVA results despite this violation.

A one-way repeated measures MANOVA was conducted to check for any statistical difference between the SAGAT scores across the three levels of the three conditions. A significant multivariate effect was observed for the gaze conditions, $F(6,132) = 20.94$, $p< .001$; Wilks' $\Lambda = 0.26$, partial $\eta^2 = 0.49$. Three follow-up repeated measures univariate ANOVAs with a Greenhouse-Geisser correction showed that SA level 1 (perception; $F(1.64,55.76) = 38.91$, $p<.001$, partial $\eta^2 = 0.53$), level 2 (comprehension, $F(1.85,62.83) = 53.87$, $p<.001$, partial $\eta^2 = 0.61$), and level 3 (projection, $F(1.64,55.86) = 24.24$, $p<.001$, partial $\eta^2 = 0.42$) are statistically significantly different between the three conditions, using a Bonferroni adjusted α level of 0.0167.

Post hoc tests revealed that all three conditions were significantly different from one another with the fixation trail having the highest level of overall TSA, followed by no gaze sharing, and fixation dot (all $p < .001$). For *Level 1*, post hoc tests showed that teams had significantly different levels of perception between all conditions with fixation trail having the highest level of perception, followed by no gaze sharing, and fixation dot (all $p < .0167$). For *Level 2* teams had significantly different levels of comprehension between all conditions with fixation trail having the highest level of perception, followed by no gaze sharing, and fixation dot (all $p < .001$). For *Level 3*, teams had significantly higher levels of projection using the fixation trail gaze compared to no gaze sharing and using the fixation dot (both $p < .001$), but there was not a difference between no gaze sharing and using the fixation dot ($p = .95$).

Shared Situation Awareness

The mean SAGAT-SSA score across participants in each condition was calculated and the results (mean and SD) are reported in Table 4.4. The fixation trail resulted in the highest SSA score, followed by no gaze sharing and the fixation dot.

A one-way repeated measures ANOVA with a Greenhouse-Geisser correction showed that the mean SAGAT-SSA scores were statistically significantly different between the three visualization techniques ($F(1.66, 55.27) = 7.78, p = .002$, partial $\eta^2 = 0.19$). Post hoc test using the Bonferroni adjustment showed that pairs exhibited a higher level of SSA using the fixation trail than the fixation dot ($p < .001$). There was no difference between no gaze sharing and both the fixation dot ($p = .36$) and the fixation trail ($p = .13$) techniques.

4.4.4 Eye Tracking Metrics

Figure 4.7 shows the mean and standard error of the six eye tracking metrics (number of fixations, fixation duration, number of saccades, saccade duration, amplitude, and velocity) across the 35 teams for each condition. Six repeated measures univariate ANOVAs with

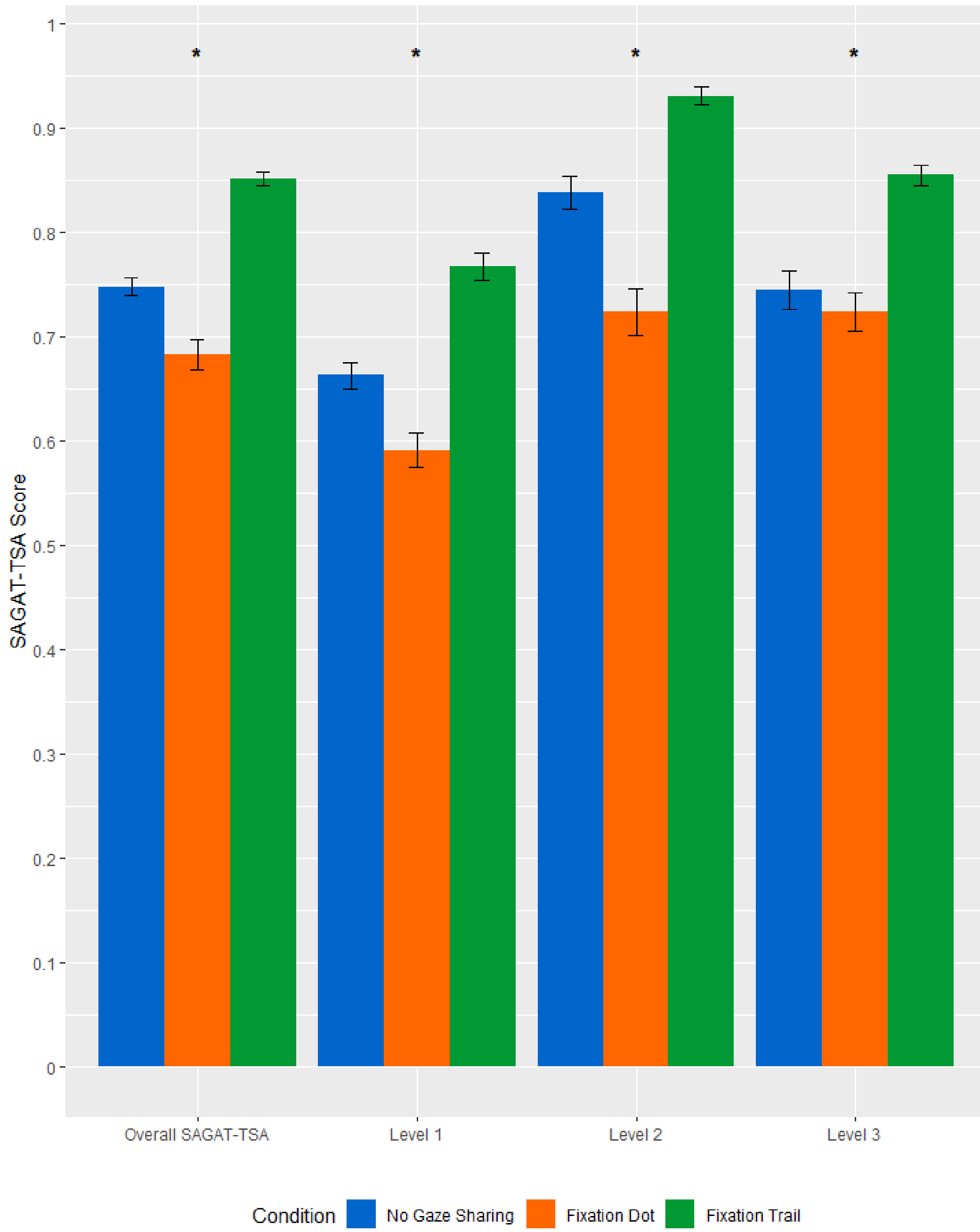


Figure 4.6: Overall SAGAT-TSA scores and by level for each conditions. An asterisk (*) indicates significant main effects in SA scores overall and by level.

a Greenhouse-Geisser correction revealed that the mean number of saccades ($F(1.65{,}56) = 17.086$, $p< .001$, partial $\eta^2 = 0.43$), saccade duration ($F(1.77{,}60.22) = 4.74$, $p = .012$, partial $\eta^2 = 0.22$), and saccade velocity ($F(1.36{,}46.35) = 1.25$, $p = .04$, partial $\eta^2 = 0.32$) differed statistically significantly between the three conditions.

Table 4.4: Mean SAGAT-SSA scores across participants using the three different visualization techniques (N=35). Standard deviations are included in parenthesis.

Visualization Technique	Both Teammates Correct	Both Teammates Incorrect in the Same Way	Total SAGAT-SSA Score
No Gaze Sharing (Control)	46.67% (12.94%)	9.52% (8.91%)	56.19% (15.28%)
Fixation Dot	45.14% (10.49%)	6.12% (7.82%)	51.24% (12.18%)
Fixation Trail	57.33% (14.66%)	8.44% (8.05%)	65.33% (14.25%)

Post hoc analysis with a Bonferroni adjustment showed significant differences between conditions. Participants exhibited the lowest number of saccades in the trail condition compared to the no gaze sharing condition ($p < .001$) and the dot condition ($p = .004$). Participants also exhibited a higher number of saccades when there was no gaze sharing compared to the dot condition ($p = .042$). Saccade durations were significantly longer in the dot condition compared to the trail condition ($p = .012$). Saccade velocities were significantly higher in the no gaze sharing condition compared to the trail condition ($p = .018$). These results suggest that the nature of the stimuli significantly influenced participants' eye movement patterns, with each condition eliciting distinct saccadic behaviors.

4.4.5 Qualitative Analysis

The qualitative analysis of the debriefing questionnaires revealed diverse opinions among the participants regarding the helpfulness of each visualization technique. A third of participants found the dot visualization technique to be helpful, citing its ability help identify areas to avoid allocating visual attention and preventing double-clicking with one participant stating, "Yes, it helped in the primary task to know where not to spend my time/eyes on." Moreover, 91% of the people who found the dot technique to be helpful expressed that the trail was

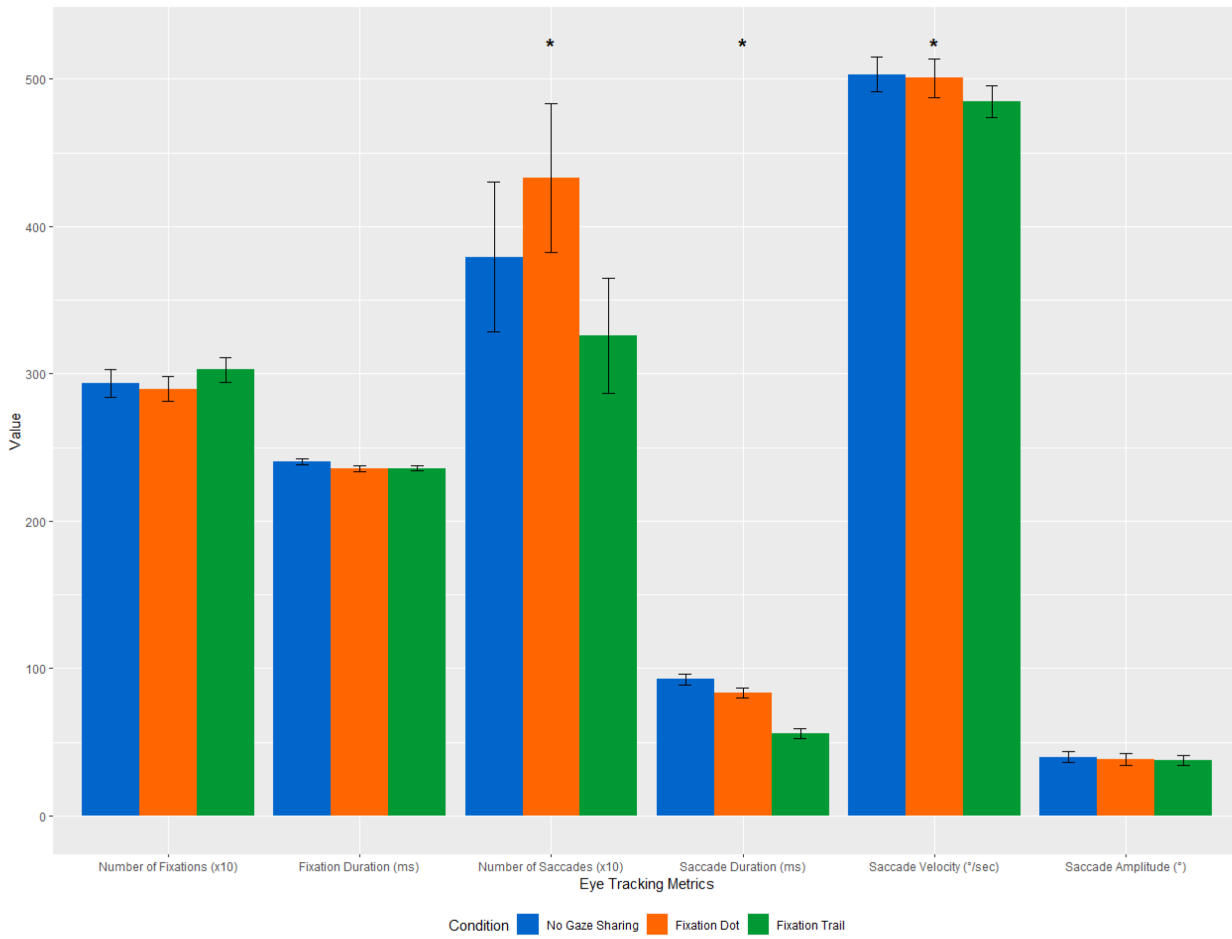


Figure 4.7: Eye tracking metrics for each condition. An asterisk (*) indicates significant main effects for a metric (ms = milliseconds, ° = degrees visual angle).

more helpful than the dot. One participant stated, "I found both techniques to be helpful. However, the trail is more helpful and gives more information than the dot". However, the remaining 66% of participants expressed dissatisfaction with the dot, with reasons ranging from its small size and shape to challenges in simultaneously focusing on the dot and the task with one participant mentioning, "It was a bit hard to keep track of my partner's gaze using the dot".

In contrast, the trail visualization technique received more favorable feedback from participants where 76% of participants found the trail visualization technique to be helpful. They appreciated its increased visibility and ease of tracking, allowing them to monitor their partner's gaze and focus on other areas of the task. One of many participants acknowledged its effectiveness, stating "The trail is more visible and easier to keep track of my partner's gaze and focus on the screen so that I can concentrate on other areas". However, 15% of participants expressed concerns about the trail being distracting and its potential to impede vision and the brightness of its color.

Table 4.5 provides a summary across all the analyses conducted—i.e., performance, workload, TSA (overall and by each level), and the eye tracking metrics—and the significant pairwise comparisons.

4.5 Discussion

We aimed to understand the effects of different gaze sharing visualization techniques on eye tracking metrics, workload, team situation awareness (TSA), and task performance of pairs working on a complex visual command-and-control task. This study directly compares two promising real-time gaze sharing techniques within a single study to provide a more holistic understanding of the potential benefits and limitations of each. Namely, knowing the scanning behaviors of teammates, guided by the fixation trail, yielded a more deliberate and efficient scanning strategy that lowered workload. Moreover, observing their partner's gaze location over time using the trail improved SSA and TSA at all levels. As a result, we surmise that the combination of lower workload and heightened SA improved overall performance when using the trail (Endsley & Kaber, 1999; Reinerman-Jones et al., 2019). Figure 4.8 summarizes the findings as to why the trail was successful and each part of the Figure will now be discussed.

Participants in the trail condition had knowledge of the teammate's position and scan behavior that allowed them to allocate their attention more efficiently, focusing on specific subtasks. This was supported by the qualitative analysis as participants reported that

Table 4.5: Summary of the post hoc test results for all dependent measures and a summary ranking of the conditions for each measure (NGS = No Gaze Sharing, Trail = Fixation Trail, Dot = Fixation Dot).

Dependent Measure		**Pairwise Comparisons**						**Rank order of each condition**
		No Gaze Sharing and Fixation Dot		**No Gaze Sharing and Fixation Trail**		**Fixation Dot and Fixation Trail**		
		p	**Significance**	***p***	**Significance**	***p***	**Significance**	
Performance	Score	<.001	*	<.001	*	<.001	*	Trail >NGS >Dot
	Reroute Accuracy	.96		.054		.002	*	Trail >NGS >Dot
NASA-TLX	Mental Demand	.004	*	<.001	*	.003	*	NGS >Dot >Trail
	Performance	.22		.093		<.001	*	Trail >NGS >Dot
	Effort	.002	*	<.001	*	.35		NGS >Dot >Trail
	Frustration	.002	*	<.001	*	.007	*	NGS >Dot >Trail
SAGAT	Overall TSA Score	<.001	*	<.001	*	<.001	*	Trail >NGS >Dot
	TSA Level 1	.006	*	<.001	*	<.001	*	Trail >NGS >Dot
	TSA Level 2	<.001	*	<.001	*	<.001	*	Trail >NGS >Dot
	TSA Level 3	.95		<.001	*	<.001	*	Trail >NGS >Dot
	SSA Score	.36		.13		<.001	*	Trail >NGS >Dot
Eye Tracking Metrics	Number of Saccades	.042	*	<.001	*	.001	*	Dot >NGS >Trail
	Saccade Duration	.84		.12		.012	*	NGS >Dot >Trail
	Saccade Velocity	.81		.018	*	.87		NGS >Dot >Trail

knowing their partner's gaze history with the trail allowed them to focus more on tasks/areas where their partner was not looking. This is also supported by the eye tracking results that

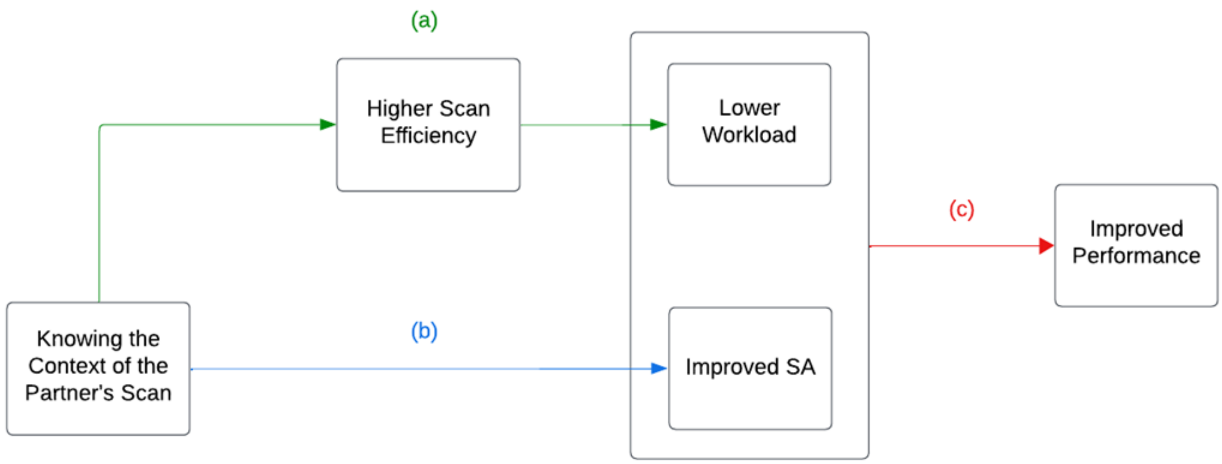


Figure 4.8: Depiction of how knowing the context of the partner's scan with the trail: (a) improved scan efficiency which lowered workload and (b) improved SA. Having both (a) and (b) resulted in (c) improved performance with the trail.

showed the trail had fewer saccades—indicating participants scanned a more focused portion of the entire display. Conversely, in the absence of gaze sharing participants engaged in broader and faster scanning across the entire display.

The findings here align with previous research indicating that shared visual information can lead to more efficient task execution (Vesper et al., 2016). Specifically, here, the increased scan efficiency lowered perceived workload (Figure 4.8a). However, research shows that reduced saccadic activity can be associated with higher mental workload (Di Stasi et al., 2012; Škvareková et al., 2020; Tsai et al., 2007), but this work centers around one individual and not teams. Here, the trail condition allowed participants to adopt the most efficient scan because they were aware of their partner's scan in real-time. Therefore, the reduced saccadic activity here can be attributed to a more strategic scanning technique rather than it being an indicator of increased workload. This was also supported by our NASA-TLX results as the fixation trail was perceived as less mentally demanding and had lower levels of frustration compared to the other conditions.

The debriefing questionnaire results also support these findings as 76% of participants found the trail visualization technique to be helpful. Participants noted that the fixation trail allowed them to get a better understanding of their partner's attentional focus and thus improved their own decision-making. The absence of gaze sharing also led to significantly higher levels of frustration compared to when there was a visualization present. This supports the work of Stein and Brennan (2004) showing that gaze sharing improves the team members' emotional well-being and satisfaction during task execution.

Knowing the context of their partner's gaze location over time also improved SA (Figure 4.8b). The fixation trail resulted in the highest TSA and SSA scores, followed by no gaze sharing, and the fixation dot with the lowest SA scores. This supports previous findings that showed the effectiveness of using a fixation trail (Akkil et al., 2016; J. Li et al., 2016; Newn et al., 2017). The higher SA scores with the fixation trail can be attributed to its ability to provide a temporal context of team members' gaze movements, allowing for a better understanding of their focus and intentions. This finding supports previous work that showed relying solely on a single point representation of the partner's gaze, as with the fixation dot, may limit the team's ability to comprehend critical aspects of the shared situation (Kumar et al., 2018).

Analyzing TSA levels at different levels of SA (i.e., perception, comprehension, and projection) provides valuable insights into the effects of gaze visualization techniques. Teams using the fixation trail exhibited the highest levels of perception. This may be because the fixation trail was more conspicuous than the dot which allowed participants to detect their partner's gaze movements more readily. The trail was thus the best in helping participants perceive critical information and identify relevant cues—i.e., factors pertaining to Level 1 SA (Endsley, 1995a).

The percentage of times both teammates were incorrect in the same way was lowest in the fixation dot condition, followed by the fixation trail, and the control. Lower percentages here are desirable as they indicate fewer shared misconceptions. Although these differences were not tested statically as the aim was to focus on total SAGAT-SSA scores, the fixation dot's lower percentage *might* imply it reduces shared errors more effectively compared to the control and fixation trail conditions. However, this technique did not lead to the highest total SAGAT-SSA score, suggesting that while it might reduce shared misconceptions, it may not sufficiently enhance overall SSA to the same extent as the fixation trail.

The fixation trail provides continuous visual cues about the trajectory of the participants' gaze, offering valuable spatial context within the shared task environment according to our qualitative analysis. This additional context provided by the fixation trail enhanced comprehension by offering insights into the participants' actions and cognitive processes within a collaborative setting. It enabled participants to perceive not only the current focus of their teammate's attention, but also the order in which they attended to various tasks. This history likely enabled participants to draw conclusions about their partners' strategy in completing the tasks and be able to adjust their own strategy accordingly.

Finally, teams using the fixation trail exhibited the highest levels of projection, suggesting that the fixation trail supports teams in accurately anticipating future states based on shared gaze information. Participants also mentioned in the debriefing questionnaire that the fixation trail allowed them to anticipate and project where their partner's attention will be directed. This cognitive projection aligns with the concept of shared mental models, where participants develop a collective understanding of the task through their teammate's gaze movements (Cannon-Bowers et al., 1993). The ability to project and plan based on teammate gaze patterns offers valuable insights to how gaze sharing can support collaboration.

For this study, the combination of lower workload and improved SA with the trail resulted in improved performance (Figure 4.8c). This is consistent with the literature, which suggests that these two outcomes improve performance (Endsley & Kaber, 1999; Reinerman-Jones et al., 2019). Even though the fixation dot resulted in lower workload than having no gaze sharing, it likely lacked the necessary level of detail to support team SA. This was also supported by the debriefing questionnaire findings where 17 participants found the dot technique to be helpful, but of these participants, 91% said that the trail was more helpful than the dot. Overall, what made the fixation trail successful here compared to the other conditions was that it allowed participants to continuously be aware of their partner's scan—i.e., where they had been looking, where they were currently looking, and where they could be looking next.

Although the fixation dot was not effective here, other work with simpler tasks has shown otherwise (Laskowitz et al., 2022; Lee et al., 2017). It is worth noting that this study is among the few studies that use gaze sharing in a complex environment that requires participants to make quick decisions and adapt in real-time. The findings here show that the fixation trail is more suitable under these circumstances. These findings underscore the fact that including gaze visualizations is not enough and that it is important to carefully design and implement gaze visualization techniques that consider context as well. However, the lack of statistical significance in other pairwise comparisons prompts further examination.

Overall, these findings provide valuable insights that can inform the design of technology and displays to support collaborative tasks (Atweh et al., 2023). Given the use of a fixation trail gaze visualization technique resulted in the best performance, TSA, and reduced workload, incorporating a continuous representation of a partner's eye movements can be beneficial for facilitating effective collaboration (Cheng et al., 2022; Y. Li et al., 2019). Designers should consider implementing visualizations that provide real-time and detailed information about the partner's gaze behavior, allowing for a more accurate understanding of their intentions and attentional focus (Fasold et al., 2021). On the other hand, relying solely on a simplistic indication of the general area of the partner's gaze may not be sufficient for supporting collaboration. The limitations identified with the dot visualization technique from the debriefing questionnaire highlight the importance of designing gaze sharing aids that are visually salient, easily interpretable, and are not distracting. Enhancing the visual salience of gaze sharing cues, while ensuring they do not interfere with the primary task, can improve the overall effectiveness and acceptance of such techniques (Newn et al., 2017).

4.6 Limitations and Future Work

A limitation of this study is the applicability of the findings in other domains. The task was designed around a set of operational scenarios and demands specific to UAV command in a controlled laboratory setting. Although there are parallels to other complex domains, some of the findings may not be applicable to all tasks. It is important for future research to build on this work in other contexts to determine the effectiveness of gaze visualization techniques in different tasks and settings. By examining a wider range of tasks, we can obtain a more comprehensive understanding of the benefits and limitations of various gaze visualization techniques across different real-world, operational environments. In addition, future work should explore additional factors that may influence the effectiveness of gaze sharing, such as team size (i.e., teams of more than two; Atweh et al., 2022), the complexity of the task, individual differences in cognitive processes, and the role of trust and communication within the team (Khan et al., 2012). This knowledge can inform the development of more robust and adaptive gaze sharing systems.

It is also important to recognize that no theoretical framework or measurement method is without limitations. Here we adopted SAGAT, but a number of researchers have raised concerns which underscore the need for a nuanced interpretation of SAGAT results (de Winter et al., 2019; Flach, 1995). However, it is crucial to recognize that Endsley's extensive body of work on SA includes the theoretical basis for using SAGAT that has been supported by other researchers (Gardner et al., 2017; Hogan et al., 2006; Hogg et al., 1993; Matthews & Beal, 2002). Thus, the empirical support for the use of SAGAT in assessing and quantifying team SA outweighs its costs for the purposes of the work presented here.

While this study used the fixation dot and fixation trail visualization techniques, it is important to note these are only two visualizations that exist. Other visualizations include cursors, heatmaps (Špakov & Miniotas, 2007), shared battle graphics (Hiniker & Entin, 1992), and network-centric approaches (Entin et al., 2006). For instance, heatmaps provide a visual representation of areas with high gaze concentration, which may offer insights into shared attention and decision-making processes, but users may find it provides too much information (Špakov & Miniotas, 2007). Comparing and evaluating these alternative techniques alongside the fixation dot and fixation trail could shed light on their potential advantages and disadvantages across visualizations. This exploration can inform the design and implementation of future technologies and displays, ensuring that they are tailored to specific operational requirements and can effectively support collaborative decision-making and attention management in complex tasks (Atweh, El Iskandarani, & Riggs, 2024).

The findings highlight the importance of considering the context and complexity of the collaborative task when designing and employing gaze visualization techniques. Different tasks may require different levels of gaze information to support effective collaboration. Furthermore, future studies should explore the interplay between gaze visualization techniques and other factors that influence SA, such as team dynamics, workload, and individual differences (Atweh et al., 2022). Understanding how these factors interact with gaze visualizations can provide valuable insights into designing technology and displays that account for the complex nature of collaborative work. For instance, examining the impact of workload on the effectiveness of gaze visualization techniques can help identify strategies to optimize the presentation and timing of gaze information to mitigate cognitive overload.

Overall, this study is among the few of its kind that compared multiple gaze visualizations techniques and integrated real-time gaze sharing technology within a simulated complex, dynamic environment. This study contributes to the knowledgebase in real-time visualization design and implementation to support team collaboration. By integrating these insights, we can potentially achieve more effective collaboration within teams, improved performance, and enhanced team SA in data-rich environments.

4.7 Conclusion

4.7.1 Gaze Sharing is Effective When Implemented Correctly

One major finding of our research emphasizes the pivotal role of visualizations as a tool to facilitate effective gaze sharing in collaborative UAV command and control tasks. Our study revealed that when done right, visualizations like the trail technique can significantly enhance team performance and SA. However, it is equally crucial to recognize that when done incorrectly (i.e., fixation dot), it can negatively affect performance and SA.

The trail visualization technique demonstrated Its effectiveness in guiding participants' gaze allocation, promoting better coordination, and improving overall task performance. By providing a clear and informative trail of their partner's gaze, participants were able to monitor and anticipate their partner's actions, leading to more effective decision-making and task execution. On the other hand, the dot visualization technique highlighted the need to consider the users' needs. Participants' dissatisfaction with the dot stemmed from its limited noticeability, potential cognitive burden, and lack of visual salience. These shortcomings negatively impacted the effectiveness of gaze sharing from a performance and SA standpoint.

The key takeaway here is that gaze sharing is indeed a valuable mechanism for collaborative tasks; however, its success hinges on the careful and thoughtful design of the visualization. To fully unlock the potential of gaze sharing, designers must ensure that the visualizations are unobtrusive, salient, and informative. Designers should prioritize the creation of visually salient and intuitive visualizations that seamlessly integrate with the task at hand, fostering a more natural and efficient collaborative environment. By providing gaze sharing cues that are easily salient and informative, team members can better perceive and comprehend their partner's intentions and actions, leading to enhanced coordination and task outcomes.

4.7.2 Multiple Analyses are Needed to Fully Understand Team Collaboration

The comprehensive evaluation approach allowed us to gain a multifaceted perspective on the effectiveness of gaze sharing visualization techniques in supporting team collaboration during the UAV task. By concurrently considering eye tracking analysis, performance metrics, NASA-TLX workload assessment, team SA assessments, and the qualitative data from the debriefing questionnaire, we were able to draw deeper insights about the strengths and limitations of each visualization technique from various vantage points.

The eye tracking analysis revealed distinct patterns of gaze behavior among participants using different visualization methods, shedding light on the underlying cognitive processes and workload distribution within teams. Performance metrics showed that the trail visualization technique outperformed other visualization techniques. However, we could not assess the differences between the different visualization techniques using just performance metrics as the NASA-TLX data was not always aligned. For example, although the dot technique yielded lower team performance as compared to no gaze sharing, it was perceived to be less mentally demanding than the absence of gaze sharing. Therefore, it was critical to collect SA data and participants' feedback from the debriefing questionnaire which allowed us to gain insights into the usability aspects, perceived benefits, and challenges associated with each visualization technique. Furthermore, collecting and analyzing SA assessments provided valuable information about team SA overall and by level and highlighted the significance of supporting all levels of SA for teams. This analysis allowed us to discover that the fixation trail was effective in fostering perception, comprehension, and projection processes among team members. Understanding the participants' SA is critical in collaborative tasks, as it directly influences their decision-making and coordination with team members.

Overall, the combination of qualitative and subjective assessments further enriched our understanding of the participants' interactions with the gaze sharing techniques. The combination of these analyses allowed us to make informed recommendations with regards to the design and implementation of gaze sharing technologies to support team collaboration.

4.8 References

Akkil, D., James, J. M., Isokoski, P., & Kangas, J. (2016). GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. *Conference on Human Factors in Computing Systems - Proceedings*, 1151–1158. https://doi.org/10.1145/2851581.2892459

Atkinson, R. B., Sidey-Gibbons, C., Smink, D. S., Askari, R., Pusic, A. L., Cho, N. L., Robertson, J. M., & Rangel, E. L. (2023). Real-Time Student Feedback on the Surgical Learning Environment: Use of a Mobile Application. *Journal of Surgical Education*. https://doi.org/10.1016/j.jsurg.2023.02.017

Atweh, J. A., El Iskandarani, M., & Riggs, S. L. (2024). Real-Time Gaze Sharing Techniques and Their Influence on Performance and Shared Situational Awareness of Teammates in UAV C2 Operations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *68*(1), 146–154. https://doi.org/10.1177/10711813241260316

Atweh, J. A., Hazimeh, R., & Riggs, S. L. (2023). Can Real-Time Gaze Sharing Help Team Collaboration? A Preliminary Examination of its Effectiveness with Pairs. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *67*(1), 716–721. https://doi.org/10.1177/21695067231193659

Atweh, J. A., Moacdieh, N. M., & Riggs, S. L. (2022). Identifying Individual-, Team-, and Organizational-Level Factors that Affect Team Performance in Complex Domains Based on Recent Literature. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1795–1799. https://doi.org/10.1177/1071181322661213

Atweh, J. A., & Riggs, S. L. (2024). ScanMatch Versus MultiMatch: A Comparison of the Sensitivity of Scanpath Similarity Metrics to Changes in Workload. *2024 IEEE Systems and Information Engineering Design Symposium (SIEDS)*, 57–62. https://doi.org/10.1109/SIEDS61124.2024.10534720

Atweh, J. A., Tabbara, I., Nasrallah, G., & Riggs, S. L. (2024). Reducing Barriers in Analyzing Eye Tracking Data: The Development of a GUI to Preprocess Eye Tracking Data. *2024 IEEE Systems and Information Engineering Design Symposium (SIEDS)*, 286–291. https://doi.org/10.1109/SIEDS61124.2024.10534671

Bai, H., Sasikumar, P., Yang, J., & Billinghurst, M. (2020). A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376550

Berger, M., Dey, D., Dandekar, A., Barati, B., Pfleging, B., & Bernhaupt, R. (2023). Empowering Driver-Passenger Collaboration: Designing In-Car Systems with a Focus on Social Connectedness, Fairness, and Team Performance. *International Journal of Human–Computer Interaction*, *39*, 3180–3202. https://doi.org/10.1080/10447318.2023.2205769

Bolton, M. L., Biltekoff, E., & Humphrey, L. (2023). The Mathematical Meaninglessness of the NASA Task Load Index: A Level of Measurement Analysis. *IEEE Transactions on Human-Machine Systems*, *53*(3), 590–599. https://doi.org/10.1109/THMS.2023.3263482

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Co-ordinating Cognition: The Costs and Benefits of Shared Gaze During Collaborative Search. *Cognition*, *106*, 1465–1477. https://doi.org/10.1016/j.cognition.2007.05.012

Bulling, A., & Gellersen, H. (2010). Toward Mobile Eye-Based Human-Computer Interaction. *IEEE Pervasive Computing*, *9*(4), 8–12. https://doi.org/10.1109/MPRV.2010.86

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared Mental Models in Expert Team Decision Making. In J. Castellan (Ed.). Lawrence Erlbaum Associates, Inc.

Cheng, S., Wang, J., Shen, X., Chen, Y., & Dey, A. (2022). Collaborative eye tracking based code review through real-time shared gaze visualization. *Frontiers of Computer Science*, *16*, 163704. https://doi.org/10.1007/s11704-020-0422-1

de Winter, J. C., Eisma, Y. B., Cabrall, C. D. D., Hancock, P. A., & Stanton, N. A. (2019). Situation Awareness Based on Eye Movements in Relation to the Task Environment. *Cognition, Technology & Work*, *21*(1), 99–111. https://doi.org/10.1007/s10111-018-0527-6

Di Gregorio, M., Romano, M., Sebillo, M., Vitiello, G., & Vozella, A. (2021). Improving Human Ground Control Performance in Unmanned Aerial Systems. *Future Internet*, *13*(8), 188. https://doi.org/10.3390/fi13080188

Di Stasi, L. L., Renner, R., Catena, A., Canas, J. J., Velichkovsky, B. M., & Pannasch, S. (2012). Towards a Driver Fatigue Test Based on the Saccadic Main Sequence: A Partial Validation by Subjective Report Data. *Transportation Research Part C: Emerging Technologies*, *21*(1), 122–133. https://doi.org/10.1016/j.trc.2011.07.002

El Iskandarani, M., Atweh, J. A., McGarry, S. P. D., Riggs, S. L., & Moacdieh, N. M. (2023). Does It MultiMatch? What Scanpath Comparison Tells us About Task Performance in Teams. *Journal of Cognitive Engineering and Decision Making*, 155534342311714. https://doi.org/10.1177/15553434231171484

Endsley, M. R. (2021). *Situation Awareness Measurement: How to Measure Situation Awareness in Individuals and Teams*. Human Factors & Ergonomics Society. https://books.google.com/books?id=jgptzgEACAAJ

Endsley, M. R., & Kaber, D. B. (1999). Level of Automation Effects on Performance, Situation Awareness and Workload in a Dynamic Control Task. *Ergonomics*, *42*(3), 462–492. https://doi.org/10.1080/001401399185595

Endsley, M. R. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting*, *32*, 97–101. https://doi.org/10.1177/154193128803200221

Endsley, M. R. (1995a). Measurement of Situation Awareness in Dynamic-Systems. *Human Factors*, *37*, 65–84. https://doi.org/10.1518/001872095779049499

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems: Situation awareness. *Human factors*, *37*, 32–64.

Entin, E. E., Hiniker, P., Grier, R., Jefferson, T., Vecchio, G., & Harkins, J. (2006). Enhancing Situational Awareness and Team Performance Using Network-Centric Technologies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(3), 482–486. https://doi.org/10.1177/154193120605000357

Fasold, F., Nicklas, A., Seifriz, F., Schul, K., Noël, B., Aschendorf, P., & Klatt, S. (2021). Gaze Coordination of Groups in Dynamic Events – A Tool to Facilitate Analyses of Simultaneous Gazes Within a Team. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.656388

Fiore, S. M., & Salas, E. (2004). Why We Need Team Cognition. American Psychological Association. https://doi.org/10.1037/10690-011

Flach, J. M. (1995). Situation Awareness: Proceed with Caution. *Human Factors*, *37*(1), 149–157. https://doi.org/10.1518/001872095779049480

Gardner, A. K., Kosemund, M., & Martinez, J. (2017). Examining the Feasibility and Predictive Validity of the SAGAT Tool to Assess Situation Awareness Among Medical

Trainees. *Simulation in Healthcare*, *12*(1), 17–21. https://doi.org/10.1097/SIH.0000000000000181

Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2020). Measuring Real-Time Team Cognition During Team Training. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *62*, 825–860. https://doi.org/10.1177/0018720819852791

Gupta, K., Lee, G. A., & Billinghurst, M. (2016). Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics*, *22*, 2413–2422. https://doi.org/10.1109/TVCG.2016.2593778

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. https://doi.org/10.1016/S0166-4115(08)62386-9

He, S.-j., Zhi, J.-y., Du, Y., & Feng, S. (2023). Effects of Design Features of Train Electronic Guidance Interface on Visual Search, Behavior, and Usability. *International Journal of Industrial Ergonomics*, *93*, 103387. https://doi.org/10.1016/j.ergon.2022.103387

Hiniker, P. J., & Entin, E. E. (1992). Examining Cognitive Processes in Command Crisis: New HEAT Experiments on Shared Battle Graphics and Time Tagging. *Proceedings of the 1992 Symposium on Command and Control Research.*

Hogan, M. P., Pace, D. E., Hapgood, J., & Boone, D. C. (2006). Use of Human Patient Simulation and the Situation Awareness Global Assessment Technique in Practical Trauma Skills Assessment. *Journal of Trauma and Acute Care Surgery*, *61*(5), 1047–1052. https://doi.org/10.1097/01.ta.0000238687.23622.89

Hogg, D. N., Torralba, B., & Volden, F. S. (1993). A Situation Awareness Methodology for the Evaluation of Process Control Systems: Studies of Feasibility and the Implication of Use. *OECD Halden Reactor Project.*

Huang, Y., Cai, M., Li, Z., Lu, F., & Sato, Y. (2020). Mutual Context Network for Jointly Estimating Egocentric Gaze and Action. *IEEE Transactions on Image Processing*, *29*, 7795–7806. https://doi.org/10.1109/TIP.2020.3007841

Jing, A., May, K., Matthews, B., Lee, G., & Billinghurst, M. (2022). The Impact of Sharing Gaze Behaviours in Collaborative Mixed Reality. *Proceedings of the ACM on Human-Computer Interaction*, *6*, 1–27. https://doi.org/10.1145/3555564

Khan, R. S. A., Tien, G., Atkins, M. S., Zheng, B., Panton, O. N. M., & Meneghetti, A. T. (2012). Analysis of Eye Gaze: Do Novice Surgeons Look at the Same Location as Expert Surgeons During a Laparoscopic Operation? *Surgical Endoscopy*, *26*, 3536–3540. https://doi.org/10.1007/s00464-012-2400-7

Kumar, A., Burch, M., & Mueller, K. (2018). Visual Analysis of Eye Gazes to Assist Strategic Planning in Computer Games. *Proceedings of the 3rd Workshop on Eye Tracking and Visualization*, 1–5. https://doi.org/10.1145/3205929.3205935

Laskowitz, S., Griffin, J. W., Geier, C. F., & Scherf, K. S. (2022, February). Cracking the Code of Live Human Social Interactions in Autism: A Review of the Eye-Tracking Literature. In C. Palmero, J. C. S. J. Junior, A. Clapés, I. Guyon, W.-W. Tu, T. B. Moeslund, & S. Escalera (Eds.). PMLR. https://proceedings.mlr.press/v173/laskowitz22a.html

Lee, G., Kim, S., Lee, Y., Dey, A., Piumsomboon, T., Norman, M., & Billinghurst, M. (2017). Improving Collaboration in Augmented Video Conference Using Mutually Shared Gaze. *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, ICAT-EGVE 2017*, 197–204. https://doi.org/10.2312/egve.20171359

Li, J., Manavalan, M., D'Angelo, S., & Gergle, D. (2016). Designing Shared Gaze Awareness for Remote Collaboration. *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, 325–328. https://doi.org/10.1145/2818052.2869097

Li, Y., Lu, F., Lages, W. S., & Bowman, D. (2019). Gaze Direction Visualization Techniques for Collaborative Wide-Area Model-Free Augmented Reality. *Symposium on Spatial User Interaction*, 1–11. https://doi.org/10.1145/3357251.3357583

Matthews, M. D., & Beal, S. A. (2002). Assessing Situation Awareness in Field Training Exercises. *Military Academy West Point NY Office of Military Psychology and Leadership*, 31.

Morgan, C., Chiappe, D., Kraut, J., Strybel, T., & Vu, K. (2012, January). An Investigation into the Potential Sources of Interference in SA Probe Techniques.

Newn, J., Velloso, E., Allison, F., Abdelrahman, Y., & Vetere, F. (2017). Evaluating Real-Time Gaze Representations to Infer Intentions in Competitive Turn-Based Strategy Games. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 541–552. https://doi.org/10.1145/3116595.3116624

Nyström, M., & Holmqvist, K. (2010). An Adaptive Algorithm for Fixation, Saccade, and Glissade Detection in Eyetracking Data. *Behavior Research Methods*, *42*(1), 188–204. https://doi.org/10.3758/BRM.42.1.188

Reinerman-Jones, L. E., Hughes, N., D'Agostino, A., & Matthews, G. (2019). Human Performance Metrics for the Nuclear Domain: A Tool for Evaluating Measures of Workload, Situation Awareness and Teamwork. *International Journal of Industrial Ergonomics*, *69*, 217–227. https://doi.org/10.1016/j.ergon.2018.12.001

Schneider, B., & Pea, R. (2017, December). Real-Time Mutual Gaze Perception Enhances Collaborative Learning and Collaboration Quality. https://doi.org/10.1007/978-3-319-45001-8_7

Škvareková, I., Pecho, P., Ažaltovič, V., & Kandera, B. (2020). Number of Saccades and Fixation Duration as Indicators of Pilot Workload. *Transportation Research Procedia*, *51*, 67–74. https://doi.org/10.1016/j.trpro.2020.11.009

Špakov, O., & Miniotas, D. (2007). Visualization of Eye Gaze Data using Heat Maps. *Elektronika Ir Elektrotechnika Medicine Technology*, *115*. https://etalpykla.vilniustech.lt/handle/123456789/149656

Stein, R., & Brennan, S. E. (2004). Another Person's Eye Gaze as a Cue in Solving Programming Problems. *Proceedings of the 6th international conference on Multimodal interfaces - ICMI '04*, 9. https://doi.org/10.1145/1027933.1027936

Sung, G., Feng, T., & Schneider, B. (2021). Learners Learn More and Instructors Track Better with Real-time Gaze Sharing. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), Article 134, 23 pages. https://doi.org/10.1145/3449208

Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task Performance and Eye Activity: Predicting Behavior Relating to Cognitive Workload. *Aviation, Space, and Environmental Medicine*, *78*(5), B176–B185.

Vesper, C., Schmitz, L., Safra, L., Sebanz, N., & Knoblich, G. (2016). The Role of Shared Visual Information for Joint Action Coordination. *Cognition*, *153*, 118–123. https://doi.org/10.1016/j.cognition.2016.05.002

Walton, J., & Gilbert, S. B. (2022). Evaluating the Effect of Displaying Team vs. Individual Metrics on Team Performance. *International Journal of Human-Computer Studies*, *160*, 102759. https://doi.org/10.1016/j.ijhcs.2021.102759

Zhang, Y., Pfeuffer, K., Chong, M. K., Alexander, J., Bulling, A., & Gellersen, H. (2017). Look Together: Using Gaze for Assisting Co-Located Collaborative Search. *Personal and Ubiquitous Computing*, *21*, 173–186. https://doi.org/10.1007/s00779-016-0969-x

Chapter 5

Real-Time Fixation Trail Gaze Sharing Visualization and Communication Dynamics in UAV Operations

5.1 Introduction

Teamwork in Unmanned Aerial Vehicle (UAV) command-and-control (C2) operations is particularly complex, demanding the integration of diverse skills and knowledge bases. Shared mental models, which refer to the common understanding of tasks, equipment, and team roles, play a pivotal role in ensuring effective coordination and decision-making (Cannon-Bowers et al., 1993). These mental models in UAV C2 teams are built and maintained through continuous communication, which is predominantly verbal, relying on spoken language to convey critical information and instructions (Nawaz et al., 2021). The dynamic and often high-stakes nature of UAV missions necessitates robust communication and a shared understanding among team members, which are critical for maintaining situation awareness (SA) and optimizing performance. While verbal communication is undoubtedly essential, it has its limitations, especially in high-stress and rapidly changing environments where quick and accurate information exchange is crucial. These limitations include the potential for miscommunication, particularly when teammates struggle to articulate precise instructions or descriptions under pressure (John et al., 2013). Delayed communication can further hinder collaboration in remote or distributed UAV centers, where time lags or bandwidth issues disrupt verbal exchanges (Cardosi & Lennertz, 2017; Mirzaei et al., 2010). Additionally, the noisy environments typical of UAV operation centers can make it difficult to hear or understand spoken instructions, further complicating coordination (Baker et al., 2021; Cummings et al., 2008).

Knowing where a team member is looking can potentially help teams to understand each other's focus of attention and improve coordination. To date, there has been limited work on gaze sharing—i.e., visualizing in real-time, where team members are looking (their eye movements) on a shared display. Despite its potential, gaze sharing in the context of UAV C2 operations remains underexplored (Kütt et al., 2019; Sung et al., 2021).

In Chapter 4, we evaluated various gaze sharing techniques (e.g., a dot versus a trail of fixations) against key performance metrics such as workload, team SA, and overall team performance (Atweh & Riggs, 2024; Atweh et al., 2023). Our findings indicated that a trailing gaze sharing technique, where team members can see the gaze path of their peers over time, was superior in enhancing team performance and SA. However, several critical questions remain unanswered: How does gaze sharing interact with traditional verbal communication methods? When is gaze sharing most effective, and does it complement or replace other forms of communication?

Chapter 5 aims to delve deeper into these questions, hypothesizing that the effectiveness of gaze sharing in UAV C2 operations may depend on the specific strategies employed by the team. We posit that the integration of both gaze sharing and verbal communication will yield superior outcomes compared to relying on verbal communication alone. By systematically exploring the dynamics of gaze sharing and team communication, we seek to uncover the conditions under which gaze sharing enhances SA and performance, thereby contributing to the development of more effective communication protocols in UAV C2 operations.

This study holds significant implications for the design of future UAV systems and the training of UAV operators, emphasizing the need to (1) bridge the gap between eye tracking research and practice and (2) have a holistic approach to communication that leverages both verbal and non-verbal cues. By advancing our understanding of gaze sharing in complex team environments, we aim to pave the way for more resilient and adaptive UAV C2 operations, ultimately enhancing mission success and operational efficiency.

5.2 Background

Effective communication is vital in UAV C2 operations, whether team members are co-located or working remotely. In these environments, real-time verbal communication—typically facilitated through intercoms, radios, or telephony—serves as the primary means of information exchange. However, the high-stakes nature of UAV missions often places immense pressure on verbal exchanges to be concise, accurate, and timely, especially when teams face rapidly changing operational demands (McDermott et al., 2005).

To mitigate misunderstandings and enhance clarity, UAV teams rely on standardized phraseology and structured protocols. Standard commands like “Affirmative,” “Negative,” or “Hold Position” are employed to ensure precision, while read-back techniques, in which the receiver repeats the instruction, confirm the accurate transmission of critical information

(Arbuckle et al., 2010; Cummings & Guerlain, 2007). These practices help maintain SA and streamline coordination in environments where information overload is a significant risk (Cheng et al., 2019).

For larger UAV teams, which may include roles like sensor operators, mission analysts, and pilots, managing communication becomes even more challenging. The need for clear role delineation and structured communication channels is amplified, as these larger teams require efficient information flow across diverse skill sets (Atweh et al., 2022). Yet, the hierarchical structures and formalized protocols designed to enhance clarity can sometimes slow response times and create communication bottlenecks, particularly in high-stress situations (Charapko et al., 2021; Giachetti et al., 2013).

In remote UAV operations, where non-verbal cues are absent, technology provides alternative methods to convey critical information. Visual indicators, such as flashing alerts or colored displays, and haptic feedback devices offer immediate, non-verbal feedback on system statuses or urgent events, allowing operators to respond swiftly without relying solely on verbal channels (Duan et al., 2019; Rizk et al., 2019; Stegagno et al., 2014).

While these communication strategies are essential, they are not without limitations. Verbal exchanges in UAV C2 operations can become strained under high workload conditions or delayed in remote settings, where situational ambiguity or lack of shared visual information can hinder effective team coordination (Tang et al., 2023). This gap suggests a need for exploring complementary communication methods, such as gaze sharing, to enhance real-time coordination in UAV teams.

As we saw in chapter 4, despite its potential benefits, the effectiveness and role of gaze sharing in UAV C2 operations remain underexplored. It is unclear whether gaze sharing can truly complement or replace traditional communication methods and how its utility might vary depending on different communication and task strategies. For instance, in some scenarios, gaze sharing might significantly reduce the need for verbal exchanges, while in others, it may serve only as a supplementary tool that enhances verbal communication.

This study aims to investigate the dynamics of gaze sharing and team communication in UAV C2 operations. We seek to understand whether gaze sharing serves as a complementary or substitutive form of communication and how its effectiveness is influenced by various communication and task strategies. By examining these factors, we hope to elucidate the conditions under which gaze sharing can most effectively enhance team performance and operational efficiency. This research will provide valuable insights into the integration of gaze sharing technologies in UAV operations and contribute to the broader understanding of team communication dynamics in high-stakes environments.

Building on this potential, this study aims to explore gaze sharing in more realistic environments where teams can communicate verbally, examining how its effects differ from conditions where verbal communication is restricted. By examining the interaction between gaze sharing and verbal communication, we hypothesize that integrating gaze sharing will complement traditional verbal communication and thus enhance team performance in UAV C2 operations by enabling smoother, more intuitive coordination. This research not only advances our understanding of non-verbal communication's role in complex team environments but also provides insight into which tasks benefit most from this technology. Additionally, it lays the groundwork for integrating eye tracking-based tools to support human performance in UAV systems.

5.3 Methods

5.3.1 Participants

This study was approved by the University of Virginia's Institutional Review Board (protocol number #3480). Twenty-four teams (48 participants) from the University of Virginia were recruited for the study (M = 23.1 years, SD = 3.49 years). Each team included one male and one female who were not acquainted with one another.

5.3.2 Experimental Setup

The setup was the same setup in Chapter 3. The only difference is the presence of an external microphone between participants to record their communications when communication was permitted (Figure 5.1) .

5.3.3 UAV Tasks

Participants were responsible for the same set of tasks (one primary and three secondary) as Chapter 3.

5.3.4 Experimental Design

A within-subjects study was conducted, where all teams completed three 10-min scenarios with one of the following three conditions: (a) no gaze sharing and communication, (b) gaze sharing with a real-time fixation trail and no communication, and (c) gaze sharing using a real-time fixation trail and communication. The fixation trail was the same one used in the previous Chapter 4 (Figure 4.2b). The order of conditions was counterbalanced across teams

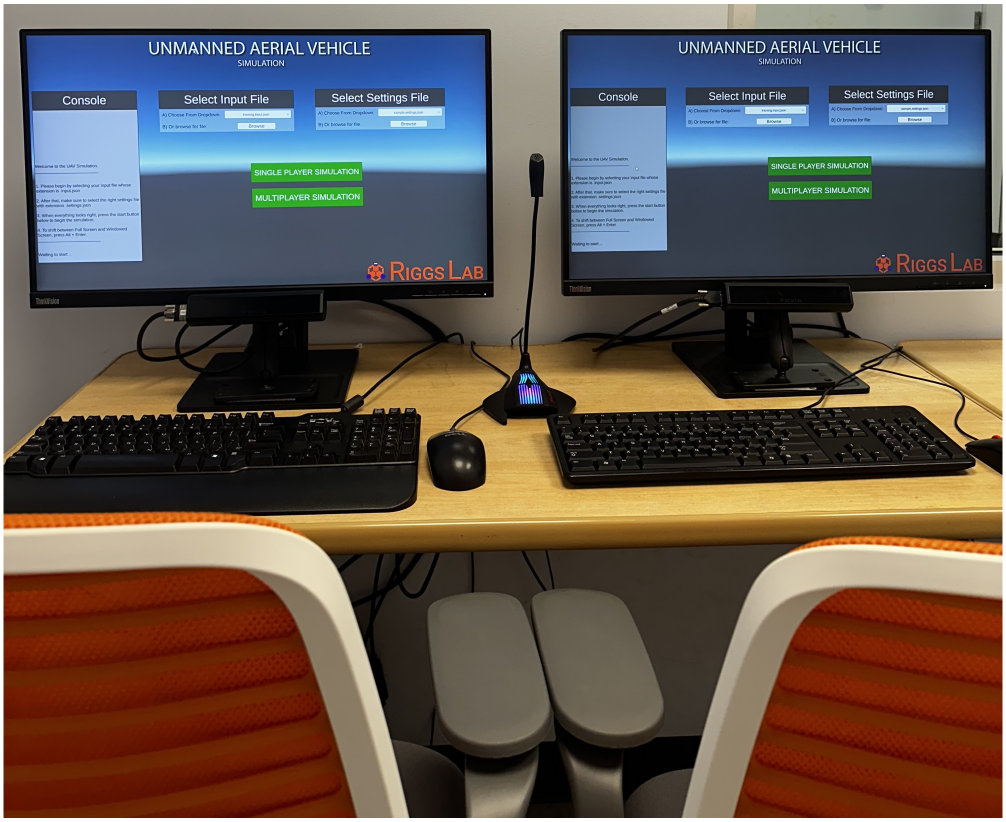


Figure 5.1: Experimental setup with the two networked desktop computers side-by-side with an external microphone in between.

to account for order effects. We maintained a consistent number of targets, reroutings, fuel leaks, and chat messages tasks across all conditions. Each instance of a task was randomized within each condition.

Performance & Workload Measures

Performance was measured using the (a) point system performance metric for overall performance, (b) response time, and (c) accuracy. Same point system for scoring performance was used here (Table 3.2). The points values were assigned to encourage participants to prioritize certain tasks (i.e., target detection). Each participant provided a subjective workload rating at the end of each condition using the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988).

5.3.5 Experimental Procedure

Participants read and signed the consent form, completed a pre-experiment questionnaire (Appendix B), and were briefed about the study goals, tasks that needed to be completed as a team, and how the testbed was networked. This meant that the same tasks occurred in real-time on both monitors, but only one teammate would have to complete every instance of a task. Participants were provided with an overview of the NASA-TLX and how it would be administered. Afterwards, the participants completed a five-minute training session together. By the end of training, participants had to demonstrate they could achieve 70% accuracy across all tasks. They were also provided with three minutes to discuss anything they deemed necessary.

After the training session, the teams completed three 10-minute scenarios. The same tasks appeared at both stations and the actions of each team member were reflected on both workstations, but a participant could not see the cursor movements of their teammate. After each scenario, participants individually completed a NASA-TLX questionnaire to assess workload (Appendix A) and were allowed to re-strategize if they wanted to. At the conclusion of the study, participants filled out a debriefing questionnaire (Table C.3 in Appendix C). The experiment session lasted 60-80 minutes and participants were compensated $15 for their time.

5.3.6 Data Analysis

Eye Tracking Data Analysis

The gaze data was screened to meet data quality requirements as outlined in ISO/TS 15007-2:2014-09, which states that at most 15% data loss is acceptable. The data loss across all participants and trials was on average 8.6% (SD = 2.2%). The same code used in previous studies to filter the eye tracking data and detect fixations and saccades was used in this study (Atweh et al., 2024).

Qualitative Data Analysis

We conducted a qualitative analysis of the debriefing questionnaires anchored in Braun and Clarke (2006)'s procedure for conventional content analysis. This protocol includes developing an overarching impression of the data, inductively eliciting initial categories or codes across the full dataset and grouping codes into defined overarching themes (Braun & Clarke, 2006). This approach aligns with the principles of human-centered design by ensuring data is grounded in participants' lived experiences. After an initial review of 10 questionnaires (5 pairs), a core team of three researchers trained in qualitative methods iteratively developed a codebook by synthesizing information relevant to our research objective. Codebook development concluded when (1) all coders categorized all relevant texts across a response with 100% agreement and (2) all coders agreed that the codebook appropriately and entirely captured all relevant data. The finalized codebook guided the remainder of the analysis.

The research team coded transcripts using NVivo® qualitative data analysis software (QSR International Pty Ltd., 2020). Coders met weekly to assess inter-rater reliability and to ensure data trustworthiness. Any outstanding coder disagreements were brought to the PI for discussion and consensus building. A response was considered reconciled when coding was completed with 100% agreement among all coders. Once coding was complete, the research team grouped codes into overarching themes via iterative group discussions.

5.4 Results

5.4.1 Performance Measures

Figure 5.2 shows the mean and standard error of performance score across the 24 teams for each condition based on the designated scoring convention (Table 3.2). Figure 5.3 presents the mean and standard error of the accuracy by task results across the 24 teams for each condition. Pairs completing the tasks using both gaze sharing and communication yielded the highest total scores (mean = 40,566 points), followed by gaze sharing and no communication (mean = 38,076 points), and the no gaze sharing and communication conditions (mean = 37,558 points).

A one-way repeated measures MANOVA was conducted to check for any statistical difference between the performance measures, (1) total score and (2) accuracy per task. A significant multivariate effect was observed for the within-subjects conditions, $F(12{,}82) = 2.26$, $p = .016$; Wilks' $\Lambda = 0.53$; partial $\eta^2 = 0.25$. Seven follow-up repeated measures univariate ANOVAs showed that both the total point score ($F(2{,}46) = 6.22$, $p = .004$, partial $\eta^2 = .21$) and accuracy for the rerouting task ($F(2{,}46) = 5.72$, $p = .006$, partial $\eta^2 = 0.2$) were statistically significantly different between the three conditions, using a Bonferroni adjusted $\alpha = 0.0083$ level by dividing the standard significance of $\alpha = 0.05$ by the number of tests, which is in this case six.

Post hoc tests using Bonferroni correction revealed that the total score performance was statistically significantly higher in the gaze sharing and communication compared to the no gaze sharing and communication condition ($p = .003$). Similar results were observed in the rerouting task accuracy scores as the pairs completing the tasks using both gaze sharing and communication yielded statistically significantly higher rerouting accuracy (mean accuracy = 71.16%) compared to when gaze sharing was absent and communication was present (mean

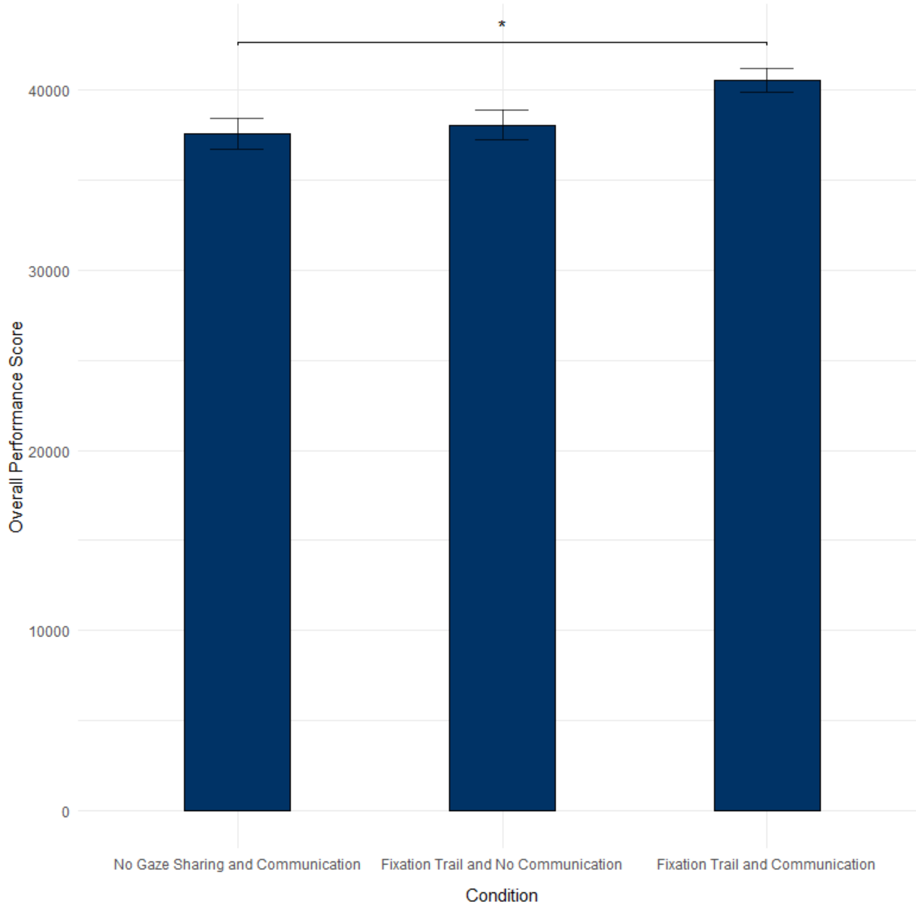


Figure 5.2: Performance scores for each condition. An asterisk (*) indicates significance.

accuracy = 46.96%; p = .003). There were no statistical differences in terms of score for all other pairwise comparisons (all p>.05).

5.4.2 NASA-TLX

Figure 5.4 shows the mean and standard error of the NASA-TLX scores for each of the six dimensions. We decided to analyze the six dimensions separately based on recent recommendations in the literature (i.e., Bolton et al., 2023). A one-way repeated measures MANOVA was conducted to check for any statistical difference between the NASA-TLX scores of the three conditions across the different dimensions. A significant multivariate effect was observed for the gaze conditions $F(12{,}82) = 2.39$, $p = .04$; Wilks' $\Lambda = 0.34$; partial $\eta^2 =$

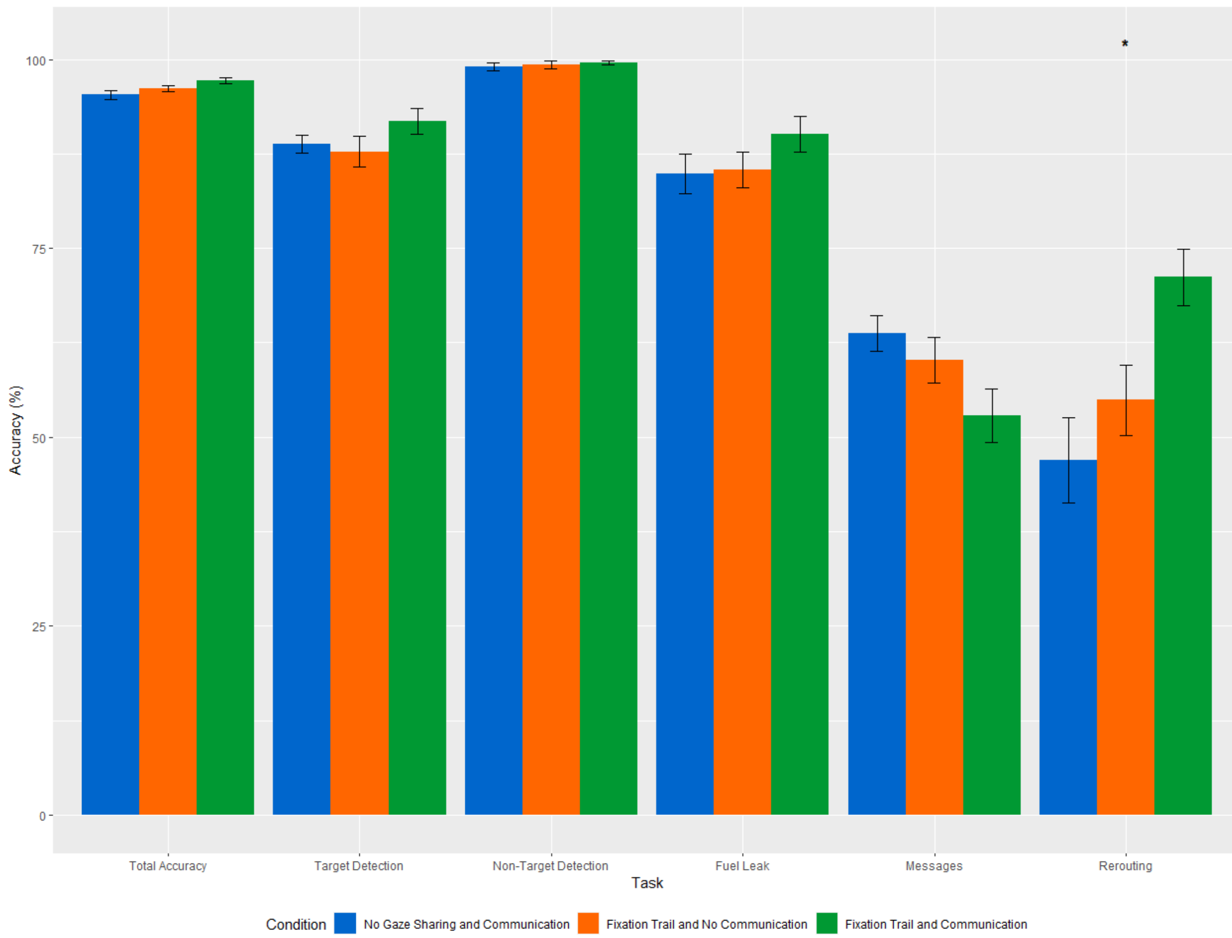


Figure 5.3: Accuracy (%) by task for each condition. An asterisk (*) indicates significant main effects in accuracy for a task.

0.25.

Six follow-up repeated measures univariate ANOVAs showed that the mental, (F(2,46) = 5.61, $p = .007$, partial $\eta^2 = 0.2$) dimension was statistically significantly different between the three conditions, using a Bonferroni adjusted α of 0.0083. Pairs expressed significantly lower mental demand when using both gaze sharing and communication compared to the no gaze sharing and communication condition ($p = .038$) and the gaze sharing and no communication condition ($p = .023$). There were no statistical differences in mental demand between the no gaze sharing and communication condition and the gaze sharing and no communication condition ($p = .9$).

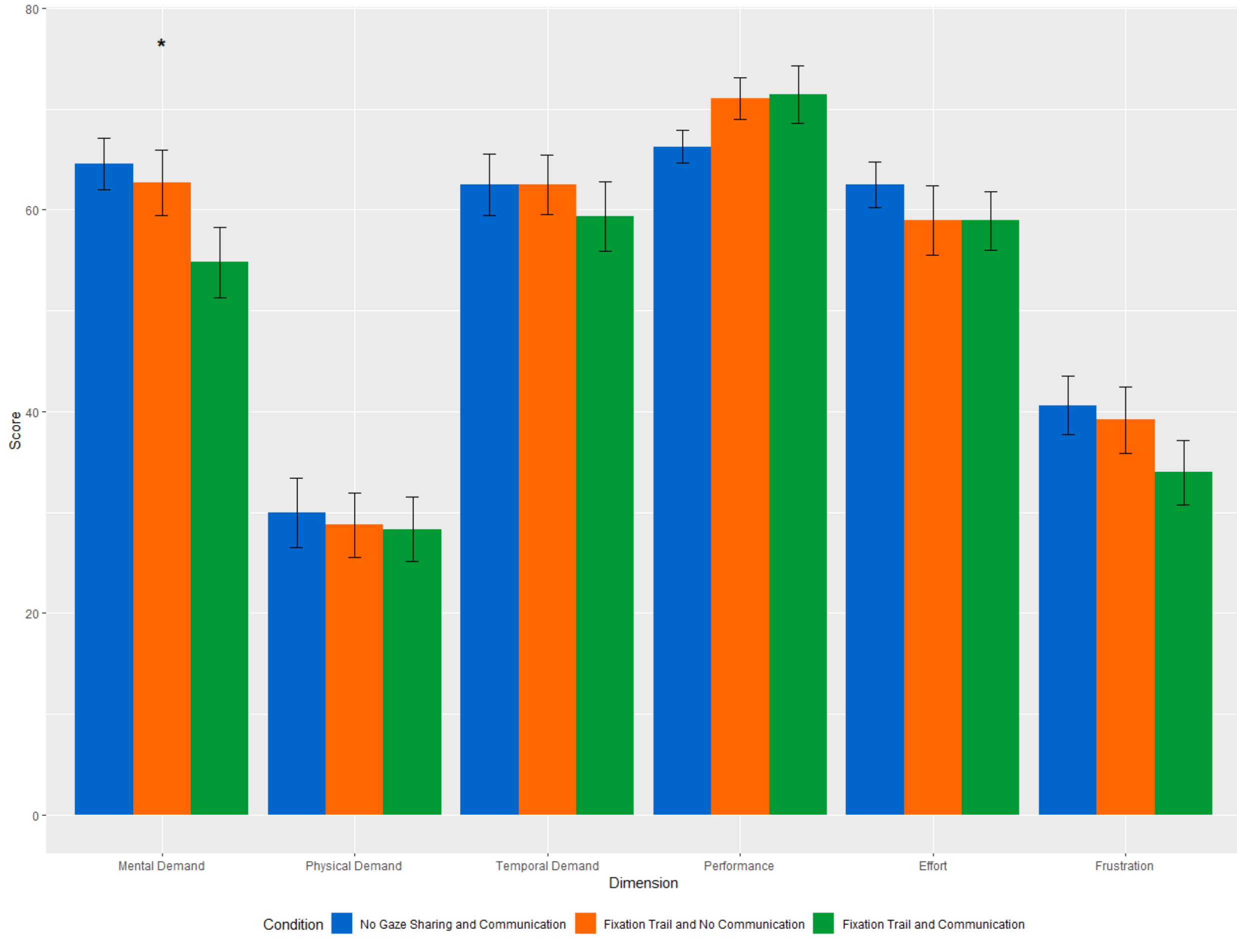


Figure 5.4: NASA-TLX scores for each dimension by condition. An asterisk (*) indicates significant main effects for a dimension.

5.4.3 Eye Tracking Analysis

Figure 5.5 shows the mean and standard error of the six eye tracking metrics (number of fixations, fixation duration, number of saccades, saccade duration, amplitude, and velocity) across the 24 teams for each condition. Six repeated measures univariate ANOVAs revealed that the mean number of fixations ($F(1.28, 29.52) = 9.59$, $p = .002$, partial $\eta^2 = 0.29$), the mean number of saccades ($F(2, 46) = 7.6$, $p = .001$, partial $\eta^2 = 0.25$), saccade duration ($F(1.028, 23.65) = 14.82$, $p < .001$, partial $\eta^2 = 0.39$), and saccade velocity ($F(2, 46) = 8.67$, $p < .001$, partial $\eta^2 = 0.27$) differed statistically significantly between the three conditions. For the number of fixations and saccade duration, we used the Greenhouse-Geisser correction to the ANOVAs due to the violation of the sphericity assumption. This correction adjusts the degrees of freedom to reduce the risk of Type I error, resulting in the decimal degrees of freedom reported.

Post hoc analysis with a Bonferroni adjustment showed significant differences between conditions. Participants exhibited the lowest number of fixations in the no gaze sharing and communication condition compared to both the trail and no communication condition ($p = .004$) and the trail and communication condition ($p = .007$). Participants had significantly higher saccades in the no gaze sharing and communication condition compared to the trail and communication condition ($p < .001$). Saccade duration was significantly longer in the no gaze sharing and communication condition compared to both the trail and no communication condition ($p = .002$) and the trail and communication condition ($p = .003$). Saccade velocity was significantly higher in the no gaze sharing and communication condition compared to the trail and communication condition ($p < .001$). Other pairwise comparisons were not statistically significant ($p > .05$).

5.4.4 Debriefing Questionnaire Thematic Analysis

Overview of Themes

The analysis identified five themes across responses that included: (1) general communication strategies, (2) initial task sharing strategies, (3) adjustments to communication strategies, (4) adjustments to task strategies, and (5) influence of gaze sharing on decision-making. Table 5.2 provides an overview of these themes with representative quotes from participants. Appendix D provides the detailed codebook that guided this thematic analysis.

General Communication Strategies

This theme explores the initial communication strategies participants decided to use dur-

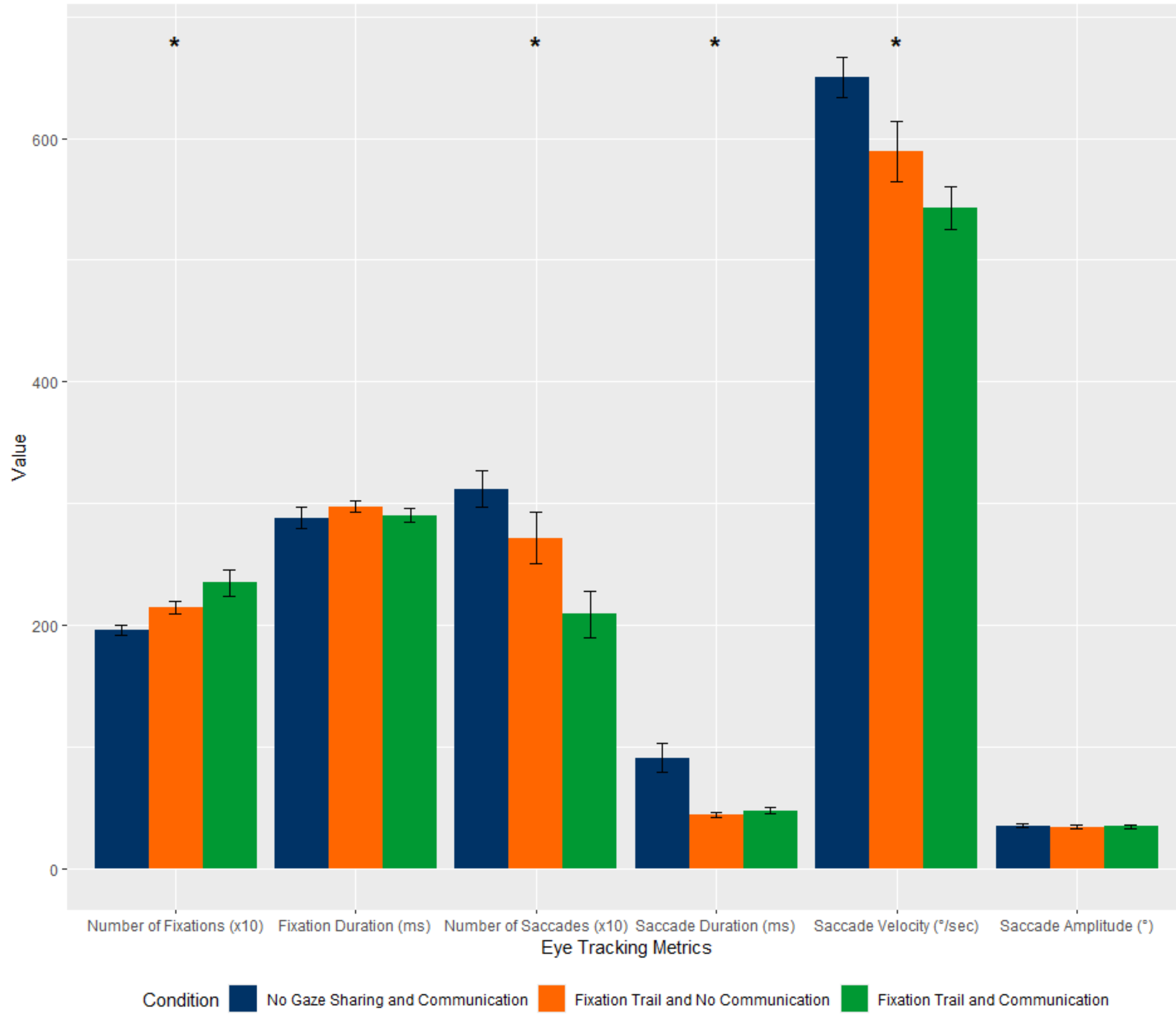


Figure 5.5: Eye tracking metrics for each condition. An asterisk (*) indicates significant main effects for a metric (ms = milliseconds, ° = degrees visual angle)

ing the tasks when communication was allowed. All participants initially decided to use verbal communication as their primary method for coordinating and managing tasks.

The strategies participants employed during the study included sharing updates on the tasks they were actively working on, such as announcing which UAV they were rerouting (e.g., "rerouting Alpha") or specifying the target they were detecting (e.g., "detecting a target at Echo"). Participants also used verbal communication to notify their teammate about chat-related updates (e.g., "There's a message for your task"). For example, one participant reflected on their approach:

> *"We used verbal communication to explain the tasks we were actively completing."*
> *– Pair 7, Participant 2 (P7, P2)*

<u>*Initial Task Sharing Strategies*</u>

This theme encompasses the initial strategies and approaches participants developed to efficiently complete the tasks. It includes methods for dividing and/or sharing tasks. The most common strategy reported by participants was task splitting, where tasks were divided among team members to manage workload effectively. All described some form of task division. For example, one participant noted:

> *"We attempted to split up tasks. I focused on rerouting UAVs and answering the questions, while my partner focused on fixing fuel leaks and identifying targets."*
> *– P7, P2*

Finally, nearly all pairs (95.83%; n=23 pairs) developed strategies for combined task management, where tasks that required joint effort or overlapped, mainly the message task, were managed collaboratively. For instance, one participant explained:

> *"We split the tasks between us, with one focusing on rerouting from NFZs [No-Fly Zones] while the other focused on the targets. We shared the tasks of fixing fuel leaks and answering questions. If the question was about NFZs or targets, the one focusing on that task would answer, otherwise, we would both check." – P14, P1*

In terms of task coordination and updates, participants highlighted the effectiveness of pre-task communication. For example, one participant said:

> *"It was important to have established a plan, but I needed to focus on my tasks, so it was helpful to not communicate during the simulation." – P10, P1*

Specific tasks focus also emerged as a critical moment where communication was necessary, particularly for tasks involving chat messages. A significant portion of participants (91.67%; n = 44 participants) noted that communication was essential in these instances. One participant explained:

> *"Communication was only necessary for the status updates with answers that I didn't know." – P2, P2*

Similarly, another participant observed:

> *"When messages came up, because we both didn't always have the answer to the questions on our own." – P23, P1*

Adjustments to Communication Strategies

Another theme in the data was participants discussing adjustments made to their communication strategies across conditions. Notably, the primary communication strategy revolved around the chat feature, which was essential for completing the chat task due to its reliance on information from all tasks. Beyond this, participants reported talking less frequently during tasks than they initially anticipated in other tasks (e.g., primary task – target detection task).

Despite the prevalent use of verbal communication, many participants reported that minimal verbal interaction was sufficient. In fact, 83.33% of the pairs (n=20) indicated that verbal communication was not necessary for most tasks. As one participant mentioned:

> *"There was really no communication besides a quick verbal remark after losing a UAV." – P19, P1*

Participants agreed that communication became especially important when managing complex tasks that required simultaneous attention. One participant noted:

> *"When there were multiple complex tasks that had to be completed at once, such as when there was a question and planes needed to be rerouted, speaking was helpful to talk through the tasks. My partner was not very communicative, and I feel that if he had spoken more, we may have performed better." – P1, P1*

Another participant echoed this sentiment, stating:

> *"We had to talk more when there were multiple reroutings or multiple targets—when 3-4 tasks required attention or a question was asked about a task the other was not focused on." – P7, P2*

Aside from communicating at critical times, participants naturally spoke more in the absence of gaze sharing. However, when both communication and gaze sharing were permissible, participants noted a reduction in the frequency of verbal communication, as gaze sharing provided sufficient context to reduce the need for explicit verbal exchanges. Team 15 described how the presence of gaze sharing eliminated the need for a verbal confirmation strategy in the target detection task:

> *"When [teammate's name] was detecting a target in the no gaze sharing and talking condition, he was saying he was going to detect the target at India for example. However, we did not have to do this strategy when gaze sharing was allowed as I could see him going to the UAV." – P15, P1*

Their teammate reflected on the effort required to maintain verbal accuracy in the target detection task, which was alleviated by gaze sharing:

> *"We were sharing the target detection task with the 16 UAVs. I found it easier not to say which target I was detecting when there was multiple because I had to actually say the correct name—you don't want to mess that up. So, when gaze sharing was allowed, I didn't have to verbally do that." – P15, P2*

Another participant echoed this and said:

> *"Because I could see where my partner's eyes were focusing, I didn't have to communicate as much if I was busy with another task I had to do." – P13, P1*

Adjustments to Task Strategies

This theme examines how participants adjusted their strategies and coordination methods based on varying conditions, including the presence or absence of communication and the fixation trail. Teams generally maintained consistent task allocation strategies across all conditions, with only minor adjustments in handling tasks that required shared attention, particularly chat messages. In the condition with no gaze sharing but with communication, teams relied heavily on verbal communication to coordinate their actions, especially when addressing chat messages. Participants would verbally inform their teammate when they encountered a question, or in some cases, both teammates would monitor and respond to chat questions simultaneously.

When both gaze sharing and communication were available, 79.16% of the teams (n=19) reported greater flexibility in task management. Participants adjusted their tasks dynamically, often taking on responsibilities that were not initially assigned to them. They could see where their partner was focusing, which allowed them to take over other tasks without needing verbal confirmation. One participant mentioned:

> *"When we could communicate, we used verbal cues to alert the person. With gaze sharing, if I saw my partner was already on a task, I'd just move on to the next one without saying anything." – P22, P1*

In the condition with gaze sharing but no communication, teams' adjustments varied significantly. Two distinct strategies emerged. Some teams found it challenging to fully leverage gaze sharing without verbal communication, leading to missed chat messages and less effective coordination. Pair 3 noted:

> *"When we only had eye tracking [gaze sharing], no verbal communication, we did not really know what to do and more questions were missed like that." – P3, P2*

Other teams successfully utilized gaze sharing to manage tasks efficiently, using their partner's gaze as a cue to adjust their own actions dynamically. This approach allowed them to handle multiple tasks simultaneously, even if they were not initially assigned to those tasks.

Influence of Gaze Sharing on Decision-Making

This theme explores the impact of gaze sharing on decision-making processes, including how it aids or hinders these aspects. Participants reported that gaze sharing significantly influenced their ability to make decisions and coordinate tasks effectively. However, not all participants found gaze sharing beneficial. Some reported that the tool could be distracting, particularly when the gaze tracker moved rapidly across the screen. One participant described this as:

> *"[It was] slightly difficult to overview my partner's performance... the gaze was useful, but it could also be distracting." – P1, P2*

Others found that the constant movement of the gaze tracker degraded their ability to maintain focus, as highlighted by a participant who stated:

> *"I also feel like it was a little bit distracting having the tracker darting around the screen constantly." – P14, P1*

Gaze sharing also played a role in enhancing trust and reliability between teammates. Seeing where their partner was focused helped participants confirm verbal communication and trust that their partner was handling specific tasks. One participant observed:

> *"Seeing their trail gaze confirms what we communicate." – P21, P1*

Conversely, the absence of gaze sharing led to increased difficulty in understanding and trusting a partner's actions, adding pressure to the decision-making process. As one participant expressed:

> *"When the gaze sharing was absent, it was more difficult to understand what my partner was focused on. This put more pressure on me as I was not able to trust/understand what he was doing." – P19, P1*

Preferred Setup

Participants were asked to rank their preferences for the three different setups: 1) No gaze sharing with verbal communication, 2) Trail gaze sharing with verbal communication, and 3)

Trail gaze sharing without verbal communication. Table 5.1 presents participants' rankings of gaze sharing conditions based on preference. The results reveal varying preferences based on how participants perceived the value of gaze sharing and verbal communication in facilitating task completion.

Table 5.1: Participant preferences among the setups

Condition	**Ranked 1st (Most Preferred)**	**Ranked 2nd**	**Ranked 3rd (Least Preferred)**
No Gaze Sharing and Communication	10 (20.8%)	9 (18.8%)	29 (60.4%)
Trail and No Communication	11 (22.9%)	23 (47.9%)	14 (29.2%)
Trail and Communication	27 (56.2%)	16 (33.3%)	5 (10.5%)

Approximately 56% of participants (n=27) ranked "Trail gaze sharing with verbal communication" as their optimal setup. Participants who preferred this setup appreciated the combination of visual and verbal cues, which allowed them to understand their partner's focus and coordinate tasks more effectively. One participant noted that this setup *"was the most optimal because the knowledge of where my teammate's focus is at the moment was an important factor to determine my actions." – P10, P2*

About 23% of participants (n=11) found "Trail gaze sharing without verbal communication" to be the most effective, with another 29% ranking it third (n=14). These participants often felt that verbal communication added unnecessary complexity, and focusing solely on gaze cues made the task simpler. As one participant explained, *"Verbal communication added too much to think about, more than was necessary to complete our tasks." – P3, P2*

Around 60% of participants (n=29) ranked "No gaze sharing with verbal communication" as the least optimal setup. These participants felt that the absence of gaze sharing made it more challenging to coordinate with their partner, even with verbal communication.

Overall, the majority (75%) indicated that the combination of the techniques enhanced shared awareness, while 67% noted improved understanding of their teammate's focus. Half of the participants felt it improved task coordination, though a small percentage (8%) reported no significant effect and 6% indicated it was distracting. When queried about SA, 44% of participants indicated the trail improved their ability to predict their teammate's actions, regardless of whether there was communication. However, 16.67% felt that it did not impact predictability.

Table 5.2: Overview of derived themes, their descriptions, examples, and illustrative quotes.

Theme Title	**Description**	**Example**	**Illustrative Quote**
General Communication Strategy	This theme explores the initial communication strategies participants decided to use during the tasks when communication was allowed.	Verbal Communication	*"We communicated verbally... We decided to say everything we were doing out loud." – P16, P2*
Initial Task Sharing Strategies	This theme encompasses the various strategies and approaches participants developed to efficiently complete the tasks. It includes methods for dividing and/or sharing tasks.	Task Splitting	*"We formed a split strategy where the host looks at the UAV's trajectory and questions, while the client looks at the targets and fuel leaks, as well as some of the questions." – P12, P2*

Theme Title	**Description**	**Example**	**Illustrative Quote**
Adjustments to Communication Strategies	This theme highlights adjustments made to their communication strategies across conditions.	Not talking as much as expected	*"During the chat tasks, communication was especially important since we needed information from all tasks to complete this one." – P24, P2*
Adjustments to Task Strategies	This theme examines how participants adjusted their strategies and coordination methods based on varying conditions, including the presence or absence of communication and the fixation trail.	Adjustments Based on the Condition	*"I found it easier not to say which target I was detecting when there was multiple because I had to actually say the correct name—you don't want to mess that up. So, when gaze sharing was allowed, I didn't have to verbally do that." – P15, P2*
Influence of Gaze Sharing on Decision-Making	This theme explores the impact of gaze sharing on decision-making processes, including how it aids or hinders these aspects.	Real-Time Knowledge of Teammate's Action	*"The absence of gaze sharing made it hard to tell what my teammate was working on; I needed extra effort to predict my teammate's action." – P20, P1*

5.5 Discussion

This study aimed to assess the role of gaze sharing in UAV team operations and its interplay with traditional communication models, exploring whether gaze sharing complements or replaces verbal communication in team tasks. Our findings suggest that gaze sharing is a powerful tool in certain contexts, though its effectiveness varies depending on task requirements and the availability of verbal communication. By discussing the results through emerging themes and their performance implications, we can provide insights into the dynamics of gaze sharing in collaborative environments and its potential for design improvements in UAV C2 systems.

5.5.1 Gaze Sharing *Might* Replace Verbal Communication

Participants exhibited initial communication and task strategies tailored to their understanding of the tasks and the resources available. Regardless of whether verbal communication was permitted, as gaze sharing was introduced, it frequently substituted verbal communication, particularly in the target detection task. This task, being the primary one with the highest points value, required efficient communication and thus making this task shared across many teams. Gaze sharing allowed teammates to discern each other's focus areas instantly, bypassing the time and cognitive effort required to formulate verbal queries or responses. For instance, participants could avoid the challenges of accurately naming UAVs or describing their positions, which are prone to miscommunication. By removing these potential bottlenecks, gaze sharing streamlined collaboration and allowed teams to allocate their cognitive resources more effectively.

To illustrate, participants reported significantly lower mental demand when using both gaze sharing and communication compared to the other conditions. This reduction in workload suggests that gaze sharing offloaded some cognitive burden, allowing participants to allocate resources more efficiently to other tasks. This is further reflected in the eye tracking data, where participants in the gaze sharing conditions exhibited significantly fewer fixations and shorter saccade durations, indicating reduced visual search effort and more efficient attention allocation during shared tasks (Atweh & Riggs, 2024).

This behavioral adjustment in communication strategy is strongly tied to the performance metrics observed in the study. Teams performed significantly better in the "gaze sharing and communication" condition compared to the "no gaze sharing and communication" condition. The ability to rely on gaze sharing for a portion of the shared tasks likely freed up cognitive bandwidth for other tasks, contributing to the overall superior performance of these teams.By observing their partner's gaze, participants could determine whether they were alone in handling shared tasks or if their teammate was actively addressing a specific portion of the task. For instance, if a participant noted their teammate's focus on detecting a target, they could redirect their attention to other tasks, such as fixing a leak or rerouting a UAV. This dynamic allocation of tasks, facilitated by gaze sharing, allowed teams to optimize their workload distribution in real-time, reducing redundancy and enhancing task efficiency (Askar et al., 2024).

This phenomenon directly ties to the accuracy results observed in the study. For the rerouting task, which is a more complex task and requires higher cognitive effort and problem-solving, teams that had access to both gaze sharing and verbal communication achieved significantly higher accuracy scores compared to teams with verbal communication alone. The ability to a shared task (e.g., target detection task) in the hands of their teammate, confident in their partner's attention to it, likely freed participants to focus entirely on the more complex rerouting task. This focused engagement may explain the improved accuracy for rerouting under the combined "gaze sharing and communication" condition.

Conversely, for simpler tasks like fixing leaks, gaze sharing did not yield a statistically significant improvement in accuracy. This is likely because participants had sufficient cognitive bandwidth to manage both the leak task and the target detection task simultaneously, even without fully relying on gaze sharing to coordinate responsibilities. The lower complexity of the leak-fixing task reduced the necessity for precise coordination via gaze sharing, explaining the lack of a significant performance difference.

These findings highlight the potential of gaze sharing for domains where communication latency or restrictions are prevalent, such as remote UAV operation centers or scenarios involving delayed communication (Cardosi & Lennertz, 2017; Magnhild & Braseth, 2020; Pandey et al., 2024). In such environments, traditional verbal coordination may be hindered by time delays or reduced bandwidth, making gaze sharing an invaluable tool for maintaining team performance. Prior research has similarly emphasized the challenges of delayed communication in remote operations and the need for alternative coordination mechanisms to support collaboration (e.g., Bernier-Vega et al., 2023; Bulfone et al., 2020). Incorporating gaze sharing in these contexts could help bridge the gap, enabling teams to sustain high levels of performance and task accuracy despite communication limitations.

5.5.2 Limitations of Gaze Sharing Alone

While gaze sharing effectively replaced verbal communication in several shared tasks (e.g., shared target detection), it was less effective for the chat message task, a secondary but still shared task requiring more nuanced understanding of teammates' progress across multiple sub-tasks. Verbal communication was necessary to clarify intentions, share updates, or coordinate responses to messages. Without the ability to discuss the messages, participants often struggled to integrate gaze information with task requirements, leading to less effective collaboration in this task.

In conditions without verbal communication, the effectiveness of gaze sharing varied significantly among teams. Teams exhibited divergent strategies: some teams assigned the chat task to a single person, resulting in lower scores when coordination failed, while others treated the chat task as shared, enabling them to collaborate more effectively even without verbal communication. These discrepancies likely contributed to the lack of significant performance differences in the "gaze sharing and no communication" condition compared to other conditions.

This highlights a key implication for designers of collaborative systems: gaze sharing, while useful for visual and spatial tasks, cannot fully substitute verbal communication in tasks requiring higher cognitive integration or contextual understanding. Task assignments and sharing strategies must be carefully considered, ensuring that tasks heavily reliant on contextual knowledge are supported by appropriate communication tools. This variation underscores the importance of designing systems that account for diverse team strategies and provide support for teams that may struggle to integrate gaze sharing effectively (Ma et al., 2024). For instance, task assignment tools or visual indicators of partner status could help teams coordinate more efficiently in no-communication environments.

5.5.3 Optimal Setup

The combination of gaze sharing and verbal communication emerged as the most effective condition. Teams leveraged gaze sharing primarily for target detection, using it to quickly identify whether they were alone in the task or which part of the shared task their teammate was handling, which freed up cognitive resources to thus tackle other tasks as seen in the workload results. Verbal communication was then reserved for the chat task, allowing teams to coordinate effectively across both shared and individual tasks. This dual strategy likely explains why participants rated this condition as their preferred one and why it yielded the highest performance among all conditions.

The findings of this study highlight the potential of gaze sharing to enhance team performance in UAV C2 systems, particularly for tasks where rapid visual coordination is critical (Szot et al., 2023; Zheng et al., 2015). However, the limitations of gaze sharing for tasks requiring contextual understanding emphasize the need for balanced system design. Designers should consider incorporating gaze sharing as a complementary tool rather than a replacement for verbal communication, with task-specific supports to maximize its utility.

5.6 Limitations and Future Work

This study has several limitations that should be noted. First, the tasks were specifically designed to simulate operational UAV C2 scenarios, which may limit the generalizability of the findings to other domains. While the results provide valuable insights into team coordination and communication strategies within UAV operations, further studies are needed to assess whether similar patterns emerge in other complex, task-oriented environments. Second, we did not fully assess the interactions between the two independent variables—gaze sharing presence and communication presence. This limitation stems from the use of a fractional factorial design rather than a full factorial design, a decision influenced by the limited funding available at the time of the study. Consequently, while the study provides robust findings within the tested conditions, it does not account for potential interaction effects that could further illuminate the interplay between gaze sharing and communication. Future research should address this gap to provide additional understanding of these variables' combined influence.

Future research should explore ways to optimize gaze sharing interfaces, investigate long-term team adaptation to such tools, and examine how varying task strategies influence the effectiveness of gaze sharing. Additionally, studies could assess the scalability of these findings in larger teams or more complex operational settings, where the dynamics of communication and gaze sharing may differ significantly (Atweh et al., 2022).

Another potential avenue is to introduce an on/off toggle for the fixation trail feature, allowing participants the flexibility to activate or deactivate the trail as needed. Some participants in this study, although limited, noted that the fixation trail could occasionally obscure other information or was not consistently helpful. Providing users with the autonomy to control this feature could possibly help mitigate potential distractions while enabling its use in moments where shared gaze information is critical, such as coordinating on shared tasks or verifying a teammate's focus.

Another important area for future research involves evaluating gaze sharing systems in the context of task interruptions, which are common in UAV C2 operations (Scott et al., 2008). Interruptions can disrupt workflow and hinder awareness, requiring operators to expend additional effort to re-engage with their tasks (Sasangohar et al., 2014). Investigating how gaze sharing features, such as the fixation trail, could support post-interruption recovery would be highly valuable. For instance, observing their partner's gaze after an interruption may help operators quickly understand the team's current focus, identify areas of attention or inattention, and reorient themselves to their role within the task. Assessing the effectiveness of gaze sharing in these scenarios could provide insights into its broader applicability for enhancing resilience and coordination in high-demand environments.

5.7 Conclusion

This study aimed to evaluate the role of gaze sharing in UAV C2 team operations, particularly its interplay with verbal communication, to understand how these tools impact team coordination, workload distribution, and performance. The findings reveal that gaze sharing can act as a powerful complement to verbal communication, particularly for tasks requiring rapid visual coordination. In certain contexts, such as shared target detection, gaze sharing even replaced verbal communication, reducing cognitive effort and miscommunication by enabling teammates to instantly assess each other's focus areas. This dynamic improved efficiency and allowed team members to allocate their cognitive resources to other critical tasks.

However, the effectiveness of gaze sharing was task-dependent. For more complex tasks like rerouting UAVs, the combination of gaze sharing and verbal communication yielded significantly higher accuracy compared to verbal communication alone. In contrast, for simpler tasks such as fixing leaks, the additional coordination provided by gaze sharing was less impactful, as participants had the cognitive bandwidth to handle both shared and individual responsibilities without relying heavily on gaze information.

The study also highlighted the limitations of gaze sharing when used alone. For tasks requiring greater contextual understanding, such as coordinating responses to chat messages, verbal communication remained essential. Teams employing gaze sharing alone displayed diverse and inconsistent strategies, underscoring the need for system designs that support contextual awareness and adaptive task strategies. These findings point to the importance of integrating gaze sharing as a complementary tool rather than a replacement for verbal communication, particularly in complex domains.

Future research should explore innovative gaze sharing displays, such as heatmaps or other visual indicators, and evaluate their applicability in diverse operational domains like air traffic control and emergency response (Schlösser et al., 2018; Zhao et al., 2024). Additionally, introducing features like on/off toggles for fixation trails could give users greater autonomy to tailor the tool to their needs, minimizing distractions while retaining its benefits. Another promising direction involves studying the use of gaze sharing systems during task interruptions, a frequent challenge in UAV C2 operations (Meyer & Schulte, 2020). Understanding how these tools can facilitate re-engagement and coordination post-interruption could expand their utility across a broader range of high-demand environments (Sasangohar et al., 2014).

Ultimately, this research contributes to the broader goal of enhancing safety and performance in collaborative systems. By refining gaze sharing technologies and integrating them into complex team environments, we can foster more effective coordination, reduce errors, and promote resilience in critical operations. These advancements hold promise for supporting teams in high-stakes settings, enabling them to achieve their objectives safely and efficiently.

5.8 References

Arbuckle, N., Taylor, B., & Kancler, D. (2010). The Effect of Communication Modality and Presence of Feedback in Command and Control Environments. *2010 International Symposium on Collaborative Technologies and Systems*, 173–179. https://doi.org/10.1109/CTS.2010.5478511

Askar, S. E., Eid, S. F., & Elshaaer, Y. I. (2024). Dynamic Pickup and Delivery Task Allocation Multi-Agent Robotic Systems: Vacancy Chain Approaches Evaluation. *2024 International Conference on Electrical, Communication and Computer Engineering (ICECCE)*, 1–6. https://doi.org/10.1109/ICECCE63537.2024.10823572

Atweh, J. A., Hazimeh, R., & Riggs, S. L. (2023). Can Real-Time Gaze Sharing Help Team Collaboration? A Preliminary Examination of its Effectiveness with Pairs. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *67*(1), 716–721. https://doi.org/10.1177/21695067231193659

Atweh, J. A., Moacdieh, N. M., & Riggs, S. L. (2022). Identifying Individual-, Team-, and Organizational-Level Factors that Affect Team Performance in Complex Domains

Based on Recent Literature. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1795–1799. https://doi.org/10.1177/1071181322661213

Atweh, J. A., & Riggs, S. L. (2024). Gaze Sharing, a Double-Edged Sword: Examining the Effect of Real-Time Gaze Sharing Visualizations on Team Performance and Situation Awareness. *Human Factors*, *67*(3), 196–224. https://doi.org/10.1177/00187208241272060

Atweh, J. A., Tabbara, I., Nasrallah, G., & Riggs, S. L. (2024). Reducing Barriers in Analyzing Eye Tracking Data: The Development of a GUI to Preprocess Eye Tracking Data. *2024 IEEE Systems and Information Engineering Design Symposium (SIEDS)*, 286–291. https://doi.org/10.1109/SIEDS61124.2024.10534671

Baker, A. L., Fitzhugh, S. M., Huang, L., Forster, D. E., Scharine, A., Neubauer, C., & Cooke, N. J. (2021). Approaches for Assessing Communication in Human-Autonomy Teams. *Human-Intelligent Systems Integration*, *3*(2), 99–128. https://doi.org/10.1007/s42454-021-00026-2

Bernier-Vega, A., Barton, K., Olson, I., Rodriguez, J., Cantu, G., & Ozcelik, S. (2023). Remote Data Acquisition Using UAVs and Custom Sensor Node Technology. *Drones*, *7*(6), 340. https://doi.org/10.3390/drones7060340

Bolton, M. L., Biltekoff, E., & Humphrey, L. (2023). The Mathematical Meaninglessness of the NASA Task Load Index: A Level of Measurement Analysis. *IEEE Transactions on Human-Machine Systems*, *53*(3), 590–599. https://doi.org/10.1109/THMS.2023.3263482

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bulfone, A., Drioli, C., Ferrin, G., & Foresti, G. L. (2020). A Scalable System for the Monitoring of Video Transmission Components in Delay-Sensitive Networked Applications. *Multimedia Tools and Applications*, *79*, 18727–18745. https://doi.org/10.1007/s11042-020-08743-7

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared Mental Models in Expert Team Decision Making. In J. Castellan (Ed.). Lawrence Erlbaum Associates, Inc.

Cardosi, K., & Lennertz, T. (2017). *Human Factors Considerations for the Integration of Unmanned Aerial Vehicles in the National Airspace System: An Analysis of Reports Submitted to the Aviation Safety Reporting System (ASRS)* (tech. rep.). https://rosap.ntl.bts.gov/view/dot/12500

Charapko, A., Ailijiang, A., & Demirbas, M. (2021). Pigpaxos: Devouring the Communication Bottlenecks in Distributed Consensus. *Proceedings of the 2021 International Conference on Management of Data*, 235–247. https://doi.org/10.1145/3448016.3452834

Cheng, X., Li, Y., & Bai, L. (2019). UAV Communication Channel Measurement, Modeling, and Application. *Journal of Communications and Information Networks*, *4*(4), 32–43. https://doi.org/10.23919/JCIN.2019.9005432

Cummings, M. L., Donmez, B., & Graham, H. D. (2008). *Assessing the Impact of Haptic Peripheral Displays for UAV Operators* (tech. rep.). MIT Humans and Automation Laboratory.

Cummings, M. L., & Guerlain, S. (2007). Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles. *Human Factors*, *49*(1), 1–15. https://doi.org/10.1518/001872007779598109

Duan, T., Punpongsanon, P., Jia, S., Iwai, D., Sato, K., & Plataniotis, K. N. (2019). Remote Environment Exploration with Drone Agent and Haptic Force Feedback. *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 167–173. https://doi.org/10.1109/AIVR46125.2019.00034

Giachetti, R. E., Marcelli, V., Cifuentes, J., & Rojas, J. A. (2013). An Agent-Based Simulation Model of Human-Robot Team Performance in Military Environments. *Systems Engineering*, *16*(1), 15–28. https://doi.org/10.1002/sys.21216

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. https://doi.org/10.1016/S0166-4115(08)62386-9

John, P., Brooks, B., Wand, C., & Schriever, U. (2013). Information Density in Bridge Team Communication and Miscommunication—A Quantitative Approach to Evaluate Maritime Communication. *WMU Journal of Maritime Affairs*, *12*, 229–244. https://doi.org/10.1007/s13437-013-0043-8

Kütt, G. H., Lee, K., Hardacre, E., & Papoutsaki, A. (2019). Eye-Write: Gaze Sharing for Collaborative Writing. https://doi.org/10.1145/3290605.3300727

Ma, X., Wang, Y., Wang, J., Xie, X., Wu, B., Li, S., & Wang, Q. (2024). Enhancing Multi-Agent System Testing with Diversity-Guided Exploration and Adaptive Critical State Exploitation. *Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis*, 1491–1503. https://doi.org/10.1145/3650212.3680376

Magnhild, K., & Braseth, A. O. (2020). Operating Autonomous Ships Remotely from Land-Based Operation Centers: The Current State-of-the-Art.

McDermott, P. L., Luck, J., Allender, L., & Fisher, A. (2005). Effective Human to Human Communication of Information Provided by an Unmanned Vehicle. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(3), 402–406. https://doi.org/10.1177/154193120504900340

Meyer, C., & Schulte, A. (2020). Operator Controlled, Reactive UAV Behaviors in Manned-Unmanned Teaming Scenarios with Selective Datalink Availability. *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*, 1673–1679. https://doi.org/10.1109/ICUAS48674.2020.9214018

Mirzaei, M., Gordon, B., Rabbath, C. A., & Sharifi, F. (2010). Cooperative Multi-UAV Search Problem with Communication Delay. *AIAA Guidance, Navigation, and Control Conference*, 8420. https://doi.org/10.2514/6.2010-8420

Nawaz, H., Ali, H. M., & Laghari, A. A. (2021). UAV Communication Networks Issues: A Review. *Archives of Computational Methods in Engineering*, *28*, 1349–1369. https://doi.org/10.1007/s11831-020-09418-0

Pandey, V. K., De, S., & Nandi, S. (2024). Automated Aerial Assessment for Seamless Adaptive Adhoc Restoration in Partially Collapsed Network. *Computer Communications*, *219*, 153–172. https://doi.org/10.1016/j.comcom.2024.03.010

QSR International Pty Ltd. (2020). *NVivo (Version 12, Release 1.6 in March 2022)* [Software]. https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Rizk, Y., Awad, M., & Tunstel, E. W. (2019). Cooperative Heterogeneous Multi-Robot Systems: A Survey. *ACM Computing Surveys (CSUR)*, *52*(2), 1–31. https://doi.org/10.1145/3303848

Sasangohar, F., Scott, S. D., & Cummings, M. L. (2014). Supervisory-Level Interruption Recovery in Time-Critical Control Tasks. *Applied Ergonomics*, *45*(4), 1148–1156. https://doi.org/10.1016/j.apergo.2014.02.005

Schlösser, C., Schröder, B., Cedli, L., & Kienle, A. (2018). Beyond Gaze Cursor: Exploring Information-Based Gaze Sharing in Chat. *Proceedings of the Workshop on Communication by Gaze Interaction*, 1–5. https://doi.org/10.1145/3206343.3206345

Scott, S. D., Wan, J., Sasangohar, F., & Cummings, M. L. (2008). Mitigating Supervisory-Level Interruptions in Mission Control Operations. *Proceedings of 2nd International Conference on Applied Human Factors and Ergonomics*, 14–17.

Stegagno, P., Basile, M., Bülthoff, H. H., & Franchi, A. (2014). A Semi-Autonomous UAV Platform for Indoor Remote Operation with Visual and Haptic Feedback. *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 3862–3869. https://doi.org/10.1109/ICRA.2014.6907419

Sung, G., Feng, T., & Schneider, B. (2021). Learners Learn More and Instructors Track Better with Real-time Gaze Sharing. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), Article 134, 23 pages. https://doi.org/10.1145/3449208

Szot, A., Jain, U., Batra, D., Kira, Z., Desai, R., & Rai, A. (2023). Adaptive Coordination in Social Embodied Rearrangement. *International Conference on Machine Learning*, 33365–33380. https://doi.org/https://proceedings.mlr.press/v202/szot23a.html

Tang, J., Duan, H., & Lao, S. (2023). Swarm Intelligence Algorithms for Multiple Unmanned Aerial Vehicles Collaboration: A Comprehensive Review. *Artificial Intelligence Review*, *56*(5), 4295–4327.

Zhao, Z., Wang, S., Wang, Q., & Shen, D. (2024). Mining Gaze for Contrastive Learning Toward Computer-Assisted Diagnosis. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(7), 7543–7551. https://doi.org/10.1609/aaai.v38i7.28586

Zheng, B., Fung, E., Fu, B., Panton, N. M., & Swanström, L. L. (2015). Surgical Team Composition Differs Between Laparoscopic and Open Procedures. *Surgical Endoscopy*, *29*, 2260–2265. https://doi.org/10.1007/s00464-014-3938-3

Chapter 6

The Role of User-Controlled Gaze Sharing in Managing Interruptions of Teammates Under Varying Task Complexity

6.1 Introduction

Unmanned Aerial Vehicle (UAV) command and control (C2) systems represent a critical component of modern complex operations, spanning domains such as defense, disaster response, and environmental monitoring (Brust et al., 2021; Khan et al., 2022). These systems often involve dynamic teamwork, requiring operators to collaborate under high workload conditions and respond to time-sensitive events (Donmez et al., 2010). Effective communication and coordination among team members are vital for maintaining shared awareness and ensuring mission success. However, in such high-stakes environments, interruptions, whether due to external events, internal task demands, or miscommunication, are common and can compromise team performance, making it imperative to explore strategies that mitigate their impact and facilitate recovery (Bozza et al., 2017; Chen & Barnes, 2014).

One promising approach to improving teamwork and mitigating the effects of interruptions is gaze sharing (Atweh & Riggs, 2024; Atweh et al., 2024). Knowing where a teammate is looking can provide valuable insights into their focus of attention, intentions, and situational priorities, thereby fostering a shared understanding and enhancing coordination. As we saw in chapters 4 and 5, gaze sharing involves visualizing, in real-time, the eye movements of team members on a shared display (Zhang et al., 2017). While this concept has shown potential in various collaborative domains, its application within UAV C2 operations remains relatively unexplored. Our previous work also showed that gaze sharing carries some level of distraction, at least to some individuals. This highlights the need to introduce and research giving the users more autonomy in this technology, such as controlling when to see their partner's eye movement using an on/off toggle (Atweh & Riggs, 2024).

To date, interruptions literature has predominantly focused on the individual level, often examining how a single operator recovers from disruptions (Grundgeiger & Sanderson, 2009). Even in team settings, interruptions are typically studied from the perspective of a supervisory role rather than as a collective team phenomenon (Sasangohar et al., 2014; Scott et al., 2008). This narrow focus overlooks the dynamic and interconnected nature of teamwork in UAV C2 operations, where interruptions experienced by one team member can ripple through the team, affecting overall coordination and performance. There is a pressing need to investigate how interruptions manifest and are managed within teams and to explore interventions that address these challenges holistically.

This chapter aims to address these gaps by examining the role of gaze sharing in facilitating interruption recovery within UAV C2 teams. Specifically, we hypothesize that (1) gaze sharing will aid teams in recovering from interruptions by providing contextual cues about each member's focus of attention, and (2) user-controlled gaze sharing systems will enhance the inclusivity and effectiveness of the tool by allowing team members to tailor its use to their individual needs and task contexts. By enabling team members to activate or deactivate gaze sharing at their discretion, we anticipate that user-controlled systems will mitigate the potential for distraction while maximizing the benefits of shared visual information.

This study holds significant implications for the design of future UAV systems and other complex team-based domains. By shedding light on the interplay between gaze sharing, interruptions, and team dynamics, our findings aim to inform the development of more effective and user-centered tools that enhance teamwork, situation awareness (SA), and resilience in the face of interruptions. Ultimately, these advancements will contribute to the broader goal of improving the safety, reliability, and performance of UAV operations in critical and high-stakes contexts.

6.2 Background

Interruptions are an inherent characteristic of complex systems, often arising due to dynamic and unpredictable environments (de Coning, 2016). They can range from external disruptions, such as alarms and system notifications, to internal cognitive distractions experienced by operators or teams. Managing interruptions effectively is critical in high-stakes domains such as healthcare, aviation, defense, and software development, where they can significantly affect safety, performance, and operational outcomes. While substantial progress has been made in understanding interruptions, significant gaps remain, particularly in areas requiring more nuanced approaches to mitigate their adverse effects.

At the individual level, interruptions disrupt cognitive processes, often leading to task delays, errors, or omissions. Foundational work, such as Trafton et al. (2003) and then by

Altmann et al. (2014), highlighted how interruptions, even momentary ones, can increase cognitive load by diverting attention from primary to secondary tasks. Memory-for-goals theory (Monk et al., 2008) provided a framework for understanding task resumption challenges, emphasizing that resumption is hindered by task complexity and interruption duration.

Recent research has advanced our understanding of interruptions and their management using eye tracking. Katidioti et al. (2016) introduced an interruption management system based on real-time pupil dilation measurement, using it to identify low-workload moments for optimal individual interruption timing. This task-independent system represents a shift towards real-time, adaptive interruption management that minimizes disruption. Similarly, Zhou et al. (2024) investigated interruptions in air traffic control, revealing that modality-specific interruptions (e.g., visual vs. auditory) have differing impacts on SA. Visual interruptions, in particular, were found to cause more significant cognitive disruptions, underscoring the need for interruption-aware interface designs.

In team environments, researchers often study interruptions that affect individual supervisory roles. Sasangohar et al. (2014) demonstrated that providing interruption recovery tools, such as interactive visual timelines, enabled supervisory operators in UAV missions to resume tasks more quickly and accurately. These findings highlight the potential of technology to support recovery in time-critical settings, such as air traffic control and first-responder operations. More recent studies have examined interruptions in collaborative and agile environments. Wiesche (2021) explored interruptions in agile software development teams, identifying three primary types: programming-related work impediments, interaction-related interruptions, and externally imposed interruptions. The study emphasized that interruptions are both a challenge and an enabler in agile settings, facilitating flexibility but also risking productivity loss.

In healthcare, Werner and Holden (2015) synthesized findings on interruptions in emergency departments (EDs), proposing a sociotechnical systems model. This model conceptualized interruptions as a process influenced by interacting system components, offering a comprehensive framework to understand and manage disruptions in complex, real-world environments. Dias et al. (2018) extended this concept by developing an intelligent interruption management system for surgical teams, leveraging real-time cognitive load monitoring to identify low-demand moments for interruptions. This proactive approach minimized disruptions and enhanced patient safety.

Despite these advancements, critical gaps remain. Many studies focus on interruptions in supervisory roles or specific team contexts, such as healthcare or software development, leaving a need for broader exploration of interruptions affecting distributed or non-hierarchical teams (Sasangohar et al., 2014). Additionally, while technologies like adaptive automation and cognitive load monitoring have shown promise, their real-world implementation remains limited, particularly in domains requiring seamless human-computer interaction.

In chapters 4 and 5, we evaluated different gaze sharing techniques—including a dot representation of gaze fixations and a trail showing gaze paths over time—against key performance metrics such as communication dynamics, workload, team SA, eye tracking metrics, and overall team performance (Atweh & Riggs, 2024; Atweh et al., 2023, 2024). Our findings revealed that the trail technique, which allows team members to see the temporal progression of their peers' gaze paths, was particularly effective in enhancing both team SA and performance. This technique enabled team members to better understand each other's focus of attention and decision-making processes, ultimately improving coordination. However, our study also highlighted critical challenges associated with gaze sharing. Notably, approximately 10% of participants reported finding the tool distracting when not actively used. This finding underscores the need for more nuanced gaze sharing mechanisms, such as user-controlled systems that allow team members to toggle gaze sharing on and off based on their immediate needs and task demands.

Further research is necessary to assess user-controlled gaze sharing systems, specifically toggling mechanisms that empower operators to dynamically control their gaze visibility. Such systems hold the potential to balance the benefits of gaze sharing with the risks of cognitive overload, thereby enhancing performance and adaptability in complex, interruption-prone operational environments. Moreover, interruption management in complex systems remains a multifaceted challenge, with research increasingly emphasizing the need for context-specific solutions. While significant progress has been made in understanding interruptions at the individual level, managing team-level disruptions—particularly in UAV command and control (C2) settings—requires further exploration. Emerging technologies, such as user-controlled gaze sharing, coupled with effective training and coordination strategies, offer promising avenues for mitigating the adverse effects of interruptions, ultimately enhancing operational resilience and team efficiency. This study addresses these gaps by focusing on the role of user-controlled gaze sharing as a mechanism to manage interruptions and improve team coordination in UAV operations.

6.3 Methods

6.3.1 Participants

Forty-two teams (84 participants) from the University of Virginia were recruited for the study (M = 22.19 years, SD = 5.26 years). Each team included one male and one female who were not acquainted with one another. Of the participants, 36 individuals (18 teams) were compensated with a $20 gift card, while the remaining 48 individuals (24 teams) received 1.5 class credits for their participation. This study was approved by the University of Virginia's Institutional Review Board (protocol number #3480).

6.3.2 Experimental Setup

The setup was the similar to the setup in Chapters 4 and 5. However, it included an additional two monitors as illustrated below (Figure 6.1).

The experimental setup consisted of four monitors, arranged to facilitate both the ongoing and interrupting tasks. Each participant had access to two monitors: a central monitor for the ongoing UAV task (labeled A1 for Participant 1 and A2 for Participant 2 as shown in Figure 6.1) and a secondary monitor for the interrupting task (labeled B1 and B2, respectively). The arrangement of monitors was B1-A1-A2-B2, ensuring clear separation of tasks while maintaining visibility of both screens (Figure 6.1).

Teams were collocated but viewed separate monitors and interacted with the system using individual mice. Monitors A1 and A2 displayed the main ongoing UAV and the design of the experimental testbed was based on the "Vigilant Spirit Control Station" the U.S. Air Force uses to develop interfaces to control multiple UAVs (Feitshans et al., 2008). The testbed was developed in Unity and ran on two desktop computers (27" 2560 x 1440 monitor; Figure 6.1). The testbed was networked (monitors A1 and A2) so participants could see in real-time inputs their teammate made in the UAV tasks (e.g., when Participant 1 clicked on the target button, Participant 2 could see the response in real-time). However, participants could not see the real-time cursor movements of their teammates. Two desktop-mounted FOVIO eye trackers (60 Hz sampling rate) recorded point-of-gaze data for each participant throughout the experiment. The Multi-Display Module in EyeWorks software enabled simultaneous tracking across both monitor A and B. The average degree of error for the FOVIO eye tracker (determined by the manufacturer) is 0.78° (SD = 0.59°). An external microphone was also used to record all verbal communication. Throughout the experiment, UAV-related ambient sounds were played in the background. These sounds were at a volume that ensured they did not interfere with participants' ability to communicate with one another.

While participants were engaged in the UAV task, the B monitors displayed random UAV-related videos unrelated to the experiment. At random intervals, participants were interrupted individually by a visual and auditory alarm, prompting them to switch from their primary UAV task to an interrupting task displayed on their secondary monitor (B1 or B2). Interruptions were staggered such that participants were never interrupted simultaneously, ensuring that their tasks never overlapped. Upon hearing the alarm, participants were required to leave the primary UAV task to address the interrupting task, temporarily relying on their teammate to manage the UAV operations.

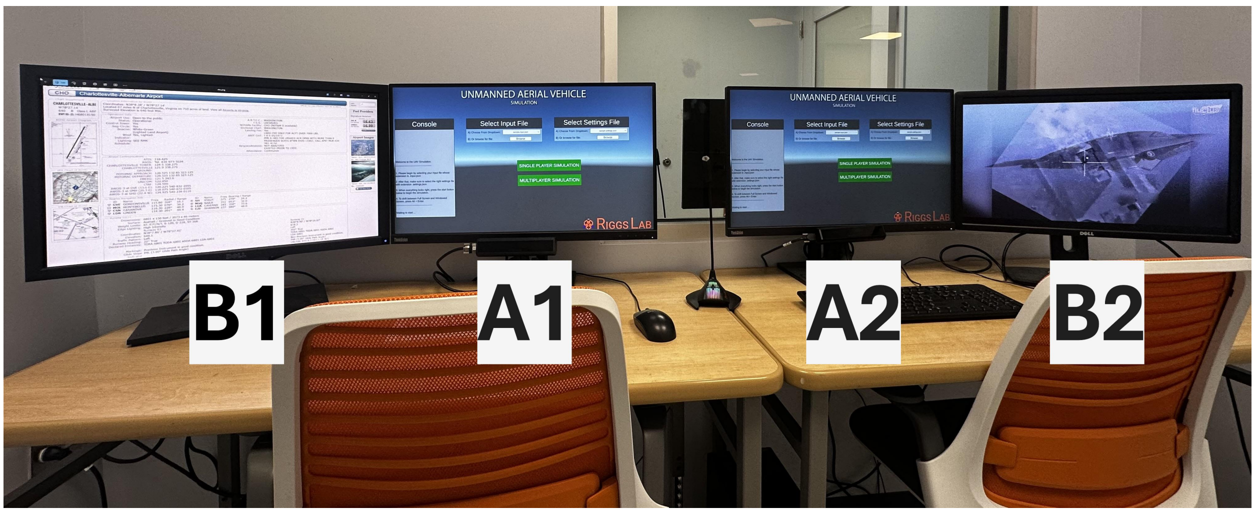


Figure 6.1: Experimental setup showing the arrangement of monitors for each participant. Monitors A1 and A2 displayed the primary UAV tasks and were networked for real-time task updates, while monitors B1 and B2 were used for interrupting tasks. The arrangement (B1-A1-A2-B2) ensured separation of tasks, with teams collocated but operating independently on separate screens. Desktop-mounted eye trackers with the Multi-Display Module in EyeWorks software recorded gaze data, and an external microphone captured verbal communication.

6.3.3 Ongoing Task (UAV Tasks)

Participants were responsible for the same set of tasks (one primary and three secondary) as the previous chapters. Each team was responsible for managing up to 16 UAVs that required them to multitask between four tasks: a primary target detection task and three secondary tasks (replying to chat messages, fixing fuel leaks, and rerouting UAVs away from no-fly-zones; Figure 2). Participants were instructed to share responsibility for both the primary UAV task and the incoming messages. The primary task was emphasized as the most important, requiring collaboration to ensure its successful completion. Sharing responsibility for the messages was also critical, as they contained questions related to all tasks and required attention from both participants. At the start of the experiment, participants were randomly assigned to Seat 1 or Seat 2. The participant in Seat 1 was responsible for handling rerouting tasks, while the participant in Seat 2 managed fuel leak tasks. This division of roles ensured that both participants contributed to distinct yet interdependent aspects of the experiment.

6.3.4 Interrupting Task

While participants were collaboratively engaged in the UAV tasks on their respective A monitors, they were randomly interrupted to complete a visual search task on their individual B monitors. When a participant was interrupted, an auditory and visual alarm was triggered, followed by a screen displaying the message, "Please attend to the following task" (Figure 6.2, Image 1). Participants were instructed to leave the UAV tasks to their teammate and focus entirely on the interrupting task until its completion (indicated by Figure 6.2, Image 5). The teammate who was not interrupted was instructed beforehand to take over all responsibilities for both the primary UAV tasks and message handling whenever their partner was interrupted. This ensured that the primary task remained operational and no critical tasks were neglected during interruptions.

The interrupting task was a visual search task designed to simulate locating key information in an aviation context. The first display indicated that the participant was approaching a specific airport (e.g., Charlottesville Airport in Figure 6.2, Image 2) and prompted them to find a particular frequency (e.g., the airport's ATIS frequency). This display remained visible for 10 seconds, providing the participant with contextual information about the task. Subsequently, an aeronautical chart of the mentioned airport was displayed (Figure 6.2, Image 3), requiring participants to visually search for the requested frequency. This display also lasted for 10 seconds. Once the search period ended, the system presented a display for the participant to input the frequency they found (Figure 6.2, Image 4). This input screen remained active for 10 seconds, regardless of whether the participant successfully located and entered the correct frequency. Participants were explicitly instructed to stay focused on the B monitor for the full duration of the interrupting task, even if they completed the task early or were unable to locate the requested frequency. This ensured that all interruptions were controlled to last exactly 30 seconds, maintaining consistency across participants and trials. Upon completion of the interrupting task, the final screen (Figure 6.2, Image 5) instructed the participant to return to the UAV tasks. Each teammate experienced seven interruptions per condition, with each interruption lasting 30 seconds. Three different airports were used in each condition.

6.3.5 Experimental Design

The study employed a 2×3 mixed factorial design to examine the effects of task complexity (Simple vs. Complex) and gaze sharing condition (Off, On, User-Controlled On/Off Button). A within-subjects design was employed with respect to the gaze sharing condition, where all teams completed three 15-minute scenarios under the following conditions: (a) gaze sharing on, (b) gaze sharing off, and (c) access to a user-controlled on/off gaze sharing toggle. Gaze sharing was implemented using a real-time fixation trail that visually represented the preceding two seconds of gaze behavior (Newn et al., 2017; Figure 6.3). As a reminder, we compared different gaze sharing visualization techniques in Chapter 4, including a fixation dot and a trail. Results indicated that the trail visualization yielded lower workload, higher SA, and better performance compared to the dot visualization. Based on these findings, the

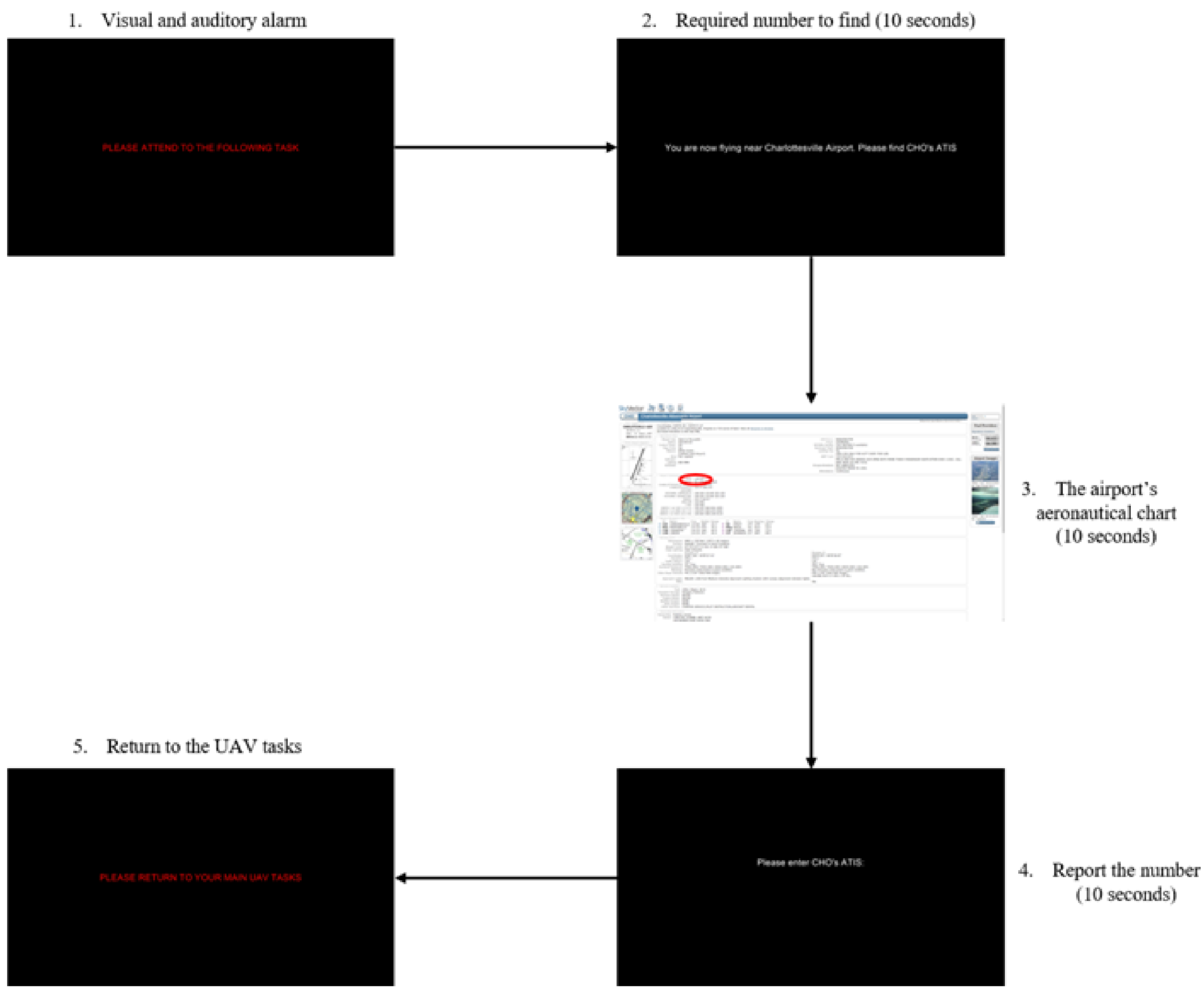


Figure 6.2: Sequence of screens displayed during the interrupting task. Image 1 shows the initial alarm and instructions to attend to the interrupting task. Image 2 indicates the approach to a specific airport (e.g., Charlottesville Airport) and provides task context. Image 3 displays the airport's aeronautical chart for the participant to visually search for the required frequency. Image 4 prompts the participant to input the located frequency. Image 5 concludes the task, instructing the participant to return to the primary UAV task.

fixation trail was chosen for this study (Atweh & Riggs, 2024). To minimize order effects, the sequence of conditions was counterbalanced across teams.

While all teams experienced all three gaze sharing conditions (within-subjects factor), task complexity was nested within teams, meaning each participant remained in their assigned role throughout the experiment. This design allowed for both within-subject comparisons across gaze sharing conditions and between-subject comparisons based on task complexity.

Additionally, the experimental design ensured consistency in the number of targets, rerouting tasks, fuel leaks, and chat messages across all conditions. To introduce variability while maintaining uniformity, the specific instances of each task type were randomized within each condition.

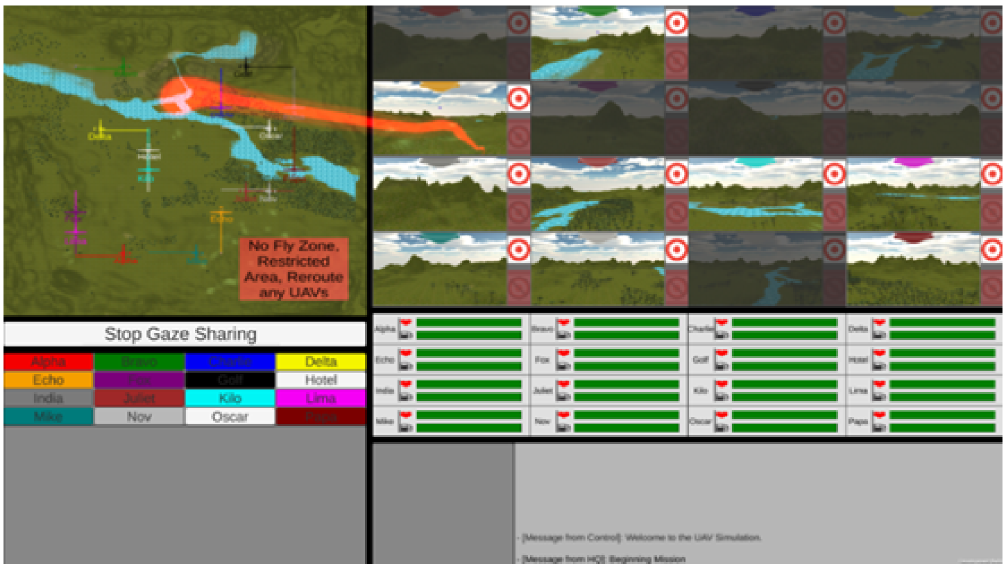


Figure 6.3: The fixation trail gaze sharing visualization technique. The display also includes a "Stop Gaze Sharing" button, allowing participants to toggle the gaze sharing feature on or off during that condition.

Training Session

Participants underwent an extensive 40–45-minute training session to ensure they were well-prepared for the experimental tasks. The training began with an introduction to the UAV tasks, during which participants could ask questions. This was followed by training on the interrupting task, where the experimenter provided detailed instructions on the visual search task and the use of aeronautical charts. The experimenter also demonstrated two example interruptions and answered participant questions to ensure thorough understanding. Although participants were tasked with collaboratively handling target detection and chat messages and dividing responsibilities for rerouting and fuel leak tasks, they were instructed to demonstrate knowledge of all tasks. This was essential because participants were required to take over all UAV tasks when their teammate was interrupted. This ensured that each participant was capable of maintaining team performance during interruptions.

After these instructional sessions, participants engaged in a 15-minute joint training session where they practiced completing the tasks together. During this session, participants were required to achieve at least 70% accuracy to proceed to the main experiment, which all participants successfully achieved. To further simulate the experimental conditions, participants were interrupted three times during this training to ensure they understood how to transition between tasks.

No participants encountered difficulties with the UAV or interrupting tasks during training; therefore, no one was excluded from the study. This comprehensive preparation ensured participants were ready to perform the tasks effectively in the experimental scenarios.

Dependent Measures

Performance evaluation encompassed a comprehensive analysis, utilizing (a) a point system performance metric to gauge overall effectiveness and (b) task-specific accuracy. The same point system for scoring performance was used here (Table 3.2). The points values were assigned to encourage participants to prioritize certain tasks (i.e., target detection). Each participant provided a subjective workload rating at the end of each condition using the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988).

Two primary dependent variables were used to assess participants' ability to recover from interruptions: interruption recovery time (resumption lag) and decision accuracy. Interruption recovery time was defined as the time elapsed between the participant's return to the primary task following the completion of the interrupting task and their first decision to address the post-interruption situation. This metric, commonly referred to as resumption lag (Trafton et al., 2005), captures the cognitive effort required to reorient and re-engage with the primary task. Resumption lag was recorded regardless of whether the first decision made was correct or incorrect. Decision accuracy was evaluated based on the correctness of the participant's first decision following an interruption. Decision accuracy was then binary-coded, with correct responses marked as 1 and incorrect responses or failure to act recorded as 0.

As mentioned, each participant experienced seven interruptions per condition. Interruptions were timed such that upon their conclusion, a new task was immediately presented within the UAV testbed to minimize idle time and maintain consistency. The tasks presented post-interruption were categorized as follows. Primary shared target detection task occurred in 3 out of 7 interruptions. Participants were required to collaborate on identifying targets in the shared video feed. Role-specific tasks occurred in 4 out of 7 interruptions where the post-interruption tasks aligned with the participant's designated responsibilities (simple vs complex task). For teammate A (responsible for the complex task - rerouting UAVs), rerouting tasks were assigned. For teammate B (responsible for the simple task - addressing fuel leaks), fuel leak tasks were assigned. This division ensured that interruptions were systematically integrated into both shared and individual task contexts.

While one teammate was engaged in the interrupting task, the non-interrupted teammate assumed full responsibility for managing the shared and individual tasks. Accuracy was measured by recording the number of tasks successfully completed by the non-interrupted teammate during their partner's absence.

We also analyzed each teammate's performance on the interrupting task aiming to understand how the difference in these scores in each condition, primarily between simple and complex tasks.

6.3.6 Experimental Procedure

Participants began the experiment by reading and signing the consent form and completing a demographic questionnaire (Appendix B). They were then briefed on the study goals, the tasks to be completed as a team, and how the testbed was networked. The explanation highlighted that tasks occurred in real-time on both A monitors but that only one teammate was required to complete each specific task instance. Participants were further briefed on their shared and individual responsibilities and trained in detail on all tasks, as outlined in the training protocol. At the beginning of each condition, the eye trackers were calibrated to ensure accurate data collection, and the simulator was launched to verify that all equipment was functioning correctly. An external microphone was used to record verbal exchanges between participants, with audio recording beginning immediately prior to the start of the conditions. Following the training session, participants completed three 15-minute scenarios, each corresponding to one of the experimental conditions (gaze sharing on, gaze sharing off, and gaze sharing toggle). After each scenario, participants individually completed a NASA-TLX questionnaire to assess perceived workload during the scenario (Appendix A). At the conclusion of the study, participants individually filled out a debriefing questionnaire to provide feedback and insights about their experience (Table C.4 in Appendix C). The total duration of the experimental session, including briefing, training, scenarios, and post-task questionnaires, ranged from 125 to 150 minutes.

6.4 Results

6.4.1 UAV Task Performance

Figure 6.4 shows the mean and standard error of the performance scores across the 42 teams for each condition based on the designated scoring convention (Table 3.2). Figure 6.5 presents the mean and standard error of the task accuracy results across the 42 teams for each condition. Pairs completing the tasks using the on/off gaze sharing toggle yielded the highest total scores (mean = 61,378 points), followed by gaze sharing is on (mean = 55,696 points), and the gaze sharing is off conditions (mean = 48,154 points).

A one-way repeated measures MANOVA was conducted to check for any statistical difference between the performance measures, (1) total score and (2) accuracy per task. A significant multivariate effect was observed for the within-subjects conditions, $F(10{,}156) = 2.31$, $p = .008$; Wilks' $\Lambda = 0.22$; partial $\eta^2 = 0.32$. Six follow-up repeated measures univariate ANOVAs showed that the total point score ($F(2{,}123) = 2.92$, $p = .0058$, partial $\eta^2 = 0.13$), accuracy for the target detection task ($F(2{,}123) = 3.09$, $p = .0049$, partial $\eta^2 = 0.15$), and accuracy for the rerouting task ($F(2{,}123) = 3.96$, $p<.001$, partial $\eta^2 = 0.64$) were statistically significantly different between the three conditions, using a Bonferroni adjusted $\alpha = 0.01$ level by dividing the standard significance of $\alpha = 0.05$ by the number of tests which is in this case six. Sphericity assumptions were met.

Post hoc tests using Bonferroni correction revealed that the total score performance was statistically significantly higher in the user-controlled condition compared to the “Gaze Sharing Is On” ($p = .044$) and “No Gaze Sharing” conditions ($p = .006$). Total performance score was significantly higher when gaze sharing is on compared to when it is off ($p = .034$). Similar results were observed in the rerouting task accuracy and the target detection accuracy scores. Teams completing the tasks with gaze sharing off yielded significantly lower target

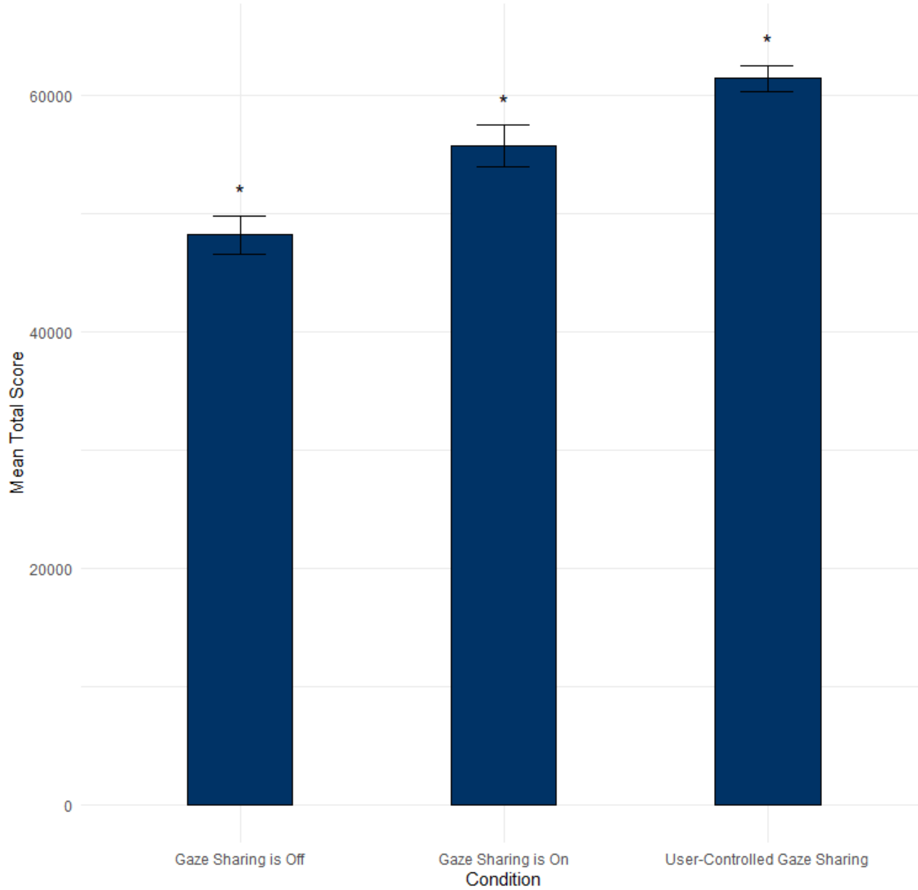


Figure 6.4: Performance scores for each condition. An asterisk (*) indicates significance.

detection accuracy (mean accuracy = 79.39%) compared to the gaze sharing is on (mean accuracy = 88.58%; p = .03) and user-controlled gaze sharing conditions (mean accuracy = 92.31%; p = .027). Similarly, teams completing the tasks with gaze sharing off yielded significantly lower rerouting accuracy (mean accuracy = 52.42%) compared to the gaze sharing is on (mean accuracy = 63.25% p = .018) and user-controlled gaze sharing conditions (mean accuracy = 67.66%; p = .009). There were no statistical differences in terms of score for all other pairwise comparisons (all p>.05).

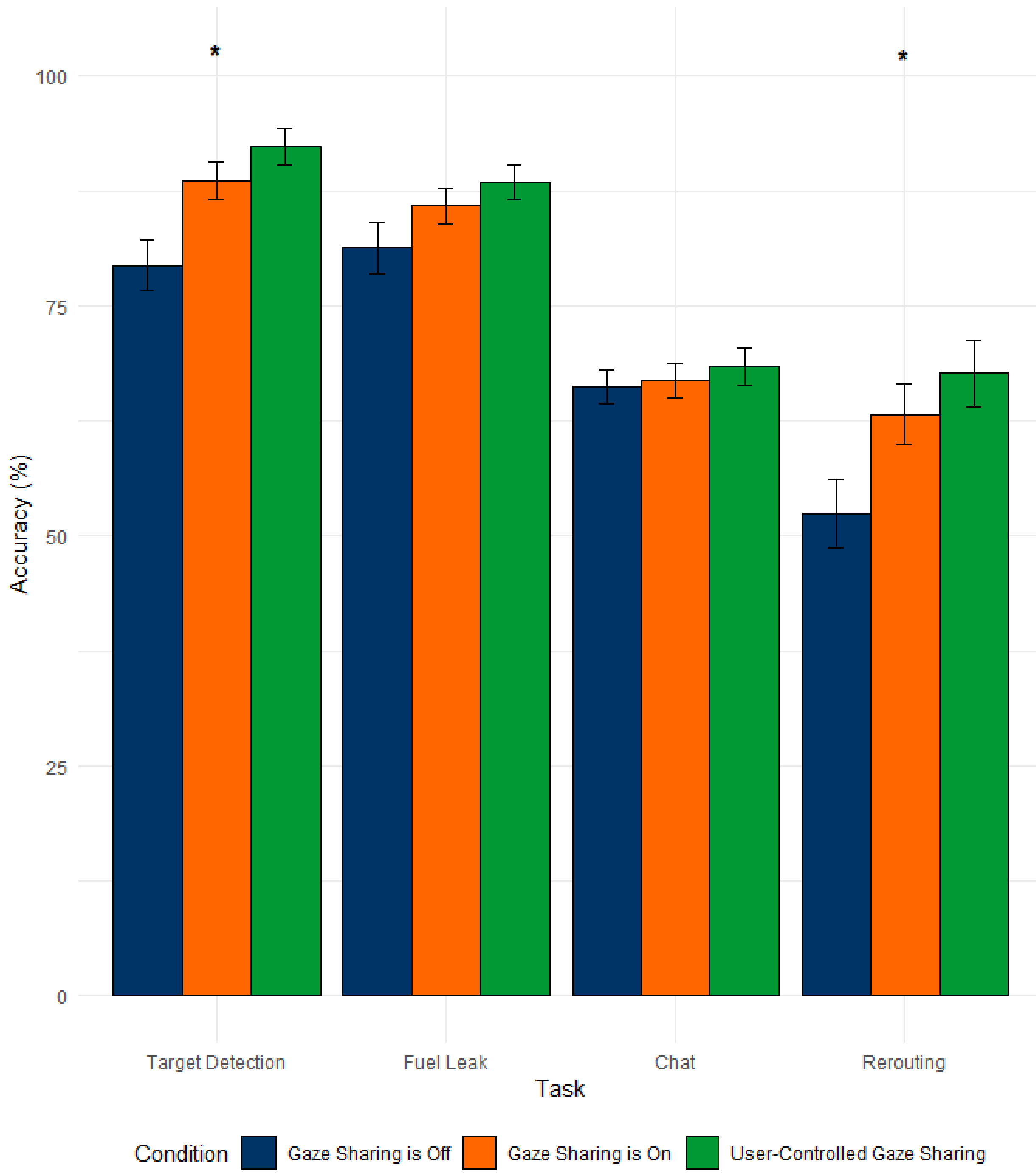


Figure 6.5: Accuracy (%) by task for each condition. An asterisk (*) indicates significant main effects in accuracy for a task.

6.4.2 NASA-TLX

Figure 6.6 shows the mean and standard error of the NASA-TLX scores for each of the six dimensions. We decided to analyze the six dimensions separately based on recent recommendations in the literature (i.e., Bolton et al., 2023). A one-way repeated measures MANOVA was conducted to check for any statistical difference between the NASA-TLX scores of the three conditions across the different dimensions. A significant multivariate effect was observed for the gaze conditions. A significant multivariate effect was observed for the gaze conditions, $F(12{,}488) = 3.22$, $p = .017$; Wilks' Λ=0.41; partial $\eta^2 = 0.63$.

Six follow-up repeated measures univariate ANOVAs showed that the mental ($F(2{,}46) = 1.71$, $p = .002$, partial $\eta^2 = 0.32$) and frustration ($F(2{,}46) = 1.12$, $p = .007$, partial $\eta^2 = 0.14$) dimensions were statistically significantly different between the three conditions, using a Bonferroni adjusted α of .0083. Pairs expressed significantly lower mental demand when using the on/off gaze sharing toggle compared to when gaze sharing is on ($p = .003$) and gaze sharing is off (p<.001). Similar results were seen with frustration as pairs expressed significantly lower frustration levels when using the on/off gaze sharing toggle compared to when gaze sharing is on ($p = .039$) and gaze sharing is off ($p = .004$). Pairs also experienced lower mental demand when gaze sharing was on all the time compared to when it was off all the time ($p = .044$). There were no statistical differences in terms of score for all other pairwise comparisons (all p>.05).

6.4.3 Interrupting Task Performance

Figure 6.7 illustrates the mean interrupting task scores across the three experimental conditions: Gaze Sharing Off, Gaze Sharing On, and Gaze Sharing On/Off Button, for both teammate roles. For Teammate 1 (Complex Task), the highest score was in the "On/Off Button" condition ($M = 5.40$), followed by the "No Gaze Sharing" condition ($M = 5.10$),

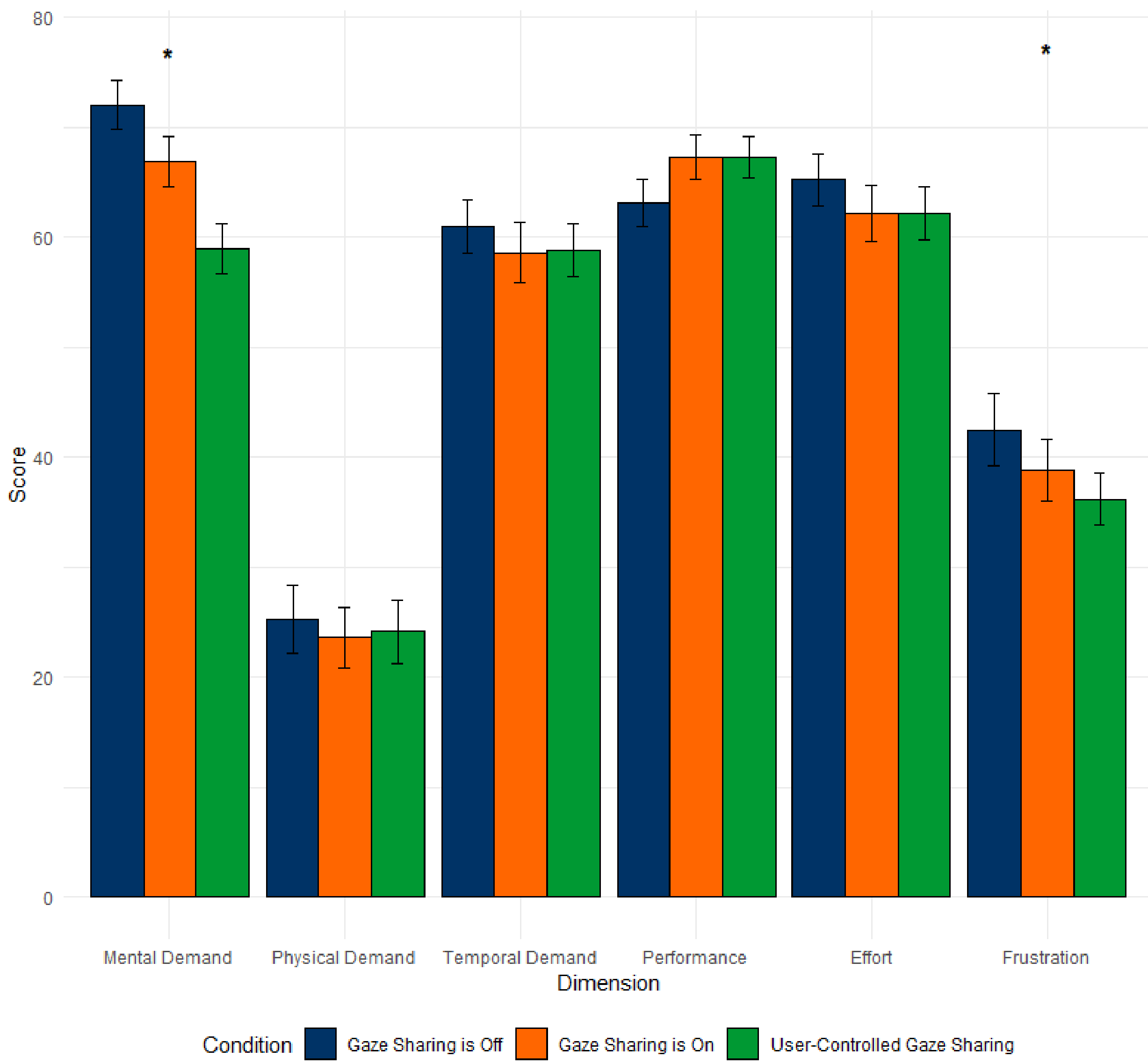


Figure 6.6: NASA-TLX scores for each dimension by condition. An asterisk (*) indicates significant main effects for a dimension.

with the lowest score in the "Gaze Sharing On" condition ($M = 4.48$). For Teammate 2 (Simple Task), the highest scores were in the "No Gaze Sharing" ($M = 5.50$) and "On/Off Button" ($M = 5.50$) conditions, while the lowest score occurred in the "Gaze Sharing On" condition ($M = 4.93$). The results suggest that interruptions had the most impact when gaze sharing was On, with performance improving when gaze sharing was either Off or could be toggled.

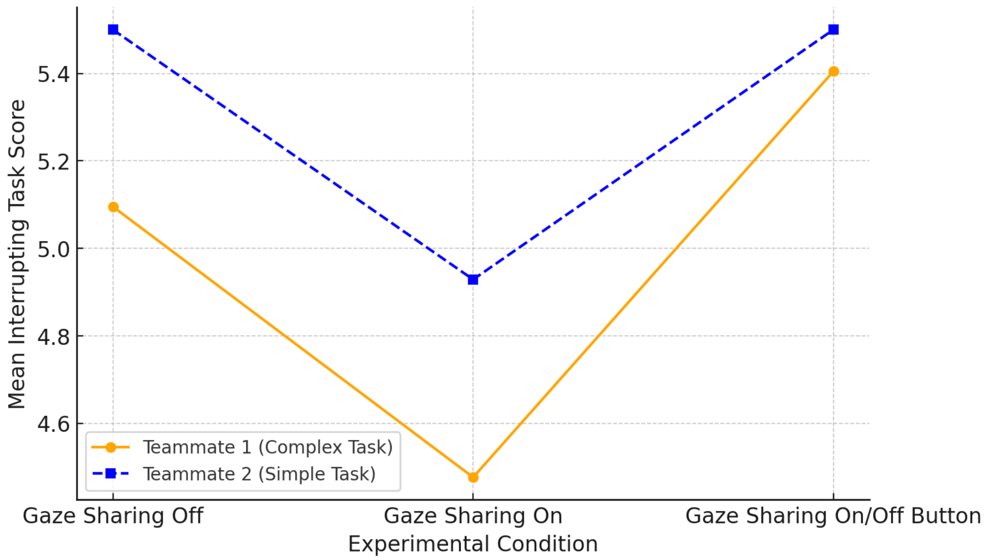


Figure 6.7: Mean interrupting task scores for each teammate across experimental conditions.

Shapiro-Wilk tests were conducted to assess the normality of interruption scores for each condition and each teammate. The results indicated that all conditions violated the assumption of normality ($p < 0.05$) for both Teammate 1 (Complex Task) and Teammate 2 (Simple Task). Given these violations, non-parametric statistical tests were used for further analysis. A Friedman test was performed to determine whether there were significant differences in interruption scores across the three experimental conditions (Gaze Sharing Off, Gaze Sharing On, and Gaze Sharing On/Off Button). The results indicated a significant effect of condition for both teammates. For the simple task, a significant difference was found ($\chi^2(2) = 13.39, p = .0012$), suggesting that the experimental conditions influenced their interruption scores. Similarly, the complex task also showed a significant effect of condition ($\chi^2(2) = 9.93, p = .0069$), indicating that interruptions varied based on gaze sharing conditions.

To identify specific differences between conditions, Wilcoxon signed-rank tests were conducted with a Bonferroni-corrected significance threshold of $p < 0.0167$. For the simple task, a significant difference was observed between "No Gaze Sharing" and "Gaze Sharing On" ($p = .0042$), as well as between "Gaze Sharing On" and "On/Off Button" ($p = .0018$). However, no significant difference was found between "No Gaze Sharing" and "On/Off Button" ($p = .94$). For the complex task, significant differences were observed between "No Gaze Sharing" and "On/Off Button" ($p = .012$) and between "Gaze Sharing On" and "On/Off Button" ($p = .0028$). However, no significant difference was found between "No Gaze Sharing" and "Gaze Sharing On" ($p = .033$).

6.4.4 Decision Accuracy After Interruptions

Shared Tasks

Table 6.1 shows the decision accuracy for the shared tasks after an interruption across different gaze sharing conditions. Decision accuracy varied across gaze sharing conditions. Participants performed best in the "User-Controlled Gaze Sharing On/Off Button" condition, with 192 correct responses out of 252. The "Gaze Sharing is On" condition also showed high accuracy (189 correct responses). The lowest accuracy was observed in the "No Gaze Sharing" condition, where only 139 responses were correct. These results suggest that access to gaze sharing improves decision accuracy in shared tasks, with real-time or user-controlled gaze sharing providing comparable benefits.

Table 6.1: Decision accuracy across different gaze sharing conditions for the shared task.

Condition	**Correct**	**Incorrect**	**Total**
Gaze Sharing is Off	139	113	252
Gaze Sharing is On	189	63	252
User-Controlled Gaze Sharing On/Off Button	192	60	252

Cochran's Q Test was conducted to determine whether and how decision accuracy varied across all gaze sharing conditions. Pairwise McNemar's Tests were then performed to identify which conditions significantly differed. For the shared task, Cochran's Q Test indicated a significant effect of Gaze Sharing Condition ($Q(2) = 45.32, p < .0001$), confirming that accuracy differed across conditions. McNemar's Tests with continuity correction showed that accuracy was significantly higher in the "Gaze Sharing Is On" condition compared to "No Gaze Sharing" ($\chi_c^2(1) = 21.77, p < .0001$) and in the "On/Off Button" condition compared to "No Gaze Sharing" ($\chi_c^2(1) = 23.67, p < .0001$). However, there was no significant difference between "Gaze Sharing Is On" and "On/Off Button" ($\chi_c^2(1) = 0.01, p = .918$), suggesting that both gaze sharing mechanisms similarly enhanced post-interruption performance in shared tasks.

Individual Tasks (Simple vs Complex)

Table 6.2 shows the first decision accuracy for the simple task (fuel leak) across different gaze sharing conditions. Decision accuracy varied across gaze sharing conditions. Participants performed best in the "Gaze Sharing is On" condition, with 154 correct responses out of 168. The "User-Controlled Gaze Sharing On/Off Button" condition also showed high accuracy (149 correct responses). The lowest accuracy was observed in the "No Gaze Sharing" condition, where only 131 responses were correct. These results suggest that access to gaze sharing improves decision accuracy in simple tasks, with real-time gaze sharing providing the highest benefit.

Table 6.2: Decision accuracy across different gaze sharing conditions for the simple task (fuel leak).

Condition	**Correct**	**Incorrect**	**Total**
Gaze Sharing is Off	131	37	168
Gaze Sharing is On	154	14	168
User-Controlled Gaze Sharing On/Off Button	149	19	168

Similarly, table 6.3 shows decision accuracy for the complex task (rerouting) across different gaze sharing conditions. Participants performed best in the "User-Controlled Gaze Sharing On/Off Button" condition, with 138 correct responses out of 168, followed by the "Gaze Sharing is On" condition with 126 correct responses. The lowest accuracy was observed in the "No Gaze Sharing" condition, where only 109 responses were correct. These results suggest that having control over gaze sharing may provide additional benefits in more cognitively demanding tasks.

Table 6.3: Decision accuracy across different gaze sharing conditions for the complex task (rerouting).

Condition	Correct	Incorrect	Total
Gaze Sharing is Off	109	59	168
Gaze Sharing is On	126	42	168
User-Controlled Gaze Sharing On/Off Button	138	30	168

A Type III Wald Chi-Square Test was conducted to examine the effects of Gaze Sharing Condition (Off, On, On/Off) and Task Complexity (Simple vs. Complex) on decision accuracy after interruptions. Results indicated significant main effects of Condition ($\chi^2(2) = 27.29, p < .001$) and Task Complexity ($\chi^2(1) = 0.95, p = .026$), as well as a significant Condition × Task Complexity interaction ($\chi^2(2) = 17.34, p < .001$) on decision accuracy.

To further investigate these effects, a Generalized Linear Mixed Model (GLMM) with a binomial logit link was performed to examine the influence of Gaze Sharing Condition and Task Complexity on decision accuracy after interruptions. Results revealed that both "Gaze Sharing Is On" ($\beta = 16.57, p < .001$) and "User-controlled Gaze Sharing On/Off Button" ($\beta = 22.36, p < .001$) conditions significantly improved decision accuracy compared to "No Gaze Sharing". Task Complexity also had a significant effect ($\beta = -9.55, p = .026$), indicating that accuracy was significantly lower in complex tasks overall. The "Gaze Sharing is On" condition did not exhibit a significant interaction with Task Complexity ($\beta = -7.2, p = .021$), suggesting it provided a similar benefit across both task types. However, having access to the user-controlled gaze sharing indicated that it was significantly more effective in complex tasks rather than simple tasks ($\beta = 12.5, p = .017$).

Cochran's Q Test was conducted separately for simple and complex tasks to determine whether and how decision accuracy varied across all gaze sharing conditions. Pairwise Mc-

Nemar's Tests were then performed to identify which conditions significantly differed.

For the fuel task (simple), Cochran's Q Test indicated a significant effect of Gaze Sharing Condition ($Q(2) = 38.17, p < .0001$), confirming that accuracy differed across conditions. McNemar's Tests with continuity correction showed that accuracy was significantly higher in the "Gaze Sharing Is On" condition compared to "No Gaze Sharing" ($\chi^2_c(1) = 21.04, p < .0001$) and in the "On/Off Button" condition compared to "No Gaze Sharing" ($\chi^2_c(1) = 16.06, p < .0001$). However, there was no significant difference between "Gaze Sharing Is On" and "On/Off Button" ($\chi^2_c(1) = 3.2, p = .074$), suggesting that both gaze sharing mechanisms similarly enhanced post-interruption performance in Simple tasks.

For the rerouting task (complex), Cochran's Q Test again revealed a significant effect of Gaze Sharing Condition ($Q(2) = 66.05, p < .001$), indicating that accuracy varied across conditions. McNemar's Tests with continuity correction found that both "Gaze Sharing Is On" and "On/Off Button" conditions significantly improved accuracy compared to "No Gaze Sharing" ($\chi^2_c(1) = 15.05, p < .001$; $\chi^2_c(1) = 27.03, p < .0001$, respectively). Unlike simple tasks, the user-controlled technology also significantly outperformed having gaze sharing on in complex tasks ($\chi^2_c(1) = 10.08, p = .0015$), suggesting that user-controlled gaze sharing provides a unique advantage under higher cognitive demands.

6.4.5 Resumption Lag

Figure 6.8 shows the mean resumption lag times (in seconds) across the three experimental conditions, for both teammate roles. For Teammate 1 (Complex Task), the highest resumption lag was observed in the "No Gaze Sharing" condition ($M = 1.75s$), followed by "Gaze Sharing On" ($M = 1.44s$) and "On/Off Button" ($M = 1.32s$). A one-way repeated measures ANOVA confirmed a significant effect of gaze sharing condition on resumption lag ($F(2, 82) = 89.07, p < .001, \text{partial } \eta^2 = 0.69$). Post hoc tests using Bonferroni correction showed that the "No Gaze Sharing" condition resulted in significantly higher resumption lag compared to "Gaze Sharing On" ($p < .001$) and "On/Off Button" ($p < .001$). Additionally, teams in the "On/Off Button" condition exhibited significantly faster resumption times compared to teams in the "Gaze Sharing On" condition ($p < .001$), confirming that user control over gaze sharing was particularly beneficial for complex tasks.

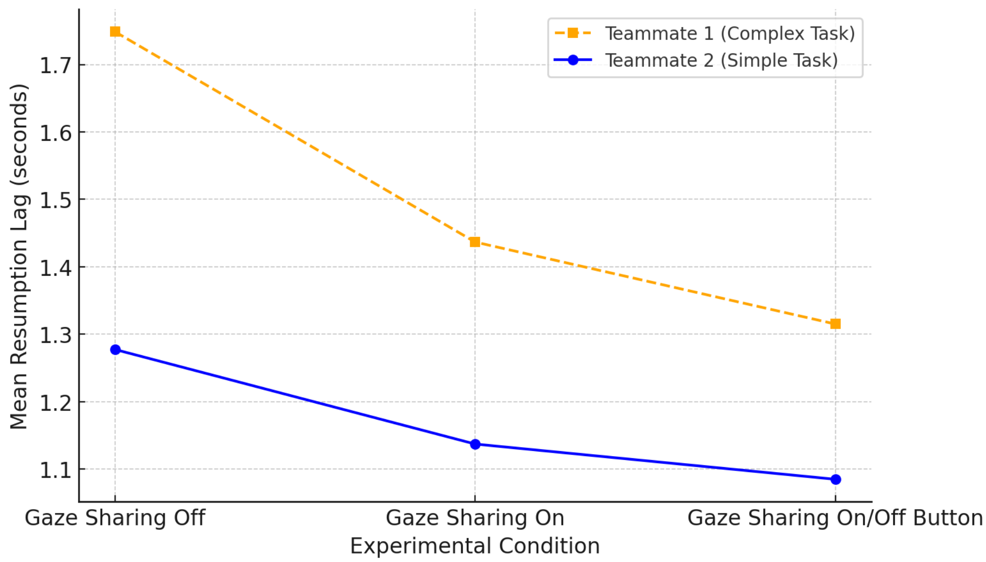


Figure 6.8: Mean resumption lag (in seconds) across the three experimental conditions for both teammate roles.

For Teammate 2 (Simple Task), the highest resumption lag was observed in the "No Gaze Sharing" condition ($M = 1.28s$), followed by "Gaze Sharing On" ($M = 1.14s$) and "On/Off Button" ($M = 1.08s$). A one-way repeated measures ANOVA confirmed that there was a statistically significant effect of gaze sharing condition on resumption lag ($F(2, 82) = 16.77, p < .001, \text{partial } \eta^2 = 0.29$). Post hoc tests using Bonferroni correction revealed that the "No Gaze Sharing" condition resulted in significantly higher resumption lag compared to both "Gaze Sharing On" ($p < .001$) and "On/Off Button" conditions ($p < .001$). However, the difference between "Gaze Sharing On" and "On/Off Button" was not statistically significant ($p = .27$), suggesting that user control over gaze sharing did not provide additional benefits in simple tasks.

6.4.6 Eye Tracking Analysis

Figure 6.9 shows the mean and standard error of the six eye tracking metrics (number of fixations, fixation duration, number of saccades, saccade duration, saccade velocity, and saccade amplitude) across the 42 teams for each condition.

Six repeated measures univariate ANOVAs revealed that the mean number of fixations ($F(1.45, 59.62) = 8.34$, $p = .003$, partial $\eta^2 = 0.22$), the mean number of saccades ($F(2, 82) = 6.91$, $p = .002$, partial $\eta^2 = 0.19$), saccade duration ($F(2, 82) = 5.45$, $p = .006$, partial $\eta^2 = 0.15$), and saccade velocity ($F(2, 82) = 7.24$, $p < .001$, partial $\eta^2 = 0.42$) differed statistically significantly between the three conditions. For the number of fixations, we used the Greenhouse-Geisser correction due to the violation of the sphericity assumption. This correction adjusts the degrees of freedom to reduce the risk of Type I error, resulting in the decimal degrees of freedom reported.

Post hoc analysis with a Bonferroni adjustment showed significant differences between conditions. The number of fixations was significantly lower in the "Gaze Sharing On" condition compared to "No Gaze Sharing" ($p = .007$) but did not differ significantly between "Gaze Sharing On" and "User-Controlled On/Off" ($p = .21$) or between "No Gaze Sharing" and "User-Controlled On/Off" ($p = .43$). The mean number of saccades was significantly lower in the "Gaze Sharing On" condition compared to "No Gaze Sharing" ($p = .004$), but no significant differences were found between "Gaze Sharing On" and "User-Controlled On/Off" ($p = .36$) or between "No Gaze Sharing" and "User-Controlled On/Off" ($p = .3$).

Saccade duration was significantly greater in "No Gaze Sharing" compared to "Gaze Sharing On" ($p = .011$), but no significant differences were found between "Gaze Sharing On" and "User-Controlled On/Off" ($p = .27$) or between "No Gaze Sharing" and "User-Controlled On/Off" ($p = .32$). Finally, the post hoc analysis also showed that saccade velocity was significantly higher in the "No Gaze Sharing" condition compared to both

"Gaze Sharing On" ($p < .001$) and "User-Controlled On/Off" ($p < .001$). Moreover, saccade velocity was significantly higher in the "Gaze Sharing On" compared to the "User-Controlled On/Off" ($p < .001$).

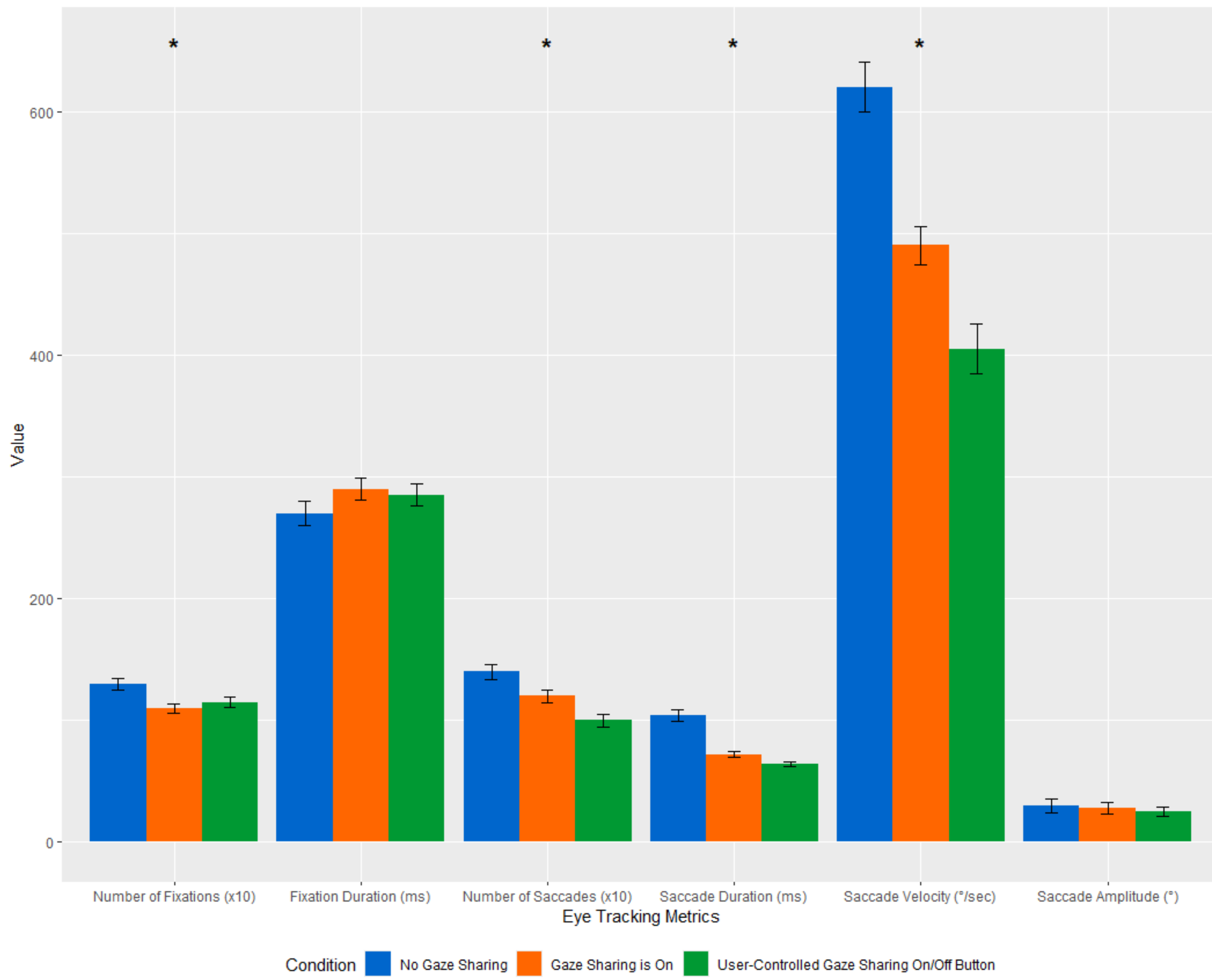


Figure 6.9: Eye tracking metrics for each condition. An asterisk (*) indicates significant main effects for a metric (ms = milliseconds, ° = degrees visual angle)

Figure 6.10 shows the mean and standard error of the pupil diameter in each condition. A repeated-measures ANOVA was also conducted to examine the effects of Gaze Sharing Condition on pupil diameter and results indicated a significant main effect, $F(1.32, 54.72) = 10.21$, $p < .001$, partial $\eta^2 = 0.26$. Post hoc analysis with a Bonferroni adjustment showed that pupil diameter was significantly higher in "No Gaze Sharing" compared to both "Gaze Sharing On" ($p = .003$) and "User-Controlled On/Off" ($p = .008$). Other pairwise comparisons were not statistically significant ($p > .05$).

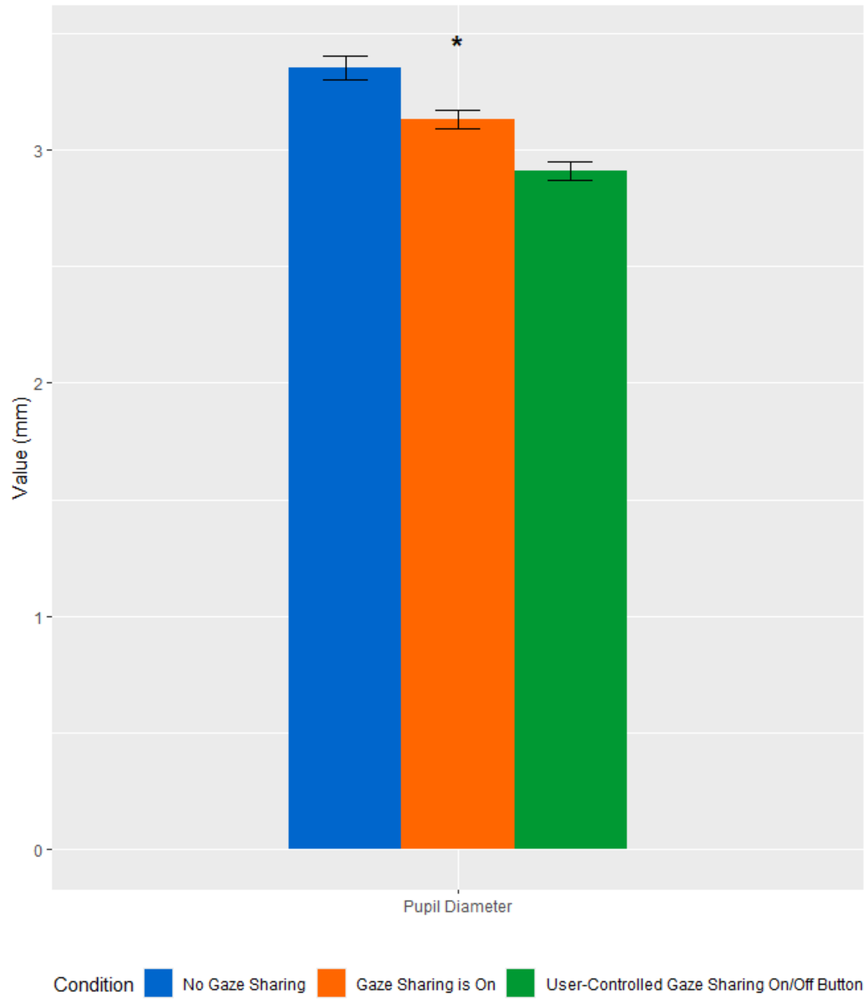


Figure 6.10: Pupil diameter (mm) for each condition. An asterisk (*) indicates significant main effects.

Figure 6.11 shows the mean and standard error of the time to fixation (TTFF) in each condition. Similarly, a repeated-measures ANOVA was conducted to examine the effects of Gaze Sharing Condition on TTFF upon returning after an interruption. Results indicated a significant main effect, $F(2, 82) = 9.41$, $p < .001$, partial $\eta^2 = 0.28$. Post-hoc tests using the Bonferroni adjustment revealed that participants in the "No Gaze Sharing" condition exhibited the longest time to first fixation, which was significantly greater than both "Gaze Sharing On" ($M = 0.72s, p < .001$) and "User-Controlled On/Off" ($p = .002$). A significant difference was also found between "Gaze Sharing On" and "User-Controlled On/Off" ($p =$

.003), with TTFF being the lowest in the latter.

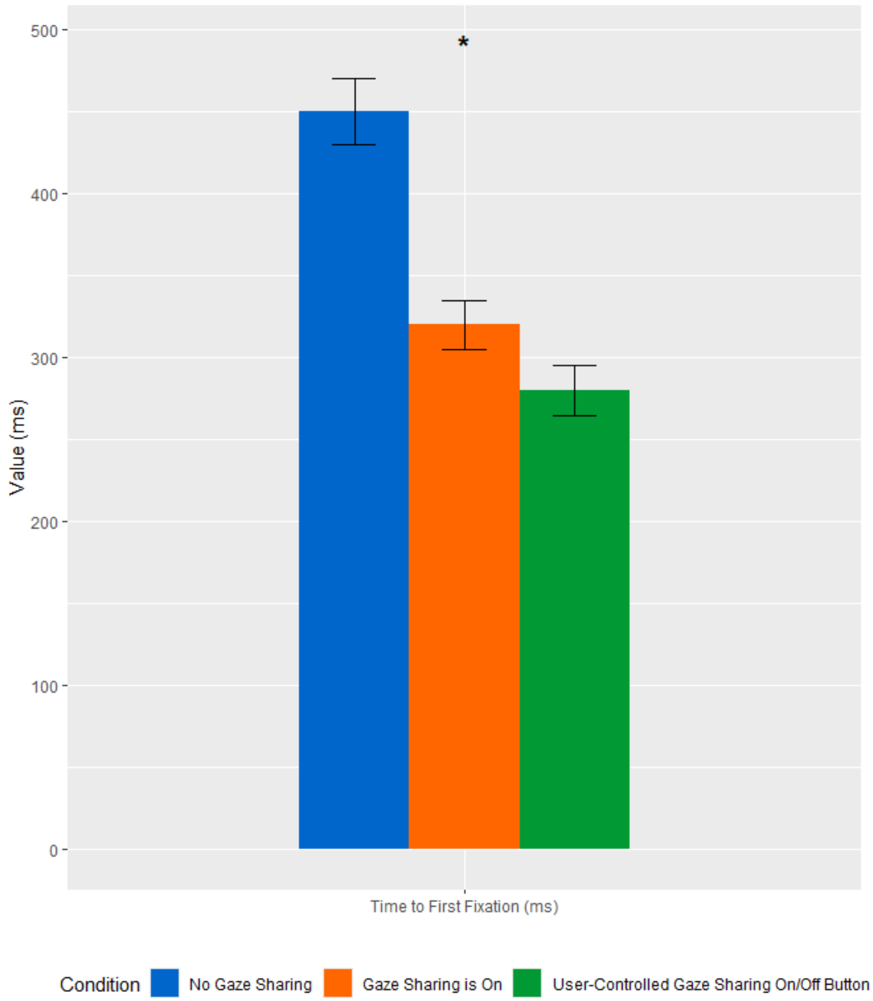


Figure 6.11: Time to first fixation (TTFF) for each condition following interruptions. An asterisk (*) indicates significant main effects.

6.4.7 Debriefing Questionnaire Analysis

Participants ranked their preference among the three gaze sharing conditions: "Gaze Sharing Always On", "No Gaze Sharing", and the "On/Off Gaze Sharing Button". Table 6.4 presents participants' rankings of gaze sharing conditions based on preference. The results indicate that the most preferred condition was the "On/Off Gaze Sharing Button", with 51.2% (n = 43) of participants ranking it as their top choice. Participants appreciated the flexibility and control this condition provided, with one noting, *"I liked having control over when I could see my teammate's gaze and when I wanted to focus on my own tasks"* – Pair 6, Participant 1 (P6, P1). Another participant emphasized the benefit of adaptability, stating, *"Sometimes I needed my teammate's gaze to coordinate better, but other times it was distracting. Having the option to toggle it on and off was the best solution"* – P32, P2.

The second most preferred condition was "Gaze Sharing Always On", selected as the top choice by 28.6% ($n = 24$) of participants. Many participants found continuous gaze sharing beneficial for maintaining awareness and reducing the need for verbal communication. One participant shared, *"It helped me stay aligned with my teammate's actions without needing additional explanations"* – P21, P1. Another remarked, *"I always knew what my teammate was looking at, which made coordination much easier."* However, some participants found the constant gaze sharing to be distracting or overwhelming, contributing to 33.3% ($n = 28$) ranking it as their least preferred condition. One participant explained, *"It felt like too much visual clutter, making it harder to concentrate on my responsibilities"* – P7, P1.

When uninterrupted, participants generally preferred gaze sharing (without specifying if continuous or user-controlled), as it provided a clear visual indicator of their teammate's presence and engagement in the task. While their teammate was away handling an interruption, the absence of their gaze on the screen reinforced the understanding that they were solely responsible for managing all tasks. Conversely, when their teammate returned, the reappearance of their gaze offered reassurance and a renewed sense of collaboration. One participant noted, *"When my teammate's gaze disappeared, I knew I had to handle everything alone, which helped me mentally prepare...But when their gaze popped back on, I knew I wasn't working alone anymore."* – P13, P2.

The least preferred condition overall was "Gaze Sharing Always Off", with only 20.2% ($n = 17$) ranking it as their top choice and the majority (50%, $n = 42$) ranking it as their least preferred. Many participants who disliked this condition expressed that the lack of gaze sharing made it harder to coordinate tasks efficiently. One participant stated, *"Without gaze sharing, I had to verbally ask my teammate [about] their action constantly, which slowed us down"* – P37, P1.

Table 6.4: Participant preferences for gaze sharing conditions

Condition	**Ranked 1st (Most Preferred)**	**Ranked 2nd**	**Ranked 3rd (Least Preferred)**
On/Off Gaze Sharing Button	43 (51.2%)	27 (32.1%)	14 (16.7%)
Gaze Sharing Always On	24 (28.6%)	32 (38.1%)	28 (33.3%)
Gaze Sharing Always Off	17 (20.2%)	25 (29.8%)	42 (50.0%)

Participants also evaluated the usefulness of gaze sharing for maintaining awareness of their teammates' actions. 33.3% ($n = 28$) of participants found it very helpful, while 22.6% ($n = 19$) considered it helpful. Another 21.4% ($n = 18$) found it moderately helpful, indicating that while it had some benefits, it was not essential for all team interactions. On the other hand, 17.9% ($n = 15$) found gaze sharing only slightly helpful, and 4.8% ($n = 4$) did not find it helpful at all. These findings suggest that while many participants appreciated the additional visual awareness, others either did not find it beneficial or preferred alternative methods of communication.

When having the option to toggle gaze sharing on and off, the reported usage varied significantly. Some participants used the button frequently, while others rarely interacted with it. 20.2% ($n = 17$) of participants reported using the button very often, and 2.4% ($n = 2$) used it often. A slightly larger group, 16.7% ($n = 14$), used the button occasionally, adjusting their usage depending on the task or workload. Meanwhile, 8.3% ($n = 7$) used it rarely, and 15.5% ($n = 13$) used it very rarely. A significant portion, 36.9% ($n = 31$), never used the button at all, suggesting that for some participants, gaze sharing either did not require toggling or was unnecessary for task performance.

Participants shared various reasons for their button usage habits, revealing key themes in how and why they interacted with the toggle function. Those who used the button frequently emphasized its value in managing interruptions and adapting to workload changes. One participant explained, *"I toggled it off when I needed to concentrate but turned it back on when I wanted to re-engage with my teammate"* – P42, P2. Another participant described using it strategically, stating, *"I used it a lot during interruptions to refocus myself before jumping back into the UAV task"* – P9, P1. These responses indicate that for some users, the ability to control gaze sharing was crucial in balancing attention between their own work and their teammates' activity.

Some participants used the button selectively, enabling gaze sharing when necessary and disabling it when it became distracting. One participant shared, *"I liked that I could turn it on to confirm where my teammate was looking in the target task, then turn it off when I was working alone"* – P28, P2. Others preferred keeping gaze sharing on for most of the time rather than frequently toggling it, as expressed by one participant: *"I just left it on since I found it easier than toggling back and forth"* – P21, P1. Conversely, participants who rarely or never used the button often cited a lack of necessity or reliance on other communication methods. One participant remarked, *"I relied more on talking to my teammate rather than watching their gaze"* – P29, P1.

6.5 Discussion

The aim of this chapter was to investigate how different gaze sharing mechanisms such as a user-controlled on/off button influence UAV team performance, workload, and interruption management. The findings demonstrate that while gaze sharing provides substantial benefits, its effectiveness depends on user control and task demands. The user-controlled gaze sharing condition emerged as the most effective, balancing the advantages of shared visual information with the need to manage cognitive load. Conversely, continuous gaze sharing, while beneficial for coordination, introduced challenges such as increased workload and difficulty managing interruptions. These insights emphasize the importance of designing adaptable gaze sharing interfaces tailored to dynamic team environments.

6.5.1 Gaze Sharing On/Off Button Optimizes the Use of Visual Cues

The results indicate that gaze sharing plays a critical role in shaping team coordination and cognitive demands, yet its efficacy is highly dependent on how it is implemented. Teams utilizing the on/off toggle condition exhibited the highest performance scores and accuracy rates, reinforcing the idea that adaptive control mechanisms provide the best balance between information sharing and workload management. These performance gains can be attributed to the strategic use of gaze sharing, which likely reduced cognitive overload while still providing critical visual context to facilitate coordination.

The higher overall performance scores and decision accuracy in the user-controlled condition indicate that teams benefited from having the option to engage with gaze sharing when needed. The ability to toggle gaze sharing likely allowed participants to reduce distractions while still leveraging shared visual cues when coordination was necessary. In contrast, teams with continuous gaze sharing performed better than those without any gaze sharing, but their cognitive workload was higher than when they had access to the On/Off button, as indicated by elevated mental demand and frustration scores. This suggests that while gaze sharing facilitates team coordination, excessive visual input without control may impose additional cognitive costs.

Target detection accuracy followed a similar pattern, showing significant differences between conditions. This finding can be explained by the nature of the UAV task, where multiple targets must be identified among the 16 UAVs. Given that target detection is a shared task, gaze sharing played a crucial role in allowing teammates to coordinate their efforts effectively. The qualitative responses support this, as participants indicated that being able to see their partner's gaze helped them allocate attention more efficiently. When gaze sharing was always on or available via the toggle, participants could quickly assess which targets their teammate was already attending to and shift their focus to other UAVs. This prevented redundant efforts and improved overall team efficiency, leading to the observed increase in target detection accuracy. The no-gaze sharing condition, in contrast, required verbal communication or independent assessments, which introduced delays and potential misallocations of attention.

The eye tracking metrics provide additional insight into how gaze sharing influenced cognitive and visual processing. The number of fixations and saccades was significantly lower in the gaze sharing conditions compared to no gaze sharing, indicating that participants required fewer visual search efforts to maintain awareness. This aligns with the idea that gaze sharing provides a direct channel for information exchange, reducing the need to scan the environment extensively. However, the continuous gaze sharing condition did not always lead to the lowest fixation and saccade counts, suggesting that uncontrolled gaze sharing may still demand attentional resources to filter relevant from irrelevant information.

Saccade velocity and pupil diameter were highest in the "No Gaze Sharing" condition, further reinforcing the notion that participants experienced greater cognitive strain when they could not rely on gaze cues for coordination. Elevated pupil diameter is a well-established indicator of increased cognitive workload, suggesting that without gaze sharing, participants had to expend more effort to track their partner's actions and distribute their own attention effectively. Saccade duration was also significantly longer in the "No Gaze Sharing" condition, which may indicate that participants engaged in more prolonged search behavior to compensate for the lack of shared visual references.

6.5.2 Gaze Sharing and Interruption Management in Simple vs. Complex Tasks

Interruptions introduced another layer of complexity to these dynamics (Abubakar et al., 2023; Aitken et al., 2021). Resumption lag times were shortest in the user-controlled condition, suggesting that participants could strategically manage their visual attention to facilitate faster recovery after task-switching. The continuous gaze sharing condition also led to faster resumption compared to no gaze sharing, further reinforcing the role of shared visual cues in maintaining awareness. However, decision accuracy after interruptions followed a different pattern—while gaze sharing consistently improved accuracy over nothing, the benefit was particularly pronounced in complex tasks when participants could control when and how to engage with gaze sharing.

For shared tasks, there was no significant difference in decision accuracy between the continuous and user-controlled gaze sharing conditions, suggesting that both facilitated post-interruption performance equally. However, for individual tasks, a distinction emerged based on task complexity. This differentiation between simple and complex tasks offers key insights into the mechanisms underlying gaze sharing effectiveness. In simple tasks, decision accuracy did not significantly differ between continuous and user-controlled gaze sharing, reinforcing the idea that gaze sharing may be less critical in low-demand contexts. In contrast, for complex tasks, user-controlled gaze sharing significantly outperformed continuous gaze sharing, indicating that the ability to manage visual information is particularly beneficial when cognitive demands are high. This explains why the on/off button condition consistently led to the highest accuracy rates—participants could mitigate unnecessary distractions while still using gaze sharing as a support mechanism when needed.

These patterns align with the differences observed in resumption lag. For simple tasks, user control over gaze sharing did not yield a significant advantage, suggesting that participants could recover effectively regardless of how gaze sharing was implemented. However, for complex tasks, resumption lag was significantly lower in the user-controlled condition, confirming that participants were able to more efficiently re-engage with the task when they had control over gaze sharing. This advantage likely stems from the ability to toggle gaze sharing on and off based on immediate task needs, allowing participants to avoid unnecessary distractions while still leveraging shared gaze information when needed.

Eye tracking metrics further support these findings. Time to first fixation after an interruption was lowest in the user-controlled gaze sharing condition, followed by continuous gaze sharing, with the "No Gaze Sharing" condition resulting in the longest delays. This indicates that gaze sharing helped participants quickly regain awareness after an interruption, with user-controlled access providing the greatest advantage. These results are particularly relevant when considering rerouting accuracy, a complex task that also exhibited significant differences between conditions. The alignment between shorter time to first fixation, faster resumption, and improved decision accuracy in complex tasks suggests that user-controlled gaze sharing plays a crucial role in helping teams manage cognitive transitions efficiently.

These findings also align with participant preferences. The majority of participants favored the user-controlled gaze sharing condition, citing its flexibility as a key factor in their ability to manage attention effectively. While some participants found continuous gaze sharing helpful, others found it overwhelming, which helps explain why continuous gaze sharing, despite offering benefits over no gaze sharing, did not always yield optimal performance. The fact that some participants rarely used the toggle button while others used it frequently suggests that gaze sharing needs vary depending on individual strategies and task demands. This highlights the importance of designing adaptable systems that accommodate diverse user needs rather than enforcing a one-size-fits-all approach.

The interaction between interruptions and gaze sharing also underscores the role of adaptability in complex, high-stakes environments. When participants had control over gaze sharing, they could toggle it off during high-cognitive-load moments and re-engage when co-ordination was necessary. This likely contributed to the faster resumption times and higher decision accuracy seen in the user-controlled condition. In contrast, when gaze sharing was always on, the lack of control may have contributed to visual overload, making it harder to regain focus after an interruption. The qualitative responses further support this interpretation, as participants emphasized that they used the toggle strategically to manage attention and workload.

Additionally, uninterrupted participants generally found gaze sharing to be useful as it provided a clear, implicit cue of their teammate's presence. When their teammate was interrupted, the disappearance of their gaze served as a signal that they were responsible for managing all tasks. Upon their teammate's return, the reappearance of gaze information offered reassurance and facilitated a seamless transition back into collaborative work. This suggests that gaze sharing was not only a tool for coordination but also an important indicator of workload distribution and teammate availability.

However, it is important to note that in this study, participants were co-located, meaning they could hear and see their teammate being interrupted. In many cases, teammates verbally acknowledged when they were leaving for an interrupting task, which helped maintain awareness of each other's availability. This raises important considerations for future work in distributed teams, where communication is more constrained, and teammates may not have direct auditory or visual confirmation of interruptions (Fischer & Mosier, 2014; Mosier & Fischer, 2021). In such settings, communication delays could create uncertainty regarding a partner's availability, potentially impacting coordination and workload management (Bernier-Vega et al., 2023; Bulfone et al., 2020). Future work should explore how gaze sharing can be adapted to remote, distributed environments, where explicit indicators of a teammate's status—such as gaze persistence or presence indicators—may help mitigate the challenges of communication delays and uncertainty in shared tasks.

Overall, these results indicate that gaze sharing is not inherently beneficial or detrimental, but rather that its effectiveness is highly dependent on how it is implemented. Together, these findings reinforce the idea that while gaze sharing generally enhances coordination, its greatest benefits emerge in high-demand scenarios where operators must balance multiple cognitive demands. The ability to control gaze sharing provides a strategic advantage, allowing users to engage with visual information selectively rather than being subjected to constant visual input that may not always be relevant. The user-controlled condition emerged as the optimal solution because it allowed participants to tailor their use of gaze sharing to their specific needs and task demands.

6.6 Limitations and Future Work

While this study provides strong evidence of the benefits of gaze sharing in UAV teams, several aspects warrant further investigation. First, although participants received extensive training on the UAV task, they were not professional UAV operators. Prior research suggests that expertise influences how individuals interact with technology, including strategies for managing attention and workload. Experienced UAV operators may develop more efficient methods for utilizing gaze sharing or may already rely on well-established coordination techniques that reduce the need for visual sharing. Future studies should examine how expertise level impacts gaze sharing effectiveness and whether more experienced teams adapt differently to its availability.

Additionally, while team composition was controlled to include one male and one female in each pair who were not acquainted with each other, research on teamwork and diversity suggests that team dynamics, familiarity, and prior experience working together all influence coordination strategies. Gender composition may affect how teammates communicate and allocate attention, and pre-existing familiarity between teammates could impact their reliance on gaze sharing. While this study did not systematically examine these factors, future work should explore how team diversity and prior collaboration influence gaze sharing utility. A broader understanding of how gaze sharing interacts with team composition variables could inform tailored implementations that maximize its benefits across different types of teams.

Another limitation concerns the design of interruptions. In this study, interruptions were fixed at 30 seconds, allowing for controlled comparisons, but in real-world UAV operations, interruptions are highly variable. Some may be momentary, requiring only a quick glance away, while others may be prolonged, forcing operators to reconstruct awareness from scratch. The impact of gaze sharing on task resumption might vary across different types and durations of interruptions, particularly in high-risk environments where re-engagement speed is critical. Future research should investigate how gaze sharing supports task resumption when interruptions range from brief distractions to extended diversions, as well as whether adaptive gaze sharing techniques could assist operators in regaining awareness more efficiently.

Beyond UAV command-and-control tasks, future research should also explore the applicability of gaze sharing in other domains that require high levels of coordination, such as air traffic control, remote surgery, or collaborative robotics. Each of these fields involves complex decision-making in team settings where nonverbal communication and shared awareness are crucial. Understanding how gaze sharing can be adapted to these domains could expand its impact beyond UAV operations.

Furthermore, this study examined a specific set of gaze sharing conditions, but many alternative techniques remain unexplored. Future work should investigate how different visualizations of gaze, such as augmented reality overlays, adaptive gaze sharing filters that emphasize relevant information, or multimodal integrations with haptic and auditory feedback, could enhance collaboration. The potential for AI-driven gaze sharing interfaces that dynamically adjust based on workload or task demands also presents a promising avenue for improving human-machine teaming.

By addressing these areas in future research, gaze sharing technologies can be further refined to enhance team coordination across a broader range of settings, ensuring that their implementation is both flexible and beneficial under real-world operational constraints.

6.7 Conclusion

This study investigated the role of gaze sharing in UAV team coordination, examining its impact on task performance, workload, and interruption management. The findings demonstrate that gaze sharing significantly enhances team efficiency, particularly when users have control over when and how to engage with shared visual information. Across all performance metrics, the user-controlled gaze sharing condition led to the highest task accuracy and lowest cognitive load, emphasizing the importance of adaptability in interface design. The results also highlight that while gaze sharing generally improves coordination, its benefits are most pronounced in complex tasks that require dynamic attention management. The ability to toggle gaze sharing allowed participants to minimize distractions while leveraging gaze information when necessary, leading to improved task performance and faster recovery from interruptions. Furthermore, eye tracking measures suggest that gaze sharing reduces the need for extensive visual search, optimizing cognitive resources and enabling more efficient decision-making.

However, the effectiveness of gaze sharing is not solely dependent on its presence; rather, its design and implementation play a crucial role in determining its impact. Continuous gaze sharing, while beneficial in some contexts, introduced additional cognitive load for some participants, reinforcing the need for flexible and user-controlled gaze sharing interfaces. These insights underscore the necessity of designing gaze sharing systems that accommodate different task demands, user preferences, and operational constraints.

By integrating these findings into future system designs, gaze sharing can be further optimized to support collaborative work in high-stakes environments. Future research should explore its applicability beyond UAV operations, investigate alternative gaze sharing techniques, and consider adaptive mechanisms that dynamically adjust based on workload and task complexity. As human-machine teaming continues to evolve, ensuring that gaze sharing technology is intuitive, adaptable, and effective will be essential for enhancing team performance in complex operational settings.

6.8 References

Abubakar, A. I., Ahmad, I., Omeke, K. G., Ozturk, M., Ozturk, C., Abdel-Salam, A. M., & Imran, M. A. (2023). Survey Energy Optimization Techniques UAV-Based Cellular Networks: From Conventional to Machine Learning Approaches. *Drones*, *7*(3), 214. https://doi.org/10.3390/drones7030214

Aitken, J., Esain, A. E., & Williams, S. (2021). Management Complexity Care Ecosystem. *Supply Chain Management*, *26*(4), 481–494. https://doi.org/10.1108/SCM-05-2020-0207

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2014). Momentary Interruptions Can Derail the Train of Thought. *Journal of Experimental Psychology: General*, *143*(1), 215. https://doi.org/10.1037/a0030986

Atweh, J. A., El Iskandarani, M., & Riggs, S. L. (2024). Real-Time Gaze Sharing Techniques and Their Influence on Performance and Shared Situational Awareness of Teammates in UAV C2 Operations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *68*(1), 146–154. https://doi.org/10.1177/10711813241260316

Atweh, J. A., Hazimeh, R., & Riggs, S. L. (2023). Can Real-Time Gaze Sharing Help Team Collaboration? A Preliminary Examination of its Effectiveness with Pairs. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *67*(1), 716–721. https://doi.org/10.1177/21695067231193659

Atweh, J. A., & Riggs, S. L. (2024). Gaze Sharing, a Double-Edged Sword: Examining the Effect of Real-Time Gaze Sharing Visualizations on Team Performance and Situation Awareness. *Human Factors*, *67*(3), 196–224. https://doi.org/10.1177/00187208241272060

Bernier-Vega, A., Barton, K., Olson, I., Rodriguez, J., Cantu, G., & Ozcelik, S. (2023). Remote Data Acquisition Using UAVs and Custom Sensor Node Technology. *Drones*, *7*(6), 340. https://doi.org/10.3390/drones7060340

Bolton, M. L., Biltekoff, E., & Humphrey, L. (2023). The Mathematical Meaninglessness of the NASA Task Load Index: A Level of Measurement Analysis. *IEEE Transactions on Human-Machine Systems*, *53*(3), 590–599. https://doi.org/10.1109/THMS.2023.3263482

Bozza, A., Asprone, D., & Manfredi, G. (2017). A Methodological Framework Assessing Disaster Resilience of City Ecosystems to Enhance Resource Use Efficiency. *International Journal of Urban Sustainable Development*, *9*(2), 136–150. https://doi.org/10.1080/19463138.2016.1244538

Brust, M. R., Danoy, G., Stolfi, D. H., & Bouvry, P. (2021). Swarm-Based Counter UAV Defense System. *Discover Internet of Things*, *1*, 1–19. https://doi.org/10.1007/s43926-021-00002-x

Bulfone, A., Drioli, C., Ferrin, G., & Foresti, G. L. (2020). A Scalable System for the Monitoring of Video Transmission Components in Delay-Sensitive Networked Applications. *Multimedia Tools and Applications*, *79*, 18727–18745. https://doi.org/10.1007/s11042-020-08743-7

Chen, J. Y., & Barnes, M. J. (2014). Human–Agent Teaming for Multirobot Control: A Review of Human Factors Issues. *IEEE Transactions on Human-Machine Systems*, *44*(1), 13–29. https://doi.org/10.1109/THMS.2013.2293535

de Coning, C. (2016). From Peacebuilding to Sustaining Peace: Implications of Complexity for Resilience and Sustainability. *Resilience*, *4*(3), 166–181. https://doi.org/10.1080/21693293.2016.1153773

Dias, R. D., Conboy, H. M., Gabany, J. M., Clarke, L. A., Osterweil, L. J., Arney, D., Goldman, J. M., Riccardi, G., Avrunin, G. S., Yule, S. J., & Zenati, M. A. (2018). Intelligent Interruption Management System to Enhance Safety and Performance in Complex Surgical and Robotic Procedures. *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis: First International Workshop*. https://doi.org/10.1007/978-3-030-01201-4_8

Donmez, B., Nehme, C., & Cummings, M. L. (2010). Modeling Workload Impact in Multiple Unmanned Vehicle Supervisory Control. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *40*(6), 1180–1190. https://doi.org/10.1109/TSMCA.2010.2046731

Feitshans, G., Rowe, A., Davis, J., Holland, M., & Berger, L. (2008). Vigilant Spirit Control Station (VSCS): The Face of COUNTER. *AIAA Guidance, Navigation and Control Conference and Exhibit*. https://doi.org/10.2514/6.2008-6309

Fischer, U., & Mosier, K. (2014). Impact Communication Delay and Medium Team Performance and Communication Distributed Teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 115–119. https://doi.org/10.1177/1541931214581025

Grundgeiger, T., & Sanderson, P. (2009). Interruptions in Healthcare: Theoretical Views. *International Journal of Medical Informatics*, *78*(5), 293–307. https://doi.org/10.1016/j.ijmedinf.2008.10.001

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. https://doi.org/10.1016/S0166-4115(08)62386-9

Katidioti, I., Borst, J. P., Bierens de Haan, D. J., Pepping, T., van Vugt, M. K., & Taatgen, N. A. (2016). Interrupted by Your Pupil: An Interruption Management System Based on Pupil Dilation. *International Journal of Human–Computer Interaction*, *32*(10), 791–801. https://doi.org/10.1080/10447318.2016.1198525

Khan, A., Gupta, S., & Gupta, S. K. (2022). Emerging UAV Technology for Disaster Detection, Mitigation, Response, and Preparedness. *Journal of Field Robotics*, *39*(6), 905–955. https://doi.org/10.1002/rob.22075

Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A. (2008). The Effect of Interruption Duration and Demand on Resuming Suspended Goals. *Journal of Experimental Psychology: Applied*, *14*(4), 299. https://doi.org/10.1037/a0014402

Mosier, K. L., & Fischer, U. M. (2021). Meeting Challenge Transmission Delay: Communication Protocols Space Operations. *Human Factors*, *65*(6), 1235–1250. https://doi.org/10.1177/00187208211047085

Newn, J., Velloso, E., Allison, F., Abdelrahman, Y., & Vetere, F. (2017). Evaluating Real-Time Gaze Representations to Infer Intentions in Competitive Turn-Based Strategy Games. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 541–552. https://doi.org/10.1145/3116595.3116624

Sasangohar, F., Scott, S. D., & Cummings, M. L. (2014). Supervisory-Level Interruption Recovery in Time-Critical Control Tasks. *Applied Ergonomics*, *45*(4), 1148–1156. https://doi.org/10.1016/j.apergo.2014.02.005

Scott, S. D., Wan, J., Sasangohar, F., & Cummings, M. L. (2008). Mitigating Supervisory-Level Interruptions in Mission Control Operations. *Proceedings of 2nd International Conference on Applied Human Factors and Ergonomics*, 14–17.

Trafton, J. G., Altmann, E. M., & Brock, D. P. (2005). Huh, What Was I Doing? How People Use Environmental Cues After An Interruption. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(3), 468–472. https://doi.org/10.1177/154193120504900354

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to Resume an Interrupted Task: Effects of Prospective Goal Encoding and Retrospective Rehearsal. *International Journal of Human-Computer Studies*, *58*(5), 583–603. https://doi.org/10.1016/S1071-5819(03)00023-5

Werner, N. E., & Holden, R. J. (2015). Interruptions in the Wild: Development of a Sociotechnical Systems Model of Interruptions in the Emergency Department Through a Systematic Review. *Applied Ergonomics*, *51*, 244–254. https://doi.org/10.1016/j.apergo.2015.05.010

Wiesche, M. (2021). Interruptions in Agile Software Development Teams. *Project Management Journal*, *52*(2), 210–222. https://doi.org/10.1177/8756972821991365

Zhang, Y., Pfeuffer, K., Chong, M. K., Alexander, J., Bulling, A., & Gellersen, H. (2017). Look Together: Using Gaze for Assisting Co-Located Collaborative Search. *Personal and Ubiquitous Computing*, *21*, 173–186. https://doi.org/10.1007/s00779-016-0969-x

Zhou, K., Wang, C., Xie, S., Zhou, Y., Zhang, X., Wang, Y., & Tang, H. (2024). Effect of Task Interruption on the Situation Awareness of Air Traffic Controllers. *PloS One*, *19*(11), e0314183. https://doi.org/10.1371/journal.pone.0314183

Chapter 7

Conclusion

The study of teams in complex systems is critical to addressing the challenges faced in high-stakes environments, such as unmanned aerial vehicle (UAV) command-and-control (C2) operations. These systems require seamless collaboration, real-time decision-making, and the ability to adapt dynamically to changing conditions (Ateş et al., 2022; Harinarayana et al., 2024; Hildmann & Kovacs, 2019). However, traditional approaches to understanding and optimizing team performance have often focused on post-hoc evaluations or individual-centric metrics, leaving gaps in our ability to assess and support teams in real-time. This dissertation addresses these challenges by exploring how cognitive engineering principles and real-time data, particularly eye tracking metrics, can enhance our understanding of team dynamics and provide actionable insights for system design.

Chapter 1 motivates the need for this research by presenting the complexities of team collaboration in UAV C2 environments and the limitations of existing methods for quantifying team performance. It emphasizes the importance of considering cognitive engineering approaches tailored to teams, rather than individuals, and highlights the potential of real-time metrics to provide a deeper understanding of team interactions. The chapter establishes the foundation for this work, advocating for a shift towards designing tools and systems that actively support team performance in real-time, operationally relevant contexts.

Chapter 2 builds on this foundation by reviewing the state of UAV C2 research and identifying key factors that influence team performance in complex systems. This review spans individual, team, and organizational levels, illustrating how these factors interact to shape outcomes in high-pressure environments. Importantly, this chapter underscores the need for real-time metrics to quantify team performance and introduces eye tracking as a promising tool to address this gap. It highlights the role of eye tracking in capturing individual and team-level factors, such as situation awareness (SA) and coordination, offering a pathway to better understand and optimize team dynamics in these critical contexts.

In summary, the aims of this dissertation were threefold, each addressing a critical aspect of understanding and enhancing team performance in complex systems.

- **Aim 1:** Understand how eye tracking can be leveraged to quantify team collaboration and identify team performance breakdowns in UAV C2 tasks.
  - RQ 1.1: How do scanpath similarity metrics (e.g., ScanMatch, MultiMatch, Multidimensional Cross-Recurrence Quantification Analysis) change as workload increases in UAV C2 tasks?
  - RQ 1.2: How do scanpath similarity metrics correlate with team performance measures (e.g., team score, response time) across different workload conditions, and can these correlations help identify performance breakdowns?
- **Aim 2:** Investigate how different gaze sharing displays (dot, trail, no gaze sharing) influence team performance, workload, and communication dynamics in complex systems.
  - RQ 2.1: How does gaze sharing influence team collaboration in more complex

systems, such as UAV C2 operations?

  - RQ 2.2: How do different gaze sharing visualization techniques (dot, trail, no gaze sharing) affect team scanning techniques, situation awareness, workload, and performance?
  - RQ 2.3: How do verbal and non-verbal communication techniques (e.g., gaze sharing) interact in UAV C2 teams, and under what circumstances do teams perceive one technique as a replacement for or a complement to another?

- **Aim 3:** Examine how user-controlled gaze sharing (via an on/off toggle) influences team collaboration and performance in UAV C2 operations, particularly in the context of frequent interruptions and varying task complexity.

  - RQ 3.1: How does user-controlled gaze sharing affect team performance compared to continuous gaze sharing and no gaze sharing displays?
  - RQ 3.2: How does gaze sharing influence teams' ability to recover from interruptions, and does its effect differ based on task complexity (simple vs. complex tasks)?

Together, these aims represent a cohesive effort to bridge gaps in the understanding of team collaboration within UAV C2 systems. The findings contribute to cognitive systems engineering by validating novel metrics, introducing innovative tools, and providing evidence-based recommendations for system design. This work not only advances the theoretical understanding of team dynamics but also offers practical solutions for improving performance, resilience, and coordination in high-stakes, technology-mediated environments.

7.1 Intellectual Merit

This dissertation addresses a specific, unexplored niche in the knowledge base: the study of team dynamics and performance in complex systems, specifically within the context of UAV operations, through the innovative application of eye tracking technologies. By integrating human factors principles with advanced analysis techniques, this research fills a critical gap in understanding how teams interact, adapt, and maintain SA under varying workload conditions.

Specifically, the present dissertation builds upon the following aspects of cognitive systems engineering and team collaboration research:

- The introduction and validation of MdCRQA and scanpath similarity metrics (like MultiMatch) as sensitive, real-time indicators of workload changes and performance.
- The exploration of gaze sharing visualizations, including the trail and dot formats, and their impact on task execution, adaptability, and teammate interactions.
- The application of real-time gaze sharing in high-stakes team environments, providing novel insights into its utility, limitations, and effects on team coordination and communication.
- The examination of interruption management strategies facilitated by gaze sharing on/off toggles, contributing to the sparse body of literature on interruption management in complex domains.

This dissertation also lays the groundwork for human factors professionals to pursue different analysis techniques for their multifaceted, operationally relevant research questions. The integration of real-time, unobtrusive eye tracking metrics into team-based studies offers a framework that can be extended to various domains, advancing both the theoretical understanding and practical applications of team performance optimization in complex systems.

7.1.1 Advancements in Team Collaboration Metrics

One of the primary contributions of this dissertation is the development and validation of innovative metrics for quantifying team collaboration in complex systems (Steitz et al., 2020) . This work highlights the potential of eye tracking technologies, particularly through MdCRQA and scanpath similarity metrics such as MultiMatch, to serve as robust tools for assessing team dynamics and performance.

Chapter 3 introduces the foundation for this work by employing these metrics to capture and analyze how UAV teams adapt their gaze behaviors under varying workload conditions. The results revealed that MdCRQA and scanpath similarity metrics are highly sensitive to changes in workload, offering a detailed understanding of team synchronization and coordination. These metrics not only correlate strongly with performance outcomes, particularly under high workload scenarios, but also provide a real-time, unobtrusive means of detecting potential performance breakdowns.

Subsequent chapters build upon this foundation. Chapters 4, 5, and 6 further contextualize the application of these metrics within the realm of gaze sharing strategies. Chapter 4 stands out as it pioneers the use of real-time gaze sharing in UAV team environments, showcasing how these metrics can evaluate the effectiveness of various visualization techniques, such as the trail and dot formats. The trail format, in particular, emerged as superior, as evidenced by its ability to facilitate better team coordination and adaptability in shared and separate task settings.

Overall, the contributions of this dissertation demonstrate the transformative potential of scanpath similarity metrics and gaze sharing in advancing our understanding of team collaboration. These metrics offer actionable insights into designing adaptive systems and interfaces that enhance team efficiency, SA, and resilience, particularly in complex, high-pressure domains like UAV operations.

7.1.2 Introducing Real-Time Gaze Sharing to Complex Systems

This dissertation makes significant strides in advancing the use of real-time gaze sharing technologies within the context of UAV team operations, offering both theoretical insights and practical applications. By exploring gaze sharing as a tool for improving team coordination and performance, this work provides a nuanced understanding of how gaze sharing visualizations can enhance collaborative efforts in high-stakes environments.

Chapter 4 is a cornerstone of this effort, as it marks one of the first studies to systematically assess gaze sharing visualizations in complex tasks (Atweh & Riggs, 2024). This chapter evaluates two distinct visualization methods—the dot and the trail—and demonstrates the superiority of the trail format in fostering effective team coordination. The findings reveal that the trail's ability to convey temporal gaze patterns significantly enhances its utility in both shared and separate task scenarios. Moreover, the study highlights how teammates interact with these visualizations to adapt their strategies and improve task performance.

Chapter 5 builds upon this foundation by investigating the role of gaze sharing in communication strategies. This chapter highlights the dual role of gaze sharing: as a substitute for verbal communication in tasks requiring shared target detection and as a complement to verbal communication in scenarios where chat messages were employed. These findings underscore the contextual nature of gaze sharing and its ability to support task-specific communication needs.

Chapter 6 extends the exploration of gaze sharing by examining its application in interruption management. A novel on/off gaze sharing toggle is introduced and assessed, revealing its potential to aid teams in mitigating the effects of interruptions. The chapter illustrates how teammates leverage gaze sharing tools to recover their SA and coordination more efficiently, adding a new dimension to the understanding of interruption recovery in team settings.

Together, these chapters demonstrate the transformative potential of real-time gaze sharing in enhancing team collaboration. By providing actionable insights into visualization design, communication dynamics, and interruption recovery, this dissertation establishes a foundation for integrating gaze sharing technologies into operational systems to optimize performance and resilience.

7.1.3 Communication Dynamics in Collaborative Systems

This dissertation provides a deeper understanding of communication dynamics in collaborative systems by examining how teams adapt their communication strategies in response to task demands, workload conditions, and the availability of shared tools such as gaze sharing visualizations. By investigating the interplay between verbal and non-verbal communication, this work uncovers how UAV teams coordinate and perform effectively in complex environments.

Chapter 5 is central to this exploration, as it delves into the role of gaze sharing tools in shaping team communication. In tasks such as shared target detection, gaze sharing served as a substitute for verbal communication, reducing the need for spoken exchanges while maintaining team performance. Conversely, in scenarios involving chat-based interactions, gaze sharing complemented verbal communication, offering an additional layer of SA that enhanced task execution. These findings illustrate the adaptability of communication strategies based on task requirements and available resources. Chapter 6 further extends this analysis by exploring how teams leverage communication dynamics during interruptions. The on/off gaze sharing toggle introduced in this chapter provided a unique opportunity to study how non-verbal cues can facilitate recovery from disruptions. Teams used the toggle to re-engage quicker in the ongoing UAV tasks, highlighting the critical role of both verbal and non-verbal communication in maintaining performance under challenging conditions.

The insights gained from these chapters underscore the importance of designing tools and systems that support flexible communication strategies. By integrating gaze sharing visualizations and other collaborative tools into team workflows, this research contributes to the development of systems that enhance communication, coordination, and overall team performance in high-stakes, complex domains.

7.1.4 Interruption Management in High-Stakes Environments

This dissertation makes a unique contribution to the underexplored area of interruption management in high-stakes team environments, particularly in the context of UAV operations. By examining how gaze sharing tools and strategies can mitigate the disruptive effects of interruptions, this work provides novel insights into team resilience, recovery, and task performance continuity.

Chapter 6 is pivotal in this regard, introducing an innovative on/off gaze sharing toggle designed to help teams re-engage following interruptions. This feature was meticulously designed and tested to provide teams with a flexible mechanism to manage interruptions proactively. The toggle allowed team members to selectively enable or disable gaze sharing, facilitating a more tailored approach to navigating disruptions based on task complexity and team roles. The findings from this study demonstrate that such toggles can serve as a critical tool for teams operating under high cognitive and temporal demands, offering a structured means of managing disruptions and minimizing performance degradation.

Additionally, this chapter explores the nuanced behavioral strategies that teams adopt in response to interruptions. Through detailed analysis, it identifies patterns of gaze behavior that facilitate rapid recovery and coordination, highlighting the interplay between individual and collective efforts. These strategies reveal how teams can dynamically recalibrate their focus and reestablish effective collaboration, even under challenging circumstances. For example, the research shows how certain roles within the team take the lead in managing disruptions, providing insights into task delegation and role fluidity during recovery phases.

The contributions of this work extend well beyond the UAV domain. The principles and findings have broader implications for other complex, interruption-prone environments such as healthcare, where surgical teams frequently encounter unexpected events; air traffic control, where operators manage multiple high-stakes tasks simultaneously; and emergency response teams, where rapid recovery and coordination are paramount. By examining the mechanisms underlying successful recovery from interruptions, this research offers a framework for designing adaptive systems and collaborative tools that enhance team performance and resilience in critical operations.

Moreover, the dissertation provides actionable insights for system designers and human factors professionals aiming to mitigate the impact of interruptions in team settings. Recommendations include the incorporation of flexible gaze sharing tools, the development of training protocols to improve interruption management strategies, and the design of interfaces that support both individual and collective awareness recovery. By focusing on these practical applications, the research bridges the gap between theoretical insights and real-world operational needs, laying the groundwork for future innovations in collaborative systems.

7.2 Broader Impact

The research presented in this dissertation offers significant contributions not only to the academic study of team dynamics and human-computer interaction but also to the practical challenges faced by teams operating in complex, high-stakes environments. By advancing our understanding of how teams interact, adapt, and perform under varying conditions, this work provides a foundation for innovations that can enhance safety, efficiency, and collaboration across diverse domains.

At its core, this research bridges the gap between theoretical exploration and real-world application. It introduces novel metrics such as MdCRQA and scanpath similarity, demonstrating their utility in capturing the intricacies of team behavior and performance. These tools enable researchers and practitioners to quantify team dynamics in ways that were previously unattainable, setting the stage for a new era of data-driven insights into collaboration and workload management.

The implications of this work extend beyond the immediate domain of UAV operations. By exploring the transformative potential of gaze sharing and adaptive displays, this dissertation highlights how cutting-edge technologies can reshape team-based systems. From improving communication strategies to designing tools that support dynamic task allocation, these findings have broad relevance for industries such as aviation, air traffic control, and emergency response, where effective teamwork is critical to success (Hofmaenner et al., 2020; Malakis & Kontogiannis, 2023; Waterson et al., 2015). Moreover, this research emphasizes the need to move beyond individual-focused approaches to human-computer interaction, instead advocating for systems that account for the complex interplay of team members in dynamic environments. The introduction of adaptive displays for teams, supported by assistive AI, represents a pivotal step in this direction, offering a vision for the future of collaborative technologies that are responsive, proactive, and team-aware (Atweh et al., 2022;

Giotopoulos et al., 2024).

Finally, the broader societal contributions of this work underscore its significance. As we face increasingly interconnected and technologically complex challenges, the ability to optimize team performance and resilience has never been more crucial. The research in this dissertation provides actionable insights that can enhance not only the safety and efficiency of critical operations but also the broader pursuit of innovation and collaboration in the modern world (Hirshfield et al., 2023). The following sections delve into the specific contributions of this dissertation, exploring its advancements in cognitive engineering, its potential to enhance team performance, its foundation for adaptive displays, and its broader societal implications.

7.2.1 Advancing Cognitive Engineering Research

This dissertation significantly advances the fields of eye tracking research and cognitive systems engineering by demonstrating how innovative metrics and methodologies can deepen our understanding of team collaboration, workload, and performance in complex systems (Coakes et al., 2008; Dietz et al., 2017). By integrating real-time eye tracking technologies with principles of cognitive engineering, this research not only validates existing theories but also challenges and extends them in the context of teams using novel technologies like gaze sharing. The introduction of MdCRQA and scanpath similarity metrics, such as MultiMatch, represents a critical advancement in how we study team dynamics. Chapter 3 lays the foundation by demonstrating that these metrics are sensitive to changes in workload and correlate strongly with performance outcomes. For example, MdCRQA captures synchronization in gaze behaviors, revealing how team members coordinate attention under both low and high workload conditions. These findings validate the use of eye tracking as a real-time, unobtrusive tool for understanding team interactions, setting the stage for more dynamic and team-centered applications of cognitive engineering principles.

Traditional cognitive engineering often focuses on individual workload, SA, and performance metrics. However, this dissertation pushes these boundaries by examining how these principles adapt when applied to teams using novel technologies (Deacon, 2020; Golden et al., 2018; Hagemann et al., 2012; Singh, 2024). For example, as seen in Chapter 4, slower and lower saccadic activity in individuals is generally associated with higher workload because individuals scan less and focus more intently under stress. However, in the context of teams using gaze sharing, the opposite trend was observed: slower and lower saccadic activity was associated with lower workload, as gaze sharing enabled team members to scan less and rely on shared visual cues for SA (Atweh & Riggs, 2024). This finding fundamentally shifts our understanding of workload indicators in team contexts and underscores the need to adapt traditional cognitive engineering metrics to account for new technologies and collaborative dynamics. It also highlights the importance of examining team-level phenomena rather than simply aggregating individual-level data, as the introduction of tools like gaze sharing fundamentally changes how workload is distributed and managed (Škvareková et al., 2020; Tsai et al., 2007; Vesper et al., 2016).

By integrating this dissertation's findings, this dissertation lays the groundwork for a new generation of cognitive systems engineering research. It highlights the importance of studying teams as dynamic systems where workload, attention, and SA are distributed across individuals and mediated by technology. The novel insights into gaze sharing and its impact on team dynamics demonstrate how new technologies can fundamentally reshape traditional cognitive engineering principles.

7.2.2 Enhancing Team Performance in Complex Systems

This dissertation offers significant contributions to understanding and enhancing team performance in complex systems, with a primary focus on UAV operations. By employing innovative methodologies and leveraging advanced eye tracking metrics, it provides actionable insights into improving SA, task coordination, and overall team resilience. Furthermore, the principles and methods developed in this research extend beyond UAV teams, offering broader applications in other high-stakes domains such as healthcare, air traffic control, and emergency response.

The core focus of this work lies in UAV C2 operations, where teams must collaborate effectively under varying workload conditions. By integrating tools like gaze sharing and metrics such as scanpath similarity, this research demonstrates how UAV teams can optimize their performance in dynamic and high-pressure environments.

Chapter 3 introduced scanpath similarity, a powerful metric for quantifying how team members align their visual attention during tasks. Using methods such as MultiMatch and MdCRQA, this chapter demonstrated that scanpath similarity is sensitive to changes in workload, offering real-time insights into team dynamics. As shown in Chapters 4 and 5, gaze sharing enables UAV teams to reduce redundant scanning and anticipate each other's actions more effectively. The trail visualization, in particular, supports shared understanding and allows teammates to coordinate seamlessly, even when tasks are distributed across the team. Moreover, Chapter 6 highlights how tools like the on/off gaze sharing toggle allow teams to recover quickly after interruptions, ensuring continuity in operations. These findings are directly applicable to UAV system design and team training, offering a foundation for developing tools and protocols that enhance collaboration, reduce cognitive load, and improve performance outcomes.

While this dissertation is rooted in UAV team studies, the methodologies and findings have broader implications for other complex domains. In healthcare, for example, eye tracking metrics such as scanpath similarity could be used to analyze and improve team coordination in operating rooms, where surgeons, nurses, and anesthesiologists must maintain precise SA. While gaze sharing may not be directly applicable due to privacy concerns or task specificity, the insights into interruption recovery and task handoffs could inform protocols for managing surgical teams under high cognitive demands. Moreover, in emergency operations or air traffic control centers, where teams must quickly process and act on dynamic information, gaze sharing tools could complement verbal communication to enhance coordination. The findings from this research could inform the design of tools that enable responders to share SA rapidly in time-critical scenarios.

By combining innovative tools, metrics, and methodologies, this dissertation lays the foundation for a new generation of research on team performance in complex systems. It highlights the importance of adapting tools and protocols to the unique demands of different domains while maintaining a rigorous and methodologically sound approach. Whether improving UAV operations or enhancing teamwork in high-stakes healthcare and emergency response scenarios, the principles and findings from this research offer actionable insights for optimizing team performance across a wide range of critical environments.

7.2.3 Laying the Foundation for Adaptive Displays for Teams

Designing adaptive displays for teams is an ambitious and largely uncharted challenge. While considerable work has been done to develop adaptive systems for individual operators, extending this approach to teams introduces a new level of complexity (Avvenuti & Vecchio, 2009; Chen & Kanfer, 2024; Mangaroska et al., 2022; Papamitsiou et al., 2020). Teams are dynamic systems where workload, attention, and performance are distributed across individuals, making it crucial to develop metrics and methodologies that account for these collective dynamics. This dissertation lays the groundwork for such systems by introducing innovative metrics and exploring how they can serve as the foundation for real-time adaptive displays.

One of the major contributions of this dissertation is the introduction and validation of metrics like MdCRQA and scanpath similarity as tools for assessing team states. Unlike traditional metrics that focus on individual workload or performance, these tools capture the interactions and synchronization between team members, providing a holistic view of team dynamics. Building on these metrics, this dissertation provides a foundation for how adaptive displays might function in team environments. For example, real-time eye tracking data can be analyzed using MdCRQA and scanpath similarity to continuously assess the team's workload and coordination. This implicit monitoring ensures that the system remains unobtrusive, allowing the team to focus on their tasks without additional distractions. Based on the metrics, adaptive displays can adjust in real-time to support the team's needs by highlighting critical information during periods of high workload and reducing clutter or unnecessary visual elements when the team shows signs of overload. These metrics could also serve as the basis for predictive systems that detect when a team is approaching a state of overload or performance breakdown. For instance, a decrease in MdCRQA metrics (e.g., MaxL or EntrV; Table 3.7) could signal that team members are losing synchronization or under high workload, prompting the display to provide additional support or alert the team to recalibrate.

The implications of this work are transformative. It pushes the boundaries of adaptive system design, shifting the focus from individuals to teams and addressing the unique challenges of team dynamics in complex systems. By introducing metrics that can capture team-level states, this dissertation provides a critical first step toward building adaptive displays that optimize team performance, resilience, and collaboration.

7.2.4 Societal Contributions

This dissertation has far-reaching societal implications, offering insights and tools that can improve the safety, efficiency, and resilience of team operations in critical, high-stakes environments. By advancing our understanding of team dynamics and introducing innovative metrics and methodologies, this work contributes to addressing some of society's most pressing challenges, particularly in domains where effective teamwork is essential for success.

As workplaces become increasingly reliant on advanced technologies, the ability to understand and optimize human-machine and human-human interactions becomes more critical (Callari et al., 2024; Stephanidis et al., 2025). This dissertation provides a framework for designing systems that enhance collaboration in these environments by integrating real-time monitoring tools that unobtrusively support team performance and offering guidelines for designing gaze sharing displays that align with team needs, reducing cognitive load and enhancing efficiency. The societal contributions of this work also lie in its ability to anticipate and address future challenges in teamwork and technology. As greater automation becomes more prevalent, the findings of this dissertation can inform the design of systems that maintain human oversight and collaboration, ensuring that teams remain effective in increasingly complex operational contexts (Kyriakou & Otterbacher, 2023; McKay, 2024; Tariq, 2025).

Beyond its immediate applications, this work contributes to educating and training the next generation of researchers, practitioners, and operators. By providing a rigorous methodological framework and actionable insights, it equips professionals in fields like military, aviation, and emergency response to understand and optimize team dynamics (Andrews et al., 2022; Atweh et al., 2022; Mathieu et al., 2017). Additionally, the interdisciplinary nature of this research fosters collaboration across fields such as human factors, AI, and cognitive systems engineering.

Ultimately, the societal contributions of this dissertation extend to enhancing the well-being and safety of people affected by teamwork in high-stakes scenarios. Whether through preventing aviation accidents or improving emergency response outcomes, the principles and findings of this work have the potential to save lives and improve the quality of critical services. By promoting effective teamwork and leveraging technology to support human capabilities, this dissertation aligns with broader societal goals of safety, efficiency, and innovation.

7.3 Future Work

The findings presented in this dissertation open the door to several promising research directions that aim to further our understanding of team collaboration in complex systems and to design tools that enhance team performance. This section synthesizes future research opportunities from each study, weaving them into a cohesive vision for advancing metrics, methodologies, and technologies that can be applied across diverse domains.

Designing adaptive displays for teams represents an uncharted frontier in cognitive systems engineering. While adaptive systems for individual operators is a growing area of research, extending these principles to teams introduces novel challenges, particularly in terms of identifying and monitoring team-level states (Fall et al., 2018; Lu et al., 2021).

This dissertation lays the groundwork for adaptive team displays by introducing metrics such as MdCRQA, which are sensitive to team workload and coordination. These metrics can serve as the basis for real-time adaptive systems that monitor the team's collective cognitive state without imposing additional burdens on individual operators. Future work should investigate whether these metrics can reliably predict performance breakdowns across varying domains and conditions, including task complexity, team size, and workload levels (Atweh & Riggs, 2025a; Atweh & Riggs, 2025b). Moreover, we need to integrate physiological data (e.g., heart rate variability, EEG) and neuroergonomics measures alongside eye tracking metrics to provide a richer, multimodal picture of team states. Future research in this area can also explore how adaptive displays can use AI to support teams proactively. For example, AI could identify signs of overload or desynchronization and provide targeted assistance, such as highlighting critical information or suggesting task reallocations.

While gaze sharing has demonstrated significant potential within UAV operations, its applicability to other domains and contexts remains underexplored. The findings across Chapters 4, 5, and 6 emphasize the need to extend gaze sharing research as follows. First, we need to investigate how gaze sharing performs in fields such as emergency response, air traffic control. For example, gaze visualizations may provide unique advantages for team decision-making in aviation settings or crisis management scenarios. Beyond fixation dots and trails, consider alternative gaze sharing techniques such as heatmaps, shared graphics, or network-centric overlays (Entin et al., 2006; Hiniker & Entin, 1992; Špakov & Miniotas, 2007). Comparing these options can provide insights into their utility and limitations across various operational settings. In remote centers and distributed teams, where delayed communication and interruptions are common, gaze sharing could help mitigate the challenges of asynchronous or disrupted coordination (Atweh & Riggs, 2024; Fischer & Mosier, 2014; Mosier & Fischer, 2021). Future studies should assess how gaze sharing tools can enhance SA and task recovery in these contexts as well.

The short-term experiments conducted in this dissertation highlight the immediate benefits of gaze sharing and advanced metrics, but long-term adaptation and scalability remain open questions. Future work needs to explore how teams adapt to gaze sharing tools and adaptive displays over extended periods, examining whether initial improvements in coordination are sustained or evolve over time. Moreover, we need to assess whether the metrics and tools validated in this dissertation can scale effectively to larger teams or more complex tasks, where communication dynamics and workload distribution may differ significantly

(Atweh et al., 2022).

As we navigate increasingly complex and interconnected environments, the pursuit of effective collaboration remains a cornerstone of success. Teams are at the heart of decision-making and problem-solving in critical domains, from ensuring public safety to managing technological systems that drive the modern world. This dissertation contributes to that pursuit by advancing our understanding of team dynamics and introducing tools and metrics that can improve coordination and resilience in high-stakes settings. The implications of this work extend beyond any single domain, offering a pathway to designing systems and technologies that empower teams to tackle the challenges of the evolving modern world.

7.4 References

Andrews, R. W., Lilly, J. M., Srivastava, D., & Feigh, K. M. (2022). The Role of Shared Mental Models in Human-AI Teams: A Theoretical Review. *Theoretical Issues in Ergonomics Science*, *24*(2), 129–175. https://doi.org/10.1080/1463922X.2022.2061080

Ateş, S. S., Uzgör, M., & Yüksek, K. (2022). UAV Tracking Module Proposal Based on a Regulative Comparison Between Manned and Unmanned Aviation. *Journal of Airline and Airport Management*, *12*(1), 29–47. https://doi.org/10.3926/jairm.206

Atweh, J. A., Moacdieh, N. M., & Riggs, S. L. (2022). Identifying Individual-, Team-, and Organizational-Level Factors that Affect Team Performance in Complex Domains Based on Recent Literature. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1795–1799. https://doi.org/10.1177/1071181322661213

Atweh, J. A., & Riggs, S. L. (2024). Gaze Sharing, a Double-Edged Sword: Examining the Effect of Real-Time Gaze Sharing Visualizations on Team Performance and Situation Awareness. *Human Factors*, *67*(3), 196–224. https://doi.org/10.1177/00187208241272060

Atweh, J. A., & Riggs, S. L. (2025a). Display Design Shapes How Eye Tracking and Workload Predict Team Performance. *2025 Systems and Information Engineering Design Symposium (SIEDS)*.

Atweh, J. A., & Riggs, S. L. (2025b). Predicting Mental Demand of Teammates Using Eye Tracking Metrics: A Machine Learning Approach. *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*. https://doi.org/10.1145/3715669.3726802

Avvenuti, M., & Vecchio, A. (2009). Mobile Visual Access To Legacy Voice-Based Applications. *Proceedings Of The 6th International Conference On Mobile Technology, Application & Systems*, 1–6. https://doi.org/10.1145/1710035.1710097

Callari, T. C., Segate, R. V., Hubbard, E. M., Daly, A., & Lohse, N. (2024). Ethical Framework Human-Robot Collaboration Future People-Centric Manufacturing: Collaborative Endeavour European Subject-Matter Experts Ethics. *Technology in Society*, *78*, 102680. https://doi.org/10.1016/j.techsoc.2024.102680

Chen, G., & Kanfer, R. (2024). The Future of Motivation in and of Teams. *Annual Review of Organizational Psychology and Organizational Behavior*, *11*(1), 93–112. https://doi.org/10.1146/annurev-orgpsych-111821-031621

Coakes, E. W., Coakes, J. M., & Rosenberg, D. (2008). Co-Operative Work Practices And Knowledge Sharing Issues: A Comparison Of Viewpoints. *International Journal Of Information Management*, *28*(1), 12–25. https://doi.org/10.1016/j.ijinfomgt.2007.10.004

Deacon, A. (2020). Team Adaptation in High Reliability Teams. *University of Calgary Digital Repository*. https://prism.ucalgary.ca/bitstreams/d911aa57-6c56-491f-a924-9bb897b4a37d/content

Dietz, A. S., Driskell, J. E., Sierra, M. J., Weaver, S. J., Driskell, T., & Salas, E. (2017). Teamwork Under Stress. *The Wiley Handbook of Teamwork*, 345–365. https://doi.org/10.1002/9781118909997.ch13

Entin, E. E., Hiniker, P., Grier, R., Jefferson, T., Vecchio, G., & Harkins, J. (2006). Enhancing Situational Awareness and Team Performance Using Network-Centric Technologies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(3), 482–486. https://doi.org/10.1177/154193120605000357

Fall, C. L., Quevillon, F., Blouin, M., Latour, S., Campeau-Lecours, A., Gosselin, C., & Gosselin, B. (2018). A Multimodal Adaptive Wireless Control Interface For People With Upper-Body Disabilities. *IEEE Transactions On Biomedical Circuits And Systems*, *12*(3), 564–575. https://doi.org/10.1109/TBCAS.2018.2810256

Fischer, U., & Mosier, K. (2014). Impact Communication Delay and Medium Team Performance and Communication Distributed Teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 115–119. https://doi.org/10.1177/1541931214581025

Giotopoulos, K. C., Michalopoulos, D., Vonitsanos, G., Papadopoulos, D., Giannoukou, I., & Sioutas, S. (2024). Dynamic Workload Management System in the Public Sector. *Information*, *15*(6), 335. https://doi.org/10.3390/info15060335

Golden, S. J., Chang, C. H., & Kozlowski, S. W. (2018). Teams in Isolated, Confined, and Extreme Environments: A Review. *Journal of Organizational Behavior*, *39*, 750–772. https://doi.org/10.1002/job.2288

Hagemann, V., Kluge, A., & Ritzmann, S. (2012). Flexibility Under Complexity: Work Contexts and Team Processes of High-Responsibility Teams. *Employee Relations*, *34*, 475–491. https://doi.org/10.1108/01425451211217734

Harinarayana, T., Krishnan, S. V., & Hota, S. (2024). Lyapunov Guidance Vector Field-Based Waypoint Following by Unmanned Aerial Vehicles. *Journal of Guidance, Control, and Dynamics*, 1–11. https://doi.org/10.2514/1.G008056

Hildmann, H., & Kovacs, E. (2019). Review: Using Unmanned Aerial Vehicles (UAVs) as Mobile Sensing Platforms (MSPs) for Disaster Response, Civil Security and Public Safety. *Drones*, *3*(3), 59. https://doi.org/10.3390/drones3030059

Hiniker, P. J., & Entin, E. E. (1992). Examining Cognitive Processes in Command Crisis: New HEAT Experiments on Shared Battle Graphics and Time Tagging. *Proceedings of the 1992 Symposium on Command and Control Research.*

Hirshfield, L. M., Wickens, C., Doherty, E., Spencer, C., Williams, T., & Hayne, L. (2023). Toward Workload-Based Adaptive Automation: The Utility of fNIRS for Measuring Load in Multiple Resources in the Brain. *International Journal of Human–Computer Interaction*, 1–27. https://doi.org/10.1080/10447318.2023.2266242

Hofmaenner, D. A., Herling, A., Klinzing, S., Wegner, S., Lohmeyer, Q., Schuepbach, R. A., & Buehler, P. K. (2020). Use of Eye Tracking in Analyzing Distribution of Visual Attention Among Critical Care Nurses in Daily Professional Life: An Observational Study. *Journal of Clinical Monitoring and Computing 2020*, 1–8. https://doi.org/10.1007/S10877-020-00628-2

Kyriakou, K., & Otterbacher, J. (2023). In Humans, We Trust: Multidisciplinary Perspectives on Requirements for Human Oversight Algorithmic Processes. *Discover Artificial Intelligence*, *3*(1), 44. https://doi.org/10.1007/s44163-023-00092-2

Lu, Y., Adrados, J. S., Chand, S. S., & Wang, L. (2021). Humans Are Not Machines—Anthropocentric Human–Machine Symbiosis For Ultra-Flexible Smart Manufacturing. *Engineering*, *7*(6), 734–737. https://doi.org/10.1016/j.eng.2020.09.018

Malakis, S., & Kontogiannis, T. (2023). Team Adaptation and Safety in Aviation. *Safety Science*, *158*, 105985. https://doi.org/10.1016/j.ssci.2022.105985

Mangaroska, K., Sharma, K., Gašević, D., & Giannakos, M. (2022). Exploring Students' Cognitive And Affective States During Problem Solving Through Multimodal Data: Lessons Learned From A Programming Activity. *Journal Of Computer Assisted Learning*, *38*(1), 40–59. https://doi.org/10.1111/jcal.12590

Mathieu, J. E., Hollenbeck, J. R., van Knippenberg, D., & Ilgen, D. R. (2017). A Century of Work Teams in the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*(3), 452–467. https://doi.org/10.1037/apl0000128

McKay, M. H. (2024). Realizing Promise AI Governance Involving Humans-in-the-Loop. In H. Degen & S. Ntoa (Eds.), *HCI International 2024 – Late Breaking Papers. HCII 2024* (Vol. 15382). Springer, Cham. https://doi.org/10.1007/978-3-031-76827-9_7

Mosier, K. L., & Fischer, U. M. (2021). Meeting Challenge Transmission Delay: Communication Protocols Space Operations. *Human Factors*, *65*(6), 1235–1250. https://doi.org/10.1177/00187208211047085

Papamitsiou, Z., Pappas, I. O., Sharma, K., & Giannakos, M. N. (2020). Utilizing Multimodal Data Through FsQCA To Explain Engagement In Adaptive Learning. *IEEE Transactions On Learning Technologies*, *13*(4), 689–703. https://doi.org/10.1109/TLT.2020.3020499

Singh, L. (2024). Impact of Induced Stress on Individual Performance, Team Coordination, and Human-AI Collaboration. *University of Southampton Thesis*. https://eprints.soton.ac.uk/487744/

Škvareková, I., Pecho, P., Ažaltovič, V., & Kandera, B. (2020). Number of Saccades and Fixation Duration as Indicators of Pilot Workload. *Transportation Research Procedia*, *51*, 67–74. https://doi.org/10.1016/j.trpro.2020.11.009

Špakov, O., & Miniotas, D. (2007). Visualization of Eye Gaze Data using Heat Maps. *Elektronika Ir Elektrotechnika Medicine Technology*, *115*. https://etalpykla.vilniustech.lt/handle/123456789/149656

Steitz, B. D., Unertl, M., K., & Levy, M. A. (2020). Characterizing Communication Patterns Among Members Of The Clinical Care Team To Deliver Breast Cancer Treatment. *Journal Of The American Medical Informatics Association*, *27*(2), 236–243. https://doi.org/10.1093/jamia/ocz151

Stephanidis, C., Salvendy, G., Antona, M., Duffy, V. G., Gao, Q., Karwowski, W., Konomi, S., Nah, F., Ntoa, S., Rau, P. L. P., Siau, K., & Zhou, J. (2025). Seven HCI Grand Challenges Revisited: Five-Year Progress. *International Journal Human–Computer Interaction*, 1–49. https://doi.org/10.1080/10447318.2025.2450411

Tariq, M. U. (2025). Navigating Ethical Frontier-Human Oversight AI-Driven Decision-Making System. In *Enhancing Automated Decision-Making Through AI* (pp. 425–448). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-6230-3.ch013

Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task Performance and Eye Activity: Predicting Behavior Relating to Cognitive Workload. *Aviation, Space, and Environmental Medicine*, *78*(5), B176–B185.

Vesper, C., Schmitz, L., Safra, L., Sebanz, N., & Knoblich, G. (2016). The Role of Shared Visual Information for Joint Action Coordination. *Cognition*, *153*, 118–123. https://doi.org/10.1016/j.cognition.2016.05.002

Waterson, P., Robertson, M. M., Cooke, N. J., Militello, L., Roth, E., & Stanton, N. A. (2015). Defining the Methodological Challenges and Opportunities for an Effective Science of Sociotechnical Systems and Safety. *Ergonomics*, *58*(4), 565–599. https://doi.org/10.1080/00140139.2015.1015622

Appendix A

NASA-TLX

Mental Demand

How mentally demanding was the task?

Very Low Very High

Physical Demand

How physically demanding was the task?

Very Low Very High

Temporal Demand

How hurried or rushed was the pace of the task?

Very Low Very High

Performance

How successful were you in accomplishing what you were asked to do?

Very Low Very High

Effort

How hard did you have to work to accomplish your level of performance?

Very Low Very High

Frustration

How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low Very High

Appendix B

Pre-Experiment Questionnaire

Table B.1: Pre-Experiment Survey in Chapters 3, 4, 5, and 6

Questions	Options
What is your age?	Free Text
What is your gender?	Male, Female, Other
What is your dominant hand?	Left, Right, Ambidextrous
If you know it, please provide your visual acuity (e.g., 20/20, 20/10)? If not, type "N/A"	Free Text
Do you wear glasses?	(1) Yes, and I am currently wearing them. (2) Yes, but I am not currently wearing them. (3) No
Do you wear contact lenses?	(1) Yes, and I am currently wearing them. (2) Yes, but I am not currently wearing them. (3) No
Are you currently wearing mascara or eye makeup of any kind?	Yes, No
What is your experience flying an Unmanned Aerial Vehicle?	No experience, Novice, Intermediate, Expert

Table B.1 – continued from previous page

Questions	Options
Do you currently have a pilot's license? If no, say "No". If yes, indicate the estimated number of hours/year you fly	Free Text
On the scale of 0-10 below, please rate how alert or sleepy you feel right now (**Note: 0 is very alert and 10 is very sleepy*).	Scale 0-10
During an average week, how many combined hours do you spend playing any types of video games?	Free Text
Would you consider yourself a novice, intermediate, or expert video game player?	Novice, Intermediate, Expert, N/A

Appendix C

Debriefing Questionnaires for Chapters 3, 4, 5, and 6

Table C.1: Debriefing Questionnaire for Chapter 3

Questions	Options
Can you describe the strategy or strategies you developed to complete the UAV tasks in the low workload scenario?	Free Text
Did your strategy change or adapt when you transitioned to the high workload scenario? Why or why not?	Free Text
Reflecting on your performance, do you believe your strategy was effective in achieving the task objectives? Why or why not?	Free Text
How would you rate the overall effectiveness of communication with your teammate during the tasks?	(a)Very effective (b) Effective (c) Neutral (d) Ineffective (e)Very ineffective
Please elaborate on your previous answer	Free Text

Table C.2: Debriefing Questionnaire for Chapter 4

Questions	Options
Please describe any strategy you developed to complete the task (e.g., how you completed individual tasks, scanning strategy, etc.).	Free Text
How effective do you think gaze sharing was as a tool in this UAV domain? Please rate your response on a scale of 1 to 5, with 1 being not effective at all and 5 being extremely effective.	Scale 1-5
Did you have any difficulties interpreting the gaze sharing information displayed using either of the visualization techniques? If so, please explain.	Free Text
Did you find one visualization technique more effective than the other for displaying gaze sharing information? If so, which one and why?	Free Text
How frequently did you refer to the gaze sharing information during the task? Please rate your response on a scale of 1 to 5, with 1 being never and 5 being constantly.	Scale 1-5
Did the gaze sharing information provided by either visualization technique impact your decision-making during the task? If so, please explain.	Free Text
Were there any limitations or drawbacks to using gaze sharing as a tool in the UAV C2 domain that you experienced during the task? If so, please explain.	Free Text

Table C.3: Debriefing Questionnaire for Chapter 5

Questions	Options
Describe the communication methods you and your teammate used during the study (e.g., verbal, non-verbal cues, hand signals).	Free Text
Were there any specific moments during the study when communication was particularly important or effective? Please describe.	Free Text
Please describe any strategy you developed to complete the task	Free Text
How did you and your teammate coordinate task allocation and strategy adjustments in different conditions?	Free Text
How did the trail feature contribute to your sense of shared awareness with your teammate during the tasks?	*(Can choose more than one option)* (1) Enhanced shared awareness. (2) Improved understanding of teammate's focus. (3) Increased task coordination. (4) No significant impact. (5) Other.

Table C.3 – continued from previous page

Questions	Options
Did the Trail gaze sharing feature improve your ability to predict or anticipate your teammate's actions or decisions?	(1) Yes, it improved predictability when communication was present with it only. (2) Yes, it improved predictability when the trail was alone and communication was absent. (3) Yes, it improved predictability when communication was both present and absent. (4) No, it did not affect predictability at all. (5) Other.
How did the presence or absence of gaze sharing (seeing your teammate's gaze trail) influence your decision-making and task coordination?	Free Text
How did the combination of trail gaze sharing and verbal communication affect your overall performance in the UAV management tasks?	*(Can choose more than one option)* (1) Enhanced performance. (2) Provided redundancy in information, which is needed. (3) Provided redundancy in information, which increased workload. (4) No significant impact. (5) Other.

Table C.3 – continued from previous page

Questions	Options
What was the optimal setup for you? Please explain and order the conditions if you prefer that. (1) No gaze sharing and verbal communication. (2) Trail gaze sharing and verbal communication. (3) Trail gaze sharing and no verbal communication.	Free Text
Please share any additional comments, suggestions, or feedback about your experience.	Free Text

Table C.4: Debriefing Questionnaire for Chapter 6

Questions	Options
Please rank the following gaze sharing conditions in order of your preference (1 = Most Preferred, 3 = Least Preferred). Gaze Sharing Always On Gaze Sharing Always Off On/Off Gaze Sharing Button	Drag and Drop
Please explain your reasoning for the previous rankings.	Free Text
When you did have the option to use the "On/Off Gaze Sharing" button, how often did you use it?	(1) Never. (2) Very Rarely. (3) Rarely. (4) Occasionally. (5) Often. (6) Very Often.
How would you describe the helpfulness of the gaze sharing feature and the on/off button during the experiment?	(1) Gaze sharing and on/off button were both helpful. (2) Only gaze sharing was helpful; the on/off button did not affect my experience. (3) The on/off button was helpful; gaze sharing alone was not helpful. (4) Neither gaze sharing nor the on/off button were helpful.

Table C.4 – continued from previous page

Questions	Options
Please explain why and how did you (or did you not) use the on/off gaze sharing button in terms of communication, task executions, and interruptions.	Free Text
How helpful did you find the gaze sharing feature in general for maintaining awareness of your teammates' actions?	(1) Not Helpful At All. (2) Slightly Helpful. (3) Moderately Helpful. (4) Helpful. (5) Very Helpful.
Please describe how you reoriented yourself with your assigned task after an interruption. Please provide any details on how the interruption affected your performance, focus, use of gaze sharing, and/or the on/off button.	Free Text

Appendix D

Chapter 5 Codebook

Table D.1: Codebook for Debriefing Questionnaire Thematic Analysis

Theme	Inductive Code	Definition	Example Quote
General Communication Strategies	Verbal Task Updates	Participant verbally communicates actions taken on tasks.	"We used verbal communication to explain the tasks we were actively completing." – P7, P2
	Chat Notifications	Participant uses verbal communication to share chat-related information.	"There's a message for your task." – P1, P1
	Confirmation Strategy	Saying out loud what one is about to do to keep the partner informed.	"He was saying he was going to detect the target at India. . . " – P15, P1
Initial Task Sharing Strategies	Task Splitting	Dividing tasks so each participant has specific responsibilities.	"I focused on rerouting UAVs and answering the questions. . . " – P7, P2
	Joint Task Management	Tasks that required collaborative input were completed together.	"We shared the tasks of fixing fuel leaks and answering questions. . . " – P14, P1
	Pre-Planning	Teams developed a strategy before starting the task.	"It was important to have established a plan. . . " – P10, P1

Continued on next page

Table D.1 – continued from previous page

Theme	Inductive Code	Definition	Example Quote
Adjustments to Communication Strategies	Minimal Verbal Communication	Participants reported that frequent talking was unnecessary.	"There was really no communication besides a quick verbal remark..." – P19, P1
	Communication at Critical Moments	Talking occurred mainly during complex or multi-task situations.	"When there were multiple complex tasks... speaking was helpful." – P1, P1
	Reduced Talk with Gaze Sharing	Gaze sharing reduced the need for verbal updates.	"I didn't have to verbally do that." – P15, P2
Adjustments to Task Strategies	Verbal Coordination in No Gaze Condition	Relying on verbal cues to coordinate actions when gaze sharing wasn't available.	"With no gaze sharing... we verbally updated each other." – P10, P1
	Dynamic Task Reallocation	Participants adjusted task roles based on partner's behavior.	"If I saw my partner was already on a task, I'd just move on..." – P22, P1
	Gaze-Based Role Adjustment	Using partner's gaze to adjust one's own task without verbal coordination.	"We used gaze sharing to see who took on what task." – P15, P1

Continued on next page

Table D.1 – continued from previous page

Theme	**Inductive Code**	**Definition**	**Example Quote**
Influence of Gaze Sharing on Decision-Making	Enhanced Coordination	Gaze helped with understanding teammate focus.	"Seeing their trail gaze confirms what we communicate." – P21, P1
	Trust and Predictability	Gaze increased confidence in what the teammate was doing.	"It was more difficult to understand what my partner was focused on…" – P19, P1
	Distraction from Gaze Sharing	Gaze trail sometimes caused distraction or hindered performance.	"It was a little bit distracting having the tracker darting around the screen…" – P14, P1