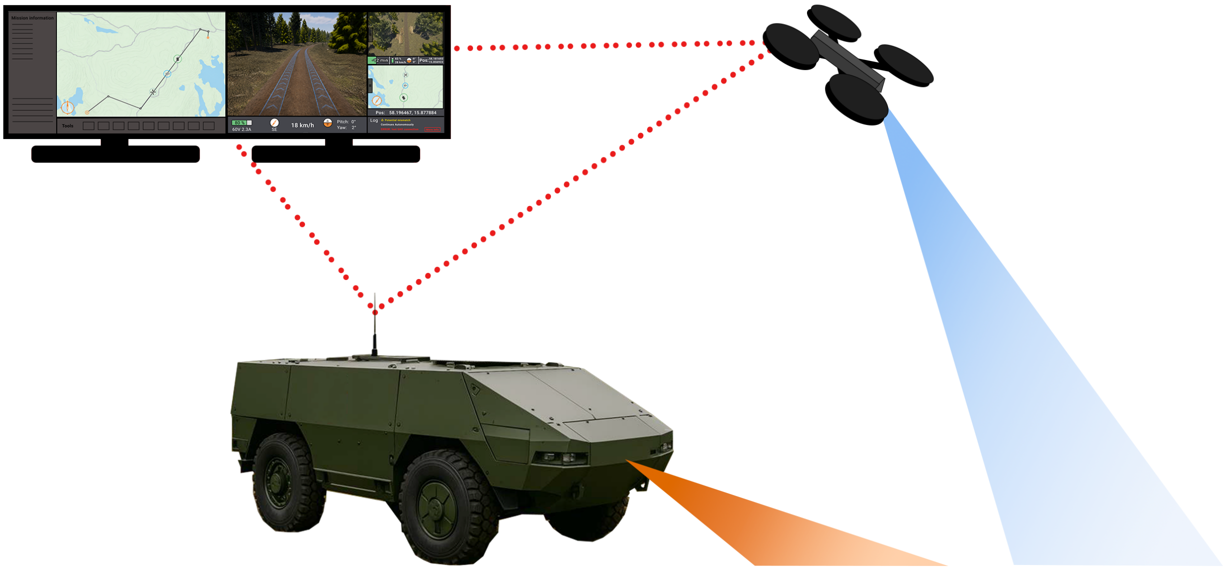


Human-Autonomy Teaming: Effects of Autonomy Level and UAV Integration on Operator Decision-Making

Master's thesis in Industrial Design Engineering

SOFIE ANDERSSON
AMIN OTHMAN

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Human-Autonomy Teaming: Effects of Autonomy Level and UAV Integration on Operator Decision-Making

SOFIE ANDERSSON
AMIN OTHMAN

Department of Industrial and Materials Science
Division of Design and Human Factors
Chalmers University of Technology
Gothenburg, Sweden 2026

Human-Autonomy Teaming: Effects of Autonomy Level and UAV Integration on Operator Decision-Making

SOFIE ANDERSSON
AMIN OTHMAN

Supervisor: Linda Pipkorn, Department of Industrial and Materials Science
Examiner: Lars-Ola Bligård, Department of Industrial and Materials Science

Master's Thesis 2026
Department of Industrial and Materials Science
Division of Design and Human Factors
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Operator-UGV-UAV system.

Typeset in LaTeX
Printed by Chalmers Digitaltryck
Gothenburg, Sweden 2026

Abstract

As autonomous systems become increasingly integrated into safety-critical operations, understanding how human operators collaborate with heterogeneous robot teams is essential for effective system design. This thesis investigates how the level of automation (LOA) of an unmanned ground vehicle (UGV) and the integration of unmanned aerial vehicle (UAV) information influence operator state, performance, and decision-making in a human-UGV-UAV team.

A controlled experiment was designed, varying UGV automation level — Assisted-Autonomous (AA-LOA), where the operator receives take-over requests (TORs), and Autonomous (A-LOA), where the system operates without operator input — and UAV information type — Conflicting or Non-Conflicting with the UGV sensor data. Thirty-three participants completed video-based simulations, representing logistics transport scenarios in which the UAV flew ahead of the UGV acting as a forward-looking sensor. Quantitative (e.g., operator response time, take-over probability, and eye-tracking gaze) and qualitative (e.g., mental workload and observations) data was collected.

Results show that AA-LOA significantly increases perceived responsibility compared to A-LOA, without a corresponding increase in mental workload. Operators integrated the UAV-view throughout the mission, and their subjective estimates of UAV reliance closely matched that of the objective eye-tracking data.

Conflicting UAV information increased the likelihood of manual intervention but did not significantly extend decision time. Despite a 50% conflict rate, operators consistently valued the UAV-view for its contextual overview and sense of safety. Demographic factors showed meaningful effects: gamers reported lower mental workload and trended toward faster decisions, while high driving frequency and strong sense of direction were the strongest predictors of longer decision times and greater confidence in own judgment during conflict scenarios.

Based on these findings, 15 design guidelines were developed addressing LOA selection, transparency, UAV integration, and temporal awareness — including UAV recap functionality and a time-delta indicator to help operators contextualize temporally misaligned information. A dual-screen operator interface concept design implementing these guidelines is presented. To conclude, the findings in this thesis suggests that assisted-autonomy combined with time-transparent UAV integration provides operators with the best conditions for effective situational awareness and decision-making in heterogeneous human-autonomy systems.

Keywords: Situational awareness, Human-robot interaction, Human-Autonomy, Level of automation, UAV reliance, UGV, Decision-making, Take-over requests

Acknowledgments

Firstly, we would like to express our deep gratitude to our supervisor, Linda Pipkorn, for guiding us through this jungle they call research. For two design engineers this new and daunting task could have been a scary and stressful project, but it instead felt safe and exiting thanks your thoughtful inputs and patient explanations.
Who knew research could be fun?

Secondly, we want to thank our examiner, Lars-Ola Bligård for his constant support not only throughout this project but throughout our years here at Chalmers, the program would not be the same without you and you've made a mark in our lives that we're forever grateful for. As you might know, research isn't always a walk in the park and getting us through the trials and tribulations where his echoing wise words, *"Trust the process"*, and trust it we forever shall.

Thirdly, we wish to thank Dag Berjsö at EIRA Systems for daring to innovate and bring in students from all kinds of disciplines to develop our skills and push us to new heights. We also want to extend a big thank you to Carl, Nicke, Peter and Sandra at Tactisim for giving it their all in helping us realize our simulator ideas. Furthermore, we would like to thank Christer Ahlström from VTI for lending us, two students, your expensive eye-tracking glasses which really served to elevate our research and taught us a lot.

Finally, we want to give a huge thank you to all the participants who took time out of their day and provided us with their insights and data, this would not have been possible without your contributions.

On a personal note, I, Amin, would like to thank and acknowledge my friends and family who've listened and supported me for the past few months while I've ranted about gaze plots, time syncs and one abbreviation after the other, it means the world to me, and I love you all to the moon and back. And while I still have you here, I'd like to express my deepest gratitude to all the wonderful people I've gotten the chance to meet through out these past years throughout my academic journey, both students and faculty alike. It really is the people that make it. I'm forever grateful, and I'll never take it for granted. Thank you all ♡

On a personal note, I, Sofie, would also like to express my gratitude to my family and friends, who have brightened my days and supported me throughout my entire journey at Chalmers. I deeply appreciate your incredible support, your belief in me, and your encouragement to find the path that feels right for me. Looking back, I truly feel that I have found that path through this journey. I feel truly fortunate to have experienced these years and to have met so many wonderful people along the way. I will carry these memories and friendships with me for the rest of my life. Thank you all.

Sofie Andersson, Amin Othman, Gothenburg, July 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

A-LOA	Autonomous Level of Automation
AA-LOA	Assisted Autonomous Level of Automation
AOI	Area of Interest
C	Conflicting UAV information
HDF	High Driving Frequency
IQR	Interquartile Range
LOA	Level of Automation
NC	Non-conflicting UAV information
SAR	Search and Rescue
SoD	Sense of Direction
TOR	Take-Over Request
UAV	Unmanned Aerial Vehicle
UGV	Unmanned Ground Vehicle

Contents

1

Introduction

1.1 Background

Research into heterogeneous robot and human teams has become a hot topic with the recent rapid development of artificial intelligence and autonomous systems (Liu et al., 2022). Collaboration between unmanned ground vehicles (UGVs) and unmanned aerial vehicles (UAVs) enables the combination of the complementary strengths of UGVs and UAVs, such as aerial overview and ground-level interaction (Grocholsky et al., 2006). This type of system can extend human capabilities by providing semi-autonomous support in critical and dangerous situations, such as search and rescue (SAR) missions in earthquake, and tsunami-stricken areas (Munasinghe et al., 2024).

The human-UGV-UAV team is a system composed of machine-to-machine and machine-to-human communication, where the operator of the system is situated at a remote control station from which they can give orders in the form of way-points to the autonomous vehicles or take over manual, remote control. The UGV is considered the actuator, and the UAV the sensor of the system (Liu et al., 2022). While increased autonomy in the UGV can reduce operator workload and improve mission efficiency, human intervention remains essential in ambiguous or high-risk situations (Parasuraman et al., 2000). In semi-autonomous SAR missions, remote operators may need to assist the UGV-UAV system with decision-making in uncertain situations, such as when its sensors are disrupted by weather or terrain. For example, the remote operator may need to decide whether to allow an autonomous UGV to continue its course or to intervene and take manual control when it reaches a steep negative incline. Such decisions may be made under time pressure and uncertainty, potentially based on multiple, and sometimes conflicting, information sources.

The benefit of UAV recognizance in allowing the UGV to plan its path further ahead, in combination with providing operators more time to react to potential situations and providing a top down perspective can not be understated; However, the geographical distance between the UAV and UGV poses a threat to operators Situational Awareness as the information being fed from its cameras is temporally misaligned with that of the UGV. For example, the UGV might initiate a Take-over request (TOR) to the operator, based on its sensor data not aligning with UAV information.

If the UAV was always above the UGV, this would mainly be because of sensor disruptions, but by adding the geographical, and thereby the temporal factor, the physical situation that the UAV has reported on could have changed drastically. These changes could be huge, but most likely they will be small frequent false alarms, and studies show that high false alarm rates can seriously undermine the effectiveness of the system, (Endsley & Jones, 2025). Worse, they could lead to operators making incorrect decisions in high-risk stressful scenarios based on expired data. The successful integration of a UAV into the Human-Autonomy team therefore hinges on the intercommunication of the system, which poses questions of what and how information should be shared and presented to the operator to support the mission, enhance operator experience and ultimately ensure optimal performance.

Considerable research has investigated the effects on operators in single machine control handovers, however less is known about the effect on operators state and performance in heterogeneous drone teams with varying levels of autonomy (LOA) and integration of UAV-views. This thesis aims to investigate what LOA in the UGV is preferred in regard to operators mental state, particularly in the cases where the information of the UAV is in conflict with that of the UGV. As this is a relatively new field with experienced operators being few and far between, other demographic factors such as gaming experience and driving frequency are used to investigate how the effects of autonomy levels and the integration of the UAVs exocentric view into the UGVs egocentric view might affect different people.

Understanding how operators incorporate the UAV information in addition to the effects of UGV LOA on decision-making can help inform the design of balanced and efficient human-UGV-UAV teams. In such systems, functions are allocated in a way that support the operators without either overloading them with information or leaving them out of the loop. Designing with operator's cognitive, emotional, and physiological condition in mind, such as, perceived responsibility, mental workload, and confidence, can greatly influence how effectively the operator can interact with and manage the system during mission execution. Performance factors such as time and quality of decision are deemed to correlate with operator state and situational awareness, so by understanding one, the other can be improved.

1.2 Project

The project is conducted at the request of the company EIRA-systems (EIRA Systems, 2026) in collaboration with the Mechanical Engineering department at Chalmers University of Technology. The company is developing a semi-autonomous UGV for a human-UGV-UAV search and rescue team, where the UGV acts as an actuator and the UAV as a sensor with a human in the loop. The tasks range from delivering supplies from point A to point B, to conducting recognizance in unknown terrain. To ensure that the human operator stays in the loop with the ability to intervene and take control when they deem it necessary, special attention must be given to ensuring that they understand the system's current state, anticipate its behavior, and intervene, without experiencing an excessive cognitive load.

The company therefore seeks to explore how human–autonomy teaming can be designed to support effective supervision, smooth control handovers, and appropriate allocation of functions between human and autonomous agents in autonomous and semi-autonomous missions.

The main limitation of the project is the time frame of 20 weeks and the fact that there is no pre-existing UI or control station. Additionally, both the UGV and the training simulator are in prototype stages, being developed in parallel to the research being conducted.

1.3 Aim

The aim of this thesis is to investigate how the automation level of a UGV and the integration of UAV information influence operator state, performance and decision-making when making control handover decisions in a human-UGV-UAV system. This is to be conducted through a 2 x 2 factors experiment, treating the UAV strictly as a sensor to the UGV, in a Sensor – Actuator system, while disregarding auxiliary factors such as the design of operator control station, UGV and UAV. Based on the findings, the thesis further aims to develop design guidelines for UAV integration, handover support, and function allocation in human–autonomy systems.

To support the aim, the following research questions were defined:

- **RQ1:** To what extent is operators state affected by UGV autonomy level?
    - **RQ1.1:** To what extent do demographic factors play a role in operator state in different UGV autonomy levels?
- **RQ2:** To what extent is operators state and performance affected by UAV integration during decision-making?
    - **RQ2.1:** To what extent do demographic factors play a role in operator state and performance during decision-making?
- **RQ3:** How can functions be appropriately allocated between the human operator, the UGV, and the UAV across different phases of the mission?

1.4 Thesis outline

The remainder of this thesis is organized as follows. Chapter 2 presents the methodology used in this project. Chapter 3 describes the experiment design, including a walkthrough of the experiment, the experimental setup, group structure, procedure, participants, and eye-tracking. Chapter 4 presents the results by answering the three research questions in sequence and includes both quantitative and qualitative analyses. Chapter 5 presents design guidelines and the proposed design concept, developed based on the findings from the experiment and the literature review. Chapter 6 presents the discussion of the aim, method, results, design guidelines, ethical, societal, and ecological aspects, AI use, and direction for future work. Finally, chapter 7 presents the conclusion.

2

Methodology

This chapter describes the methodological approach used to structure and conduct the project. It begins by briefly outlining the purpose of each methodological step and the outcomes it was intended to produce, followed by a detailed description of how these steps are implemented throughout the project. An overview of the methodology is presented in Figure 2.1. A GANTT-chart of the timeline is shown in Appendix A.

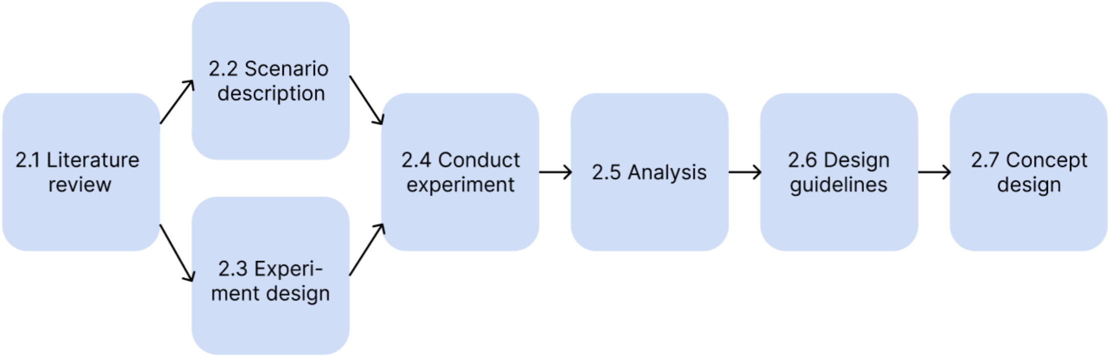


Figure 2.1: Methodology overview

2.1 Literature review

The project began with a literature review focused on human–robot interaction, function allocation, control handover, trust, and responsibility in semi-automated robotic systems. The literature review provided insights into previous work and findings within the academic field, on which to build new theories and experiments.

The search strategy centered around a 2022 meta review of UAV-UGV collaboration by Munasinghe et al., 2024 from which referenced papers were selected and reviewed. The focus was mainly on chapter 4.2.1, which outlines scenarios where the UAV acts as a sensor and the UGV acts as an actuator and decision maker. Using this paper as a foundation, the literature search was expanded by "snowballing" through referenced paper, mapping out the wider academic landscape of UGV-UAV cooperation.

2.2 Scenario description

Following the literature review, a scenario description was developed to provide a clear and consistent representation of the context in which the experiments were conducted. The experimental scenario, illustrated in Figure 2.2, depicts a UGV logistics transport from point A to point B, where the UGV might encounter scenarios that require intervention from an operator. The scenario description was created through describing texts on a timeline, and the use of brainstormed storyboards (See Appendix B).

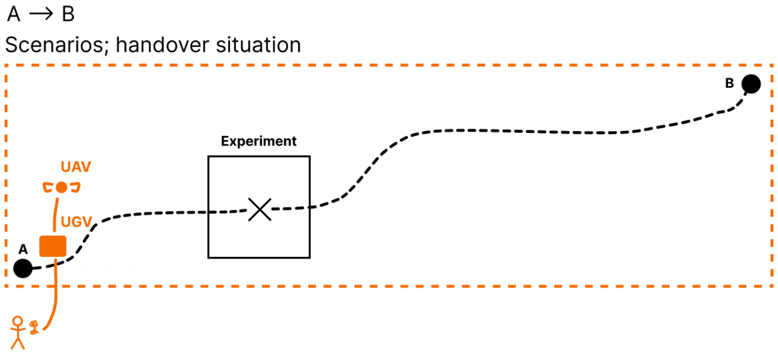


Figure 2.2: Illustration of scenario

2.3 Experiment design

In parallel with the scenario development, the experiment design was developed iteratively. Initially, a series of scenario sketches was created to explore and refine concept ideas, see Appendix B. The variables to be tested were specified based on the research questions and literature review. A prototype of the scenarios and conditions was then created and handed over to the company Tactisim, which was commissioned to reconstruct them in Unreal Engine, a real-time 3D creation tool, for use in research. The experiments were designed using two levels of UGV automation and two types of UAV information, resulting in a 2 × 2 factorial experiment design, see Figure 2.3. Outcomes include Mental Workload & Responsibility, confidence & Reliability, Decision time & Takeover probability, and Gaze share.

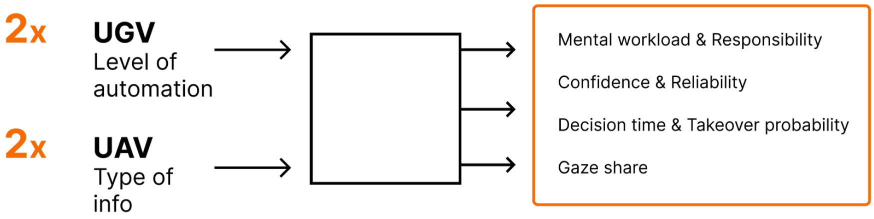


Figure 2.3: Experiment design

The experiment consisted of simulations of the UGV and UAV in different scenarios with different conditions. In addition to the simulations and Take-Over Requests (TOR), participants were asked to answer questions with the keyboard. Videos, TOR and questionnaire items were compiled and presented through the use of HTML and JavaScript, enabling the study to be executed directly within a web browser. Participant data such as demographic factors, choices, times and text-input was directly output into a .csv-file after each experiment, while, gaze data was stored locally for later processing.

Before the experiment started, there was a practice session with two TOR to familiarize the participant with the interface and how to make a decision. This was motivated by previous research showing a significant decrease in people's TOR response times after the first and second TOR (Gold et al., 2018 and Zhang et al., 2019).

The two automation levels in the experiment were Assisted-Autonomous, where the operator receives take-over requests, and Autonomous, where the system makes all decisions itself and the operator monitors it, unable to intervene. The different types of UAV information were designed to either be Conflicting or Non-Conflicting with the information provided by the UGV.

A function allocation of the system was used when designing the systems, see Appendix C. The function allocation first describes the system and then its agents, with its functional purpose being to *Fetch items semi-autonomously across different terrain*. To achieve this, responsibility was distributed among the operator for mission control, the UGV for mission execution, and the UAV for mission assistance, following the principle from the literature review that the UGV serves as the actuator and the UAV as a sensor. Additionally, values and priority measures, purpose-related functions, object-related functions and physical objects were listed for the system and its agents.

The experimental questions were derived from the literature and refined through discussions with the project supervisor. To measure participants' mental workload for each mission, the NASA-Task Load Index (NASA-TLX) was used. NASA-TLX evaluates perceived workload using a multidimensional rating scale (Hart & Staveland, 1988). Only the mental demand subscale was used in this study. To measure Situational Awareness (SA), three questions were aimed at measuring Endless' three hierarchical SA levels (Endsley & Jones, 2025):

Level 1 – Perception of the elements in the environment,
Level 2 – Comprehension of the current situation,
Level 3 – Projection of future status

2.4 Conduct experiment

The experiment was conducted with 33 participants. The participants were divided into four groups with different orders of the A-LOA and AA-LOA mission, and order of starting with Conflict or Non-Conflict scenarios. Quantitative and qualitative data were collected throughout the execution of the experiment, in which participants were exposed to sequential take-over requests and answered questions. The experiment was executed on a computer screen, with a keyboard for answering questions and making decisions, and eye-tracking glasses for gathering gaze data.

2.5 Analysis

The experiment was followed by a systematic analysis of the data gathered. During the simulations, quantitative data were collected on responses, timestamps, and gaze data. Qualitative data were collected through written questions throughout the experiment. Additionally, comments from the participants during and after the experiment were written down. The quantitative data were analyzed through box plot comparisons, histograms and stacked area plots, as well as, tree plots, and logit-fit. Different statistical tests were performed to analyze significance in the results, all of which are listed in Table 2.1 below.

Table 2.1: Comparison of Analytical Methods

	Wilcoxon signed-rank test	**Mann–Whitney U test**	**Mixed-effects models**
Type	Non-parametric	Non-parametric	-
Data structure	Paired / one-sample	Two independent groups	Hierarchical / clustered / longitudinal with continuous outcome
Key assumptions	Symmetric distribution of differences; ordinal or continuous outcome; paired observations	Independent samples; ordinal or continuous outcome; same distributional shape under H_0	Random effects normally distributed or continuous outcome; observations conditionally independent; normality of residuals and random effects
Handles repeated measures?	Yes (paired)	No	Yes

For the qualitative analysis, the KJ-method (Bligård, 2015) was used. For each question, all answers were written on individual Post-it notes in Figma and then rearranged and clustered into groups based on similar content. The groups were named based on the Post-it notes in each group. For the question regarding whether participants would prefer a system with the UAV-view and why, responses were first sorted into yes/no categories and, for "no" responses, further categorized based on the reported reason. Four different reasons were identified in the data. For the question addressing *how useful* they found the UAV information, responses were grouped on a five-point scale ranging from *not helpful at all* to *very helpful*, based on the exact wording of the participants' responses. A similar approach was used for the question addressing whether participants viewed the UAV display due to its usefulness or due to uncertainty, ranging from *useful* to *unsure* on a four-point scale. The notes from the experiment were reviewed separately to identify participants' opinions on potential interface improvements and to understand what was perceived as unclear.

2.6 Design guidelines

Design guidelines were developed based on the findings. The guidelines translated the findings into actionable design insights for a Human-Autonomy system, focusing on information presentation, handover design, and function allocation.

2.7 Concept design

A concept design that implements the guidelines was created and presented to illustrate their practical application. This can be seen as an evaluative step, where the guidelines are tested and reflected upon through their translation into a concrete design proposal. The concept design was developed through brainstorming sessions exploring different inputs, outputs, and additional features, guided by design principles and findings derived from the literature review (see Appendix L).

3

Experiment Design

This chapter will present the experiment design, followed by the experimental setup, group structure, experiment procedure, participants and eye tracking.

The experiment consisted of video simulations of the UGV driving, a UAV-view in the top right corner of the screen, subsequent take-over requests, and questionnaire items. The UAV was flying approximately 10 seconds ahead of the UGV. The experiment followed a 2 × 2 design, consisting of two Levels of Automation (LOA) and two different conditions regarding the UAV information: Conflicting and Non-Conflicting UAV information, see Figure 3.1. The two LOAs were Autonomous (A-LOA), in which the participant did not make any decisions and was unable to intervene, and Assisted-Autonomous (AA-LOA), where the participant received TORs from the system. In the TORs, the participants must decide whether to allow the UGV to continue operating autonomously or to take over manual control. Each LOA was represented with a separate mission in the experiment.

In both LOAs, there were two scenarios: Truck and Logs, which could be potential obstacles in the way of the UGV. But when the UGV drove towards a potential obstacle in the AA-LOA mission, there would never actually be an obstacle in the way. However, the UAV-view might display one, leading to Conflicting (C) information between the UGV and UAV information. If the UAV displayed congruent information with the UGV, with no obstacle in the way, the information was Non-Conflicting (NC). In the AA-LOA mission, a Hinder (H) condition (see Appendix D) with obstacles both in the UGV-view and UAV-view was included for variation and was not used for analysis.

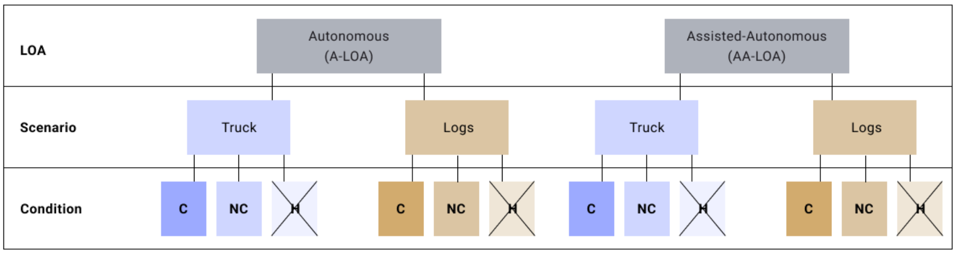


Figure 3.1: Overview of LOAs, scenarios and conditions

The overall structure of the experiment, see Figure 3.2, consisted of a practice session with two TORs, an initial questionnaire with questions about demographics, one mission for the A-LOA system and one for the AA-LOA system, an intermediate questionnaire between the two missions, and a final questionnaire. Both missions consisted of 10 videos each and were split into C, NC and H conditions. After each video, the participants were asked about their confidence in the decisions being made and their perceived reliability of the system. In the intermediate questionnaire, there were questions pertaining to mental workload, trust and responsibility. In the final questionnaire, the same questions as in the intermediate questionnaire were asked again for the other LOA system, with additional questions regarding the autonomous agents, reliability, Situational Awareness and UAV integration.

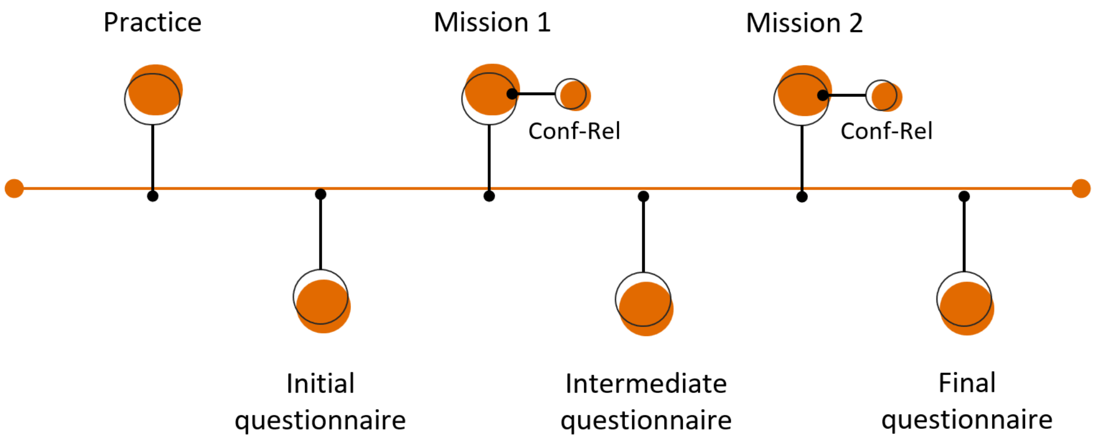


Figure 3.2: Overview experiment procedure

3.1 Experiment walkthrough

The interface in the experiment consisted of a main UGV-view to the left, a UAV-view in the upper right corner and a map in the lower right corner. The experiment was conducted such that while the UGV drove forward, the UAV operated ahead of the UGV. This means that the participants could first see a potential obstacle in the UAV-view, as in Figure 3.3, and then it would appear in the UGV-view a few seconds later when it arrived.

Figure 3.3: Screen interface

The UGV would always proceed to stop in front of the potential obstacle. In the A-LOA mission, the UGV would stop only a few seconds before continuing. Whereas, in the AA-LOA mission, the system would send a TOR. In the TOR, the UAV-view switched to a UAV-Recap, i.e. a photo of how it looked when the UAV was there earlier, see Figure 3.4, which would either match the UGV information or not. The participant then had to choose either "Continue Autonomously" to let the UGV continue its intended route demarcated by the blue line, or "Take over" to control it manually. The participants would never actually have to control it manually in this experiment as the scenario ended right after the decision was made. Then the two questions about confidence and reliability regarding this video were asked.

Figure 3.4: Take Over Request with UAV-Recap

The two different scenarios included in both LOA missions were:

- A truck occasionally obstructs the route.
- Logs stacked on either side of the road, occasionally blocking the roadway.

3.1.1 Truck scenario

The truck scenario was based on a narrow passage with a large white truck that was either obstructing the UGV's path or was parked to the right of the narrow passage. In the Non-Conflic condition, the UGV and UAV exhibited congruent information of a free passage, see TOR in Figure 3.5. In the Conflict condition, the UGV-view was unchanged, but the UAV-view showed the truck obstructing the way (see Figure 3.6). The Hinder condition, which was not used for analysis, can be seen in Appendix D. The difference between the UAV-Recaps is highlighted in Figure 3.7 and 3.8.

Figure 3.5: TOR - Truck Non-Conflict condition

Figure 3.6: TOR - Truck Conflict condition

Figure 3.7: UAV-view Truck Non-Conflict

Figure 3.8: UAV-view Truck Conflict

3.1.2 Logs scenario

The logs scenario was based on a narrow passage with logs on both sides of the UGVs intended path, where the logs may have fallen down on the path. Similar to the truck scenario, this scenario consisted of three different conditions. The first case was Non-Conflic, showing congruent information in the UGV-view and UAV-view (see Figure 3.9). The second case was Conflict, showing the same UGV-view as the Non-Conflict scenario, but with conflicting information on the UAV-view (see Figure 3.10). The Hinder condition can be seen in Appendix D, and the difference between the UAV-Recaps are shown in Figure 3.11 and 3.12.

Figure 3.9: TOR - Logs Non-Conflict condition

Figure 3.10: TOR - Logs Conflict condition

Figure 3.11: UAV-view Logs Non-Conflict

Figure 3.12: UAV-view Logs Conflict

3.2 Experimental setup

The experimental setup, see Figure 3.13, included a keyboard, eye tracking glasses, and a screen that mirrored the display of a laptop on which the experiments were conducted in a web browser. The keyboard was used for making decisions and answering questions during the experiment. The participants were seated in a chair with the armrests positioned just below the edge of the table. The Pupil-Labs Invisible eye-tracking glasses were connected to a OnePlus 8 and the participants' eyes were approximately 77 cm from the screen. Four squares were positioned at the edges of the screen and used for calibration of the eye-tracking glasses. Participants were asked to fixate on each square one at a time during the calibration procedure. The screen was 59.5x33.5x cm, see Figure 3.14.

Figure 3.13: Experiment setup

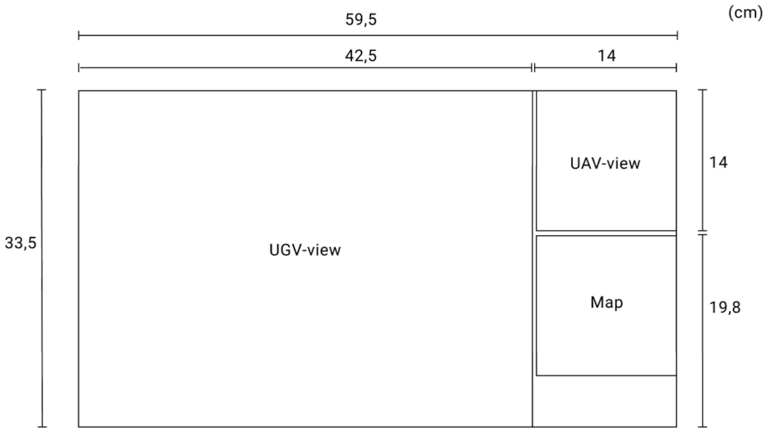


Figure 3.14: Screen measurements

3.3 Group structure

The experiment consisted of four different experimental groups (A1, A2, B1, and B2), see Table 3.1, defined by the combination of condition order (A/B) and LOA order (1/2). The participants were randomly assigned to groups.

Table 3.1: Overview of experimental groups and conditions

Group	**Condition Order**	**System Automation Level Order**
A1	Starting with Conflict	Autonomous → Assisted
A2	Starting with Conflict	Assisted → Autonomous
B1	Starting with Non-conflict	Autonomous → Assisted
B2	Starting with Non-conflict	Assisted → Autonomous

Experimental Group A started with two Conflict conditions, while Experimental Group B started with two Non-Conflicting condition (see Figure 3.15). In addition, Group A1 and B1 started with the A-LOA system, whereas Group A2 and B2 started with the AA-LOA system.

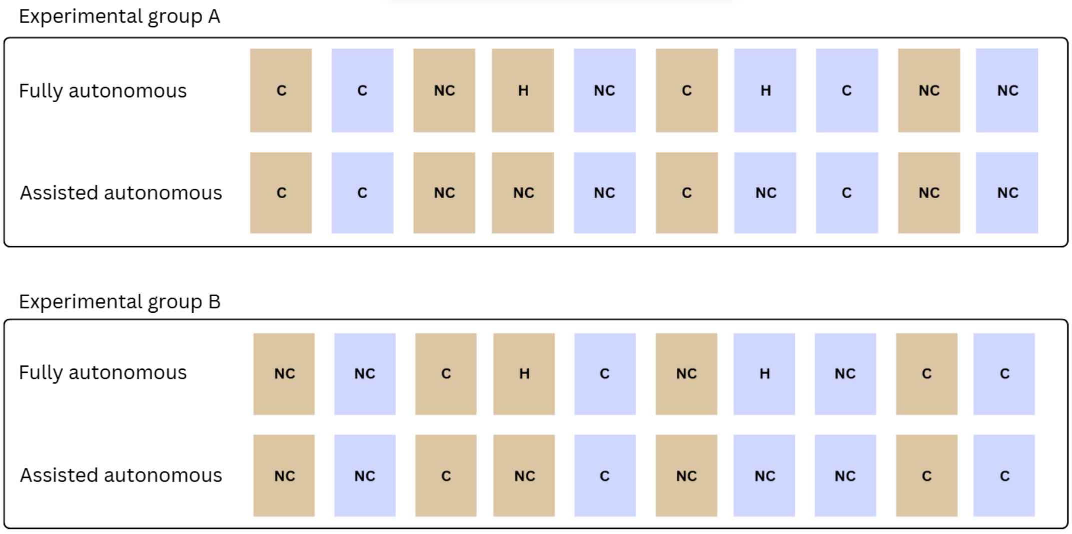


Figure 3.15: Experimental structure across conditions A1, A2, B1, and B2

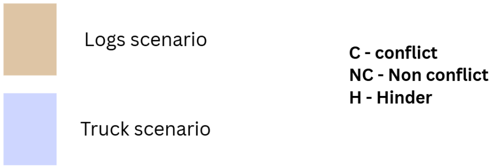


Figure 3.16: Contexts

3.4 Experiment procedure

Upon arrival, participants were provided with two identical consent forms (See Appendix E). Participants signed both forms, keeping one copy themselves, while the researchers retained the other. The participant's ID (01-33) was written on the participant's copy, while the researchers' copy remained anonymous and was not linked to individual results. Participants were informed that they could contact the researchers at any time after the experiment to request removal of their data, using their participant ID.

Before the experiment began, participants received a short introduction explaining the experimental setup and their role in the task. They were told that the system included a UGV, a UAV and that they would serve as the human "operator". Two mediating objects, see Figure 3.17, were used to show the participants how the UAV moves ahead and acts as a reconnaissance sensor for the UGV. Participants were informed that the system was designed to start driving autonomously, but that in some situations, the system would make decisions independently, while in other situations it might require assistance from the "operator". The UGV would then send a take-over request to the operator, with two options. The participants were told that they could either choose to take over and operate manually or let the UGV continue autonomously. A practice session was conducted to familiarize participants with the interface and to demonstrate how to choose "Continue Autonomously" or "Take-over" with the keyboard. The practice showed two take-over situations. It was explained that time would not be measured during the practice, but that it would be measured during the actual experiment. Participants were also encouraged to ask questions at any time if anything was unclear.

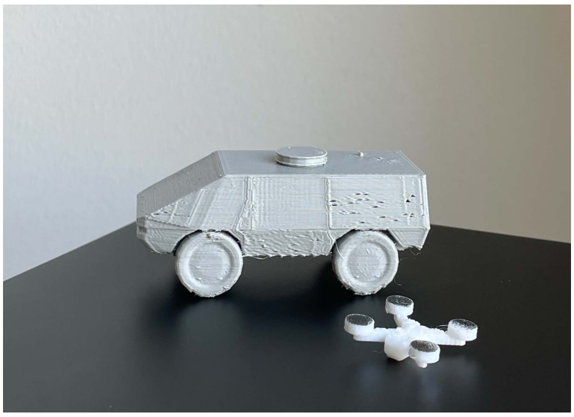


Figure 3.17: UGV and UAV mediating objects

When starting an experiment, the participant ID was filled in, see Figure 3.18. Then the experiment started with the initial questionnaire (Appendix F), including questions about demographics, which were completed by the researcher asking the questions and using the computer mouse. The demographic questions included the participant's age, gender, information about their driving license and driving habits, gaming experience and military experience.

The eye-tracking glasses were subsequently fitted and calibrated before the video simulations were presented. Along with this, the participants were asked to move the chair armrests to the edge of the table and to sit parallel to the screen. They were also instructed to avoid turning or tilting their heads and to remain relatively still so as to not compromise the eye-tracking data.

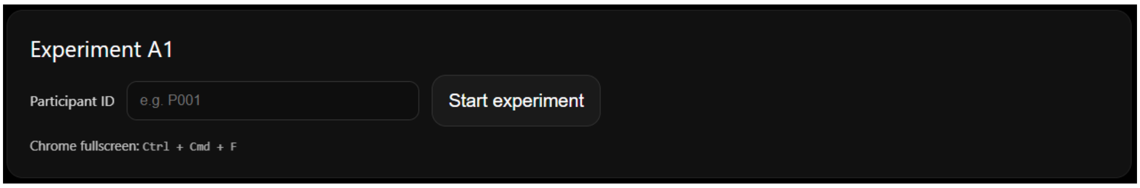


Figure 3.18: Participant ID Screen

The subsequent questions were answered by the participants themselves using the keyboard. Instructions for entering responses were displayed on the screen, and at least one researcher was present throughout the session to address any questions that might arise. After completing four questions regarding their sense of direction, participants proceeded to a calibration screen (see Appendix J) which was used to verify the gaze data. This calibration screen was repeated at the end of the experiment and was used for post hoc adjustment of gaze data.

Next, either the A-LOA mission or the AA-LOA mission began, depending on the experimental group, and the videos were shown sequentially. Figure 3.19 shows what the environment looked like when the UGV was driving. At the end of each video, the participant faced a scenario where the UGV stopped, and either sent a TOR (if AA-LOA) or continued (if A-LOA). Each scenario was followed by a decision questionnaire about how confident they were with the decision (1-7) and how reliable they perceived the system to be (1-7), see the questions in Appendix G.

Figure 3.19: Screen interface

After the first mission, an intermediate questionnaire was presented with questions regarding their perceived mental workload, how responsible they felt for the outcome of the mission, and how they would describe their trust in the system (see Appendix H). Each mission was approximately 10 minutes long.

The second mission then started, with the other LOA. To clearly communicate the two different missions to the participant, a "Mission started" and a "Mission completed" image were shown before and after each mission (see Appendix K). After the second mission, a final questionnaire was presented (see Appendix I). The eye-tracking recording was stopped and the glasses were then removed. The final questionnaire included the same questions as in the intermediate questionnaire, and additional questions assessing how reliable participants perceived the UGV and UAV information to be, their situational awareness (perception, comprehension, and projection) and how much they rated the different agents' contributions to solving the situation in terms of perception and decision-making.

Lastly, participants answered three questions in free-text regarding how helpful they found the UAV information when making decisions, whether they would have preferred a system without the UAV-view, and why they chose to look at it. Additional observational notes were taken by a researcher when participants commented on aspects of the experiment during the session. A discussion often arose after the experiments, which was documented in the notes. The participants were then thanked and compensated with a cinema ticket.

3.5 Participants

In total, there were 33 participants, divided into experiments A1, A2, B1, and B2 (see Table 1). Participants were randomly assigned to the experimental groups.

Condition	**Number of Participants**
A1	8
A2	8
B1	8
B2	9
Total	**33**

Table 3.2: Distribution of participants across experimental conditions.

A participant overview summarizing the responses to the background questions in the experiment is presented in Figure 3.20. The 33 participants ranged in age from 19 to 58 years, with a mean age of 27.7 years. There were in total 19 men, 13 women, and one non-binary person. Regarding driving frequency, 16 participants reported never or rarely driving, while 17 reported driving between monthly and daily. Most participants had held a driving license for 1-10 years, seven participants did not have a driver's license, and five had held a license for more than 10 years.

Most participants (27) reported no military experience, whereas six participants had either completed basic military training, were currently active, or had previous military experience. In terms of gaming experience, 18 participants identified as gamers, ranging from casual to hardcore gamers, while 15 participants described themselves as novice or non-gamers.

The participants answered a set of four questions regarding their sense of direction and spatial navigation abilities on a scale of 1 – 7, with 19 participants ranking in the high group and 14 in the low.

- Q1: My sense of direction is very good.
- Q2: I am very good at reading maps.
- Q3: I can usually remember a new route after I have traveled it only once.
- Q4: I have a very good "mental map" of my environment.

Table 3.3 shows the definitions for each demographic group used for the analysis.

Table 3.3: Definitions of Demographic Factors

Factor	**Definition**
Gaming Experience	**Gamer**: Casual, Core, Hardcore **Non-Gamer**: Never, Novice
Driving Experience	**High driving frequency**: Monthly, Weekly, Daily **Low driving frequency**: Never, Rarely
Sense of Direction	**High sense of direction**: Avg. SoD ≥ 5 **Low sense of direction**: Avg. SoD < 5
Military Experience[1]	**No Military Experience**: None **Military Experience**: Other

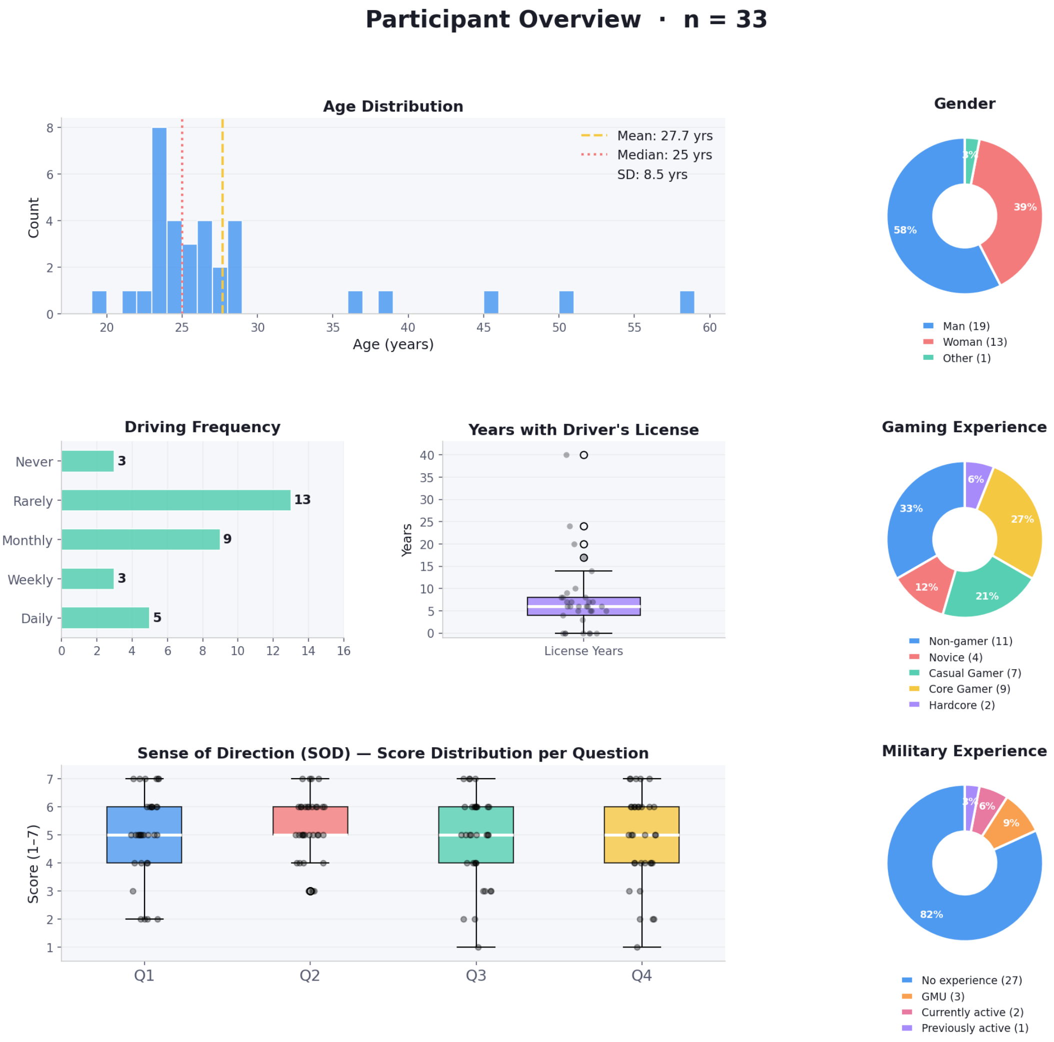


Figure 3.20: Demographics

3.6 Eye-tracking

Eye-tracking was performed using Pupil Invisible glasses from Pupil Labs, see Figure 3.21. These were connected to a OnePlus 8, which was used to calibrate the eye-tracking glasses, record videos, and store both the videos and gaze data. During the calibration, the participants were instructed to avoid horizontal head movements and to keep their heads as still as possible throughout the experiment. To calibrate the glasses, they were asked to look at the center of each marker located in the corners of the screen, while the researcher adjusted a calibration circle to align the estimated gaze position from the eye-tracking glasses with the participants' actual point of focus.

After approximately half of the experiments had been conducted, an additional eye-tracking validation procedure was introduced to verify the accuracy of the gaze data. This validation consisted of a black screen displaying four numbers, presented both before and after the experiment, and each participant was instructed to look at each number. The purpose was to make it possible to detect whether the gaze coordinates had drifted over time. This procedure was introduced starting from participant 16.

At a later stage, for participants 18-33, a revised validation screen was implemented containing five numbers on a black background. These numbers were positioned at locations corresponding to the corner of the UAV-view, the corner of the map, the center of the UGV view, and two additional positions within the UGV view. This modification was introduced to make it easier to determine exactly where the participants were looking on the different displays. See Appendix J for both versions, including one with how the numbers were placed in relation to the windows on the interface.

Figure 3.21: Eye-tracking Set-Up

4

Results

The results include how different levels of UGV autonomy affect the operator state and performance, to what extent the operators rely on UAV visual information when making control handover decisions, the effect of demographic factors, and how functions can be appropriately allocated between the human operator, UGV, and UAV. Operator state was assessed through measures of perceived responsibility, mental workload, confidence, and reliability. Operator performance was assessed through measures of decision time and choices.

4.1 Effects of Automation level on Operator State

This section presents the quantitative results pertaining to the following questions:

To what extent is operators state affected by UGV autonomy level?

To what extent do demographic factors play a role in operator state?

4.1.1 Mental workload & Perceived Responsibility

The Mental Workload scores for the Autonomous and Assisted-Autonomous systems is presented as box plot comparison of the NASA-TLX 20-point scale for each LOA, shown in Figure 4.1. The question can be seen in Appendix H. The data is treated as within-subjects and the plots show the median, spread, interquartile range (IQR), and outliers. The mean for the Autonomous system was 6.46, and 6.27 for the Assisted-Autonomous system. A Wilcoxon signed-rank test showed no significant difference in perceived Mental Workload between A-LOA and AA-LOA ($p = 0.952$).

The Perceived Responsibility is presented as a box plot comparison of the self reported 7-point Likert scale within each LOA, and can be seen in Figure 4.2. The mean for the A-LOA was 2.94 and 3.85 for the AA-LOA. It is clear that participants experienced a significantly higher sense of responsibility in the AA-LOA, which is expected as they were actively involved in the decision-making process. The results were statistically significant according to the Wilcoxon test ($p < 0.001$).

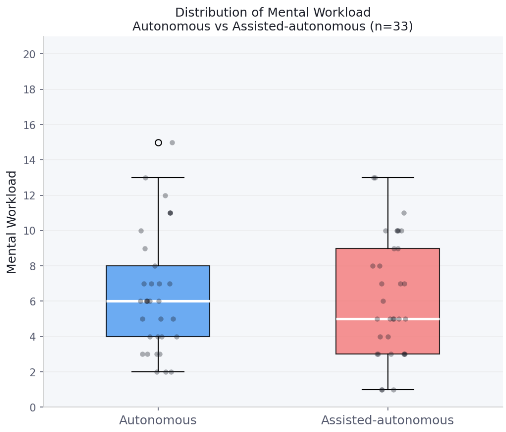


Figure 4.1: Mental workload

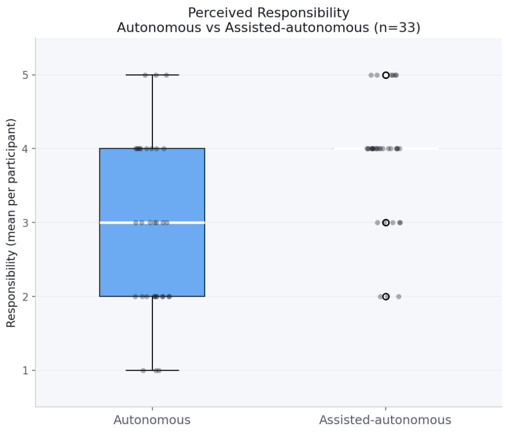


Figure 4.2: Perceived Responsibility

The pattern of increased Perceived Responsibility in AA-LOA was persistent within each demographic factor and no factor stood out significantly from the rest.
One interesting finding in analyzing the demographic factors is that Gamers rated the Mental Workload lower by $\approx 2-3$ points, or 13.5 percentage points, in comparison to Non-Gamers ($p = 0.022$), shown in Figure 4.3.

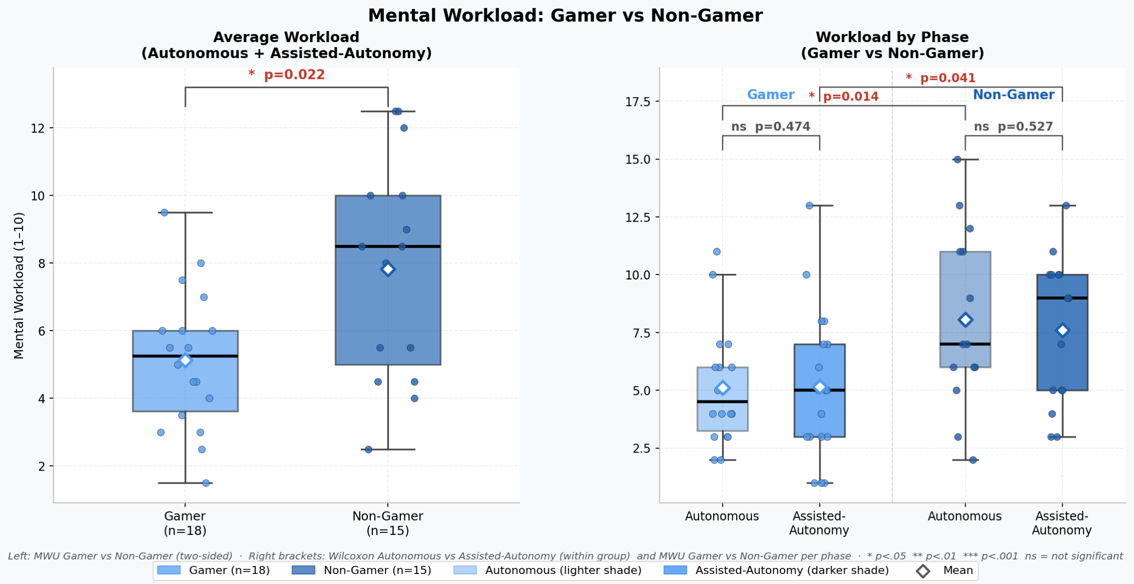


Figure 4.3: Mental Workload - Gamers vs Non-Gamers

4.1.2 Confidence & Perceived Reliability

Confidence and Perceived Reliability was scored on a 7-point Likert scale after each scenario and is visualized as box plot comparisons, see Figure 4.4. The questions were *How confident are you with the decision?* and *How reliable was the system (UGV-UAV) in this scenario*, which can be seen in Appendix G.
Confidence level showed no significant differences between LOA. The mean score was 6.109 and 6.186 for A-LOA and AA-LOA respectively, this increase is likely due to noise ($p = 0.8198$).

Perceived Reliability scores strongly indicated that participants felt that the A-LOA was more reliable than the AA-LOA ($p = 0.0019$). The mean Perceived Reliability was 5.148 and 4.655 respectively.
Comparing the Perceived Reliability for demographic factors between LOA showed similar results to scores presented previously in Figure 4.4, with a significant decrease within the AA-LOA. The exception to this was the Low Driving Frequency and Low SoD groups, showing no significance between conditions, scoring both levels similarly ≈ 4.5, see Figure 4.5.

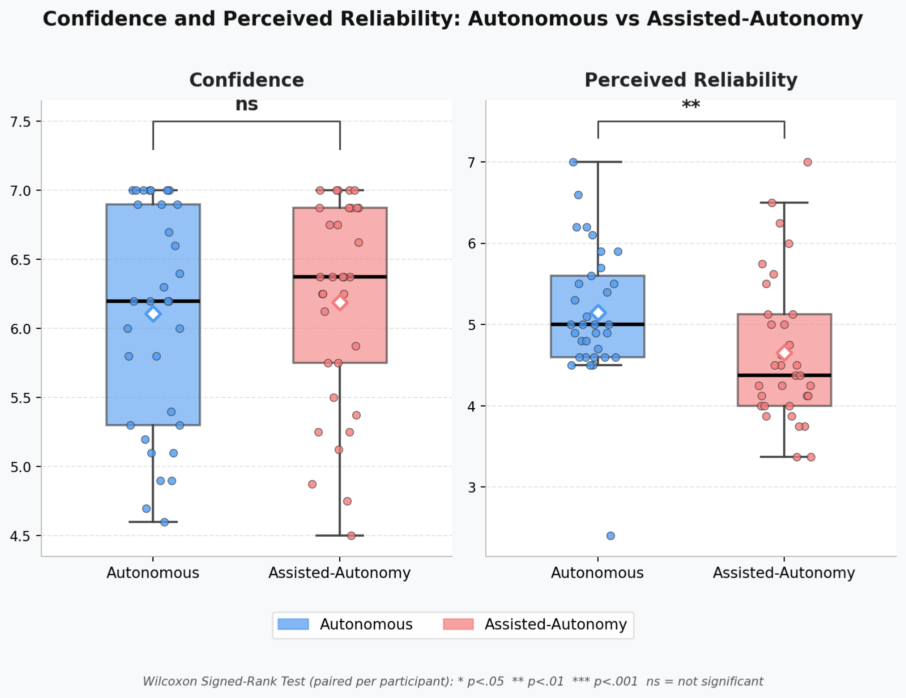


Figure 4.4: Confidence and reliability

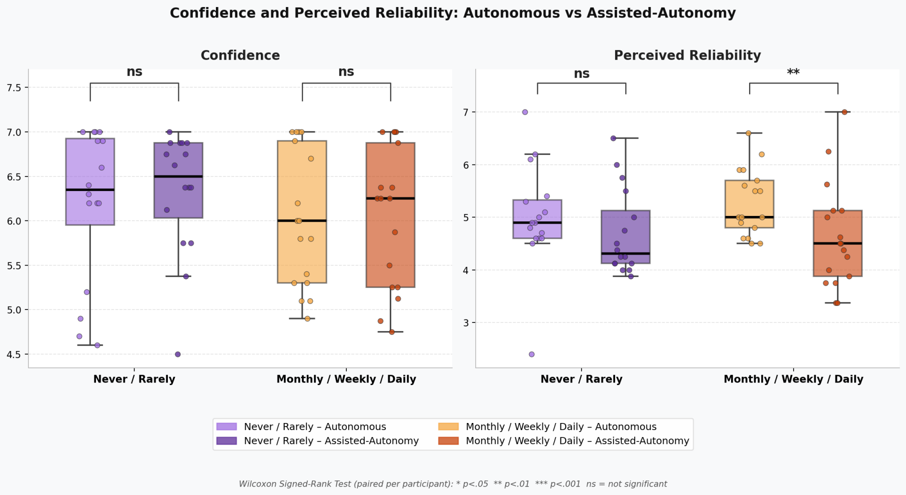


Figure 4.5: Confidence and Reliability: Driving

4.1.3 Summary: LOA Effect on State

Mental Workload showed no significant difference between LOA, except for Gamers consistently rating their workload 13.5 percentage points lower than Non-Gamers.

Perceived Responsibility results, strongly support the notion that operators feel more responsible in AA-LOA when they have to make decisions.

Confidence scores indicate that operators feel similarly about the decisions being made in both LOA.

Perceived Reliability indicates that operators consistently perceive A-LOA as more reliable then AA-LOA.

4.2 Effects of UAV Integration on Operator State and Performance

The subsequent part of the analysis pertaining to the operator state and performance introduces the additional condition of Conflict or Non-Conflict, which will be the basis of comparison within the AA-LOA. The A-LOA does not allow for operator decisions and as such, can not be used for performance comparisons.

4.2.1 Confidence & Perceived Reliability

Splitting the Assisted-Automation condition into the Conflict and Non-Conflict conditions shows a significant number of participants reporting higher Confidence ($p = 0.0067$) in their own decisions in the Conflict condition compared to when the system made the decision, however this trend seems reverse in the Non-Conflict condition ($p = 0.070$), see left side of Figure 4.6, and Table 4.1.

Table 4.1: Confidence - Mean and Standard Deviation

Confidence	NC - Auto	NC - Assisted	C - Auto	C - Assisted
Mean	6.591	6.364	5.386	6.015
Std.	0.509	0.593	1.684	1.004

There is negligible difference in Perceived Reliability within the two conditions ($NC : p = 0.1733, C : p = 0.7698$), however the effect between LOA still remains, as the score for both conditions within the levels are consistent in being higher and lower respectively, see Table 4.2. The mean score between the conditions drops from ≈ 6.5 to ≈ 3.0, see right side of Figure 4.6 .

Table 4.2: Perceived Reliability - Mean and Standard Deviation

Perceived Reliability	NC - Auto	NC - Assisted	C - Auto	C - Assisted
Mean	6.576	6.379	3.008	2.932
Std.	0.748	0.875	1.681	1.650

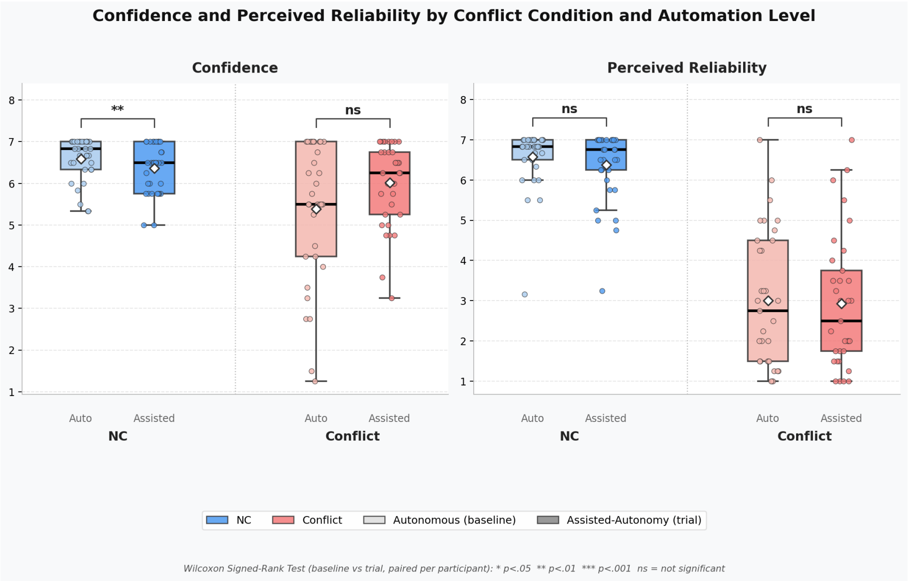


Figure 4.6: Confidence and reliability: C/NC

When introducing the NC/C conditions, High Driving Frequency (HDF) showed a significant effect on Confidence, see Figure 4.7. In the Non-Conflict condition, HDF tended to feel more confident in the decision making of the Autonomous system ($p = 0.0250$), whereas in the Conflict condition HDF placed more confidence in their own decision making ($p = 0.0284$).
A similar pattern was observed for participants with a High SoD, showing significant differences between conditions: higher reliance on automation in the Non-Conflict condition and increased confidence in their own judgment during Conflict condition.

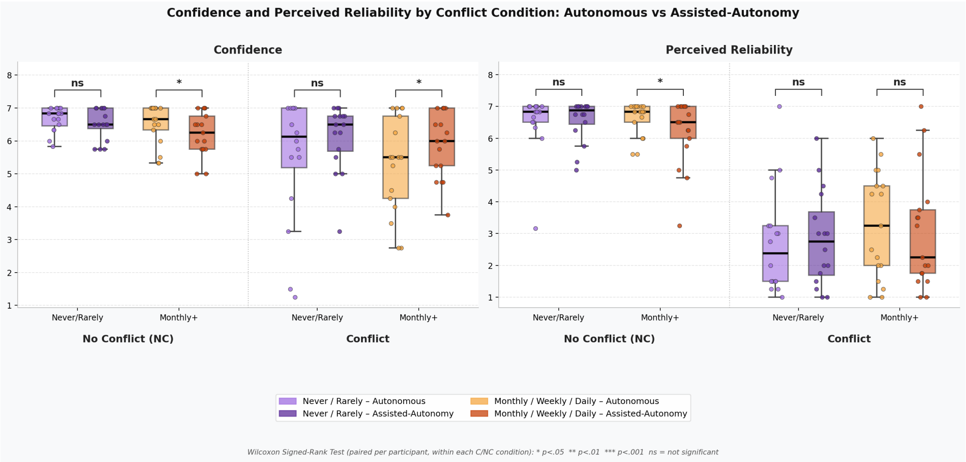


Figure 4.7: Confidence and Reliability: Driving - C/NC

4.2.2 Decision Time & Take-Over Probability

Decision Time is analyzed with a mixed-effects model clustered on participant id, presented in a box plot comparison in Figure 4.8 and contrary to expectations, there is no significant increase in Decision Time in Conflict over Non-Conflict condition. Take-Over Rate in the Conflict condition is substantially higher (31.8%) compared to the Non-Conflict condition (7.6 %). According to the mixed-effects model, the difference between conditions is significant ($p < 0.001$), indicating that participants were more likely to intervene when presented with conflicting information.

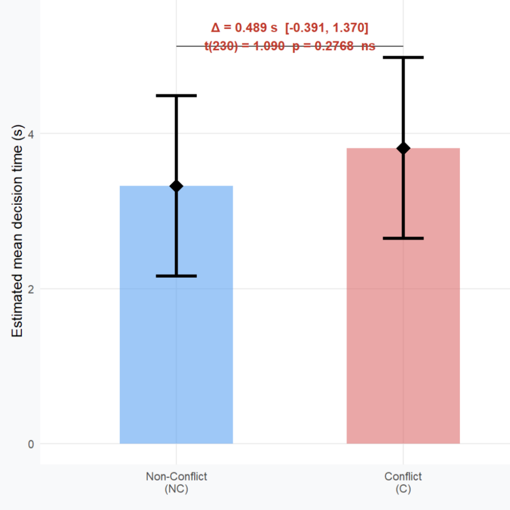


Figure 4.8: Decision Time

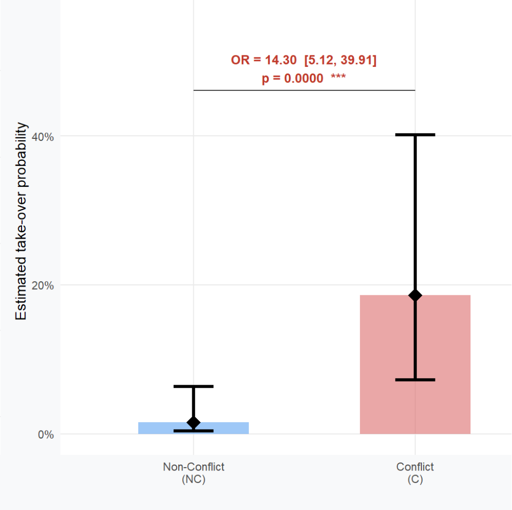


Figure 4.9: Decision Choice

Comparing the average time of each demographic factor against the average of all participants combined, shows that no group stands out a statistically significant amount, however the pattern is that Gaming, Low Driving Freq. and Low SoD trend toward faster decisions, while Non-Gamers, High Driving Freq. and High SoD trend toward slower decisions, as seen in Table 4.3.

Further analysis of NC/C condition did not yield any significant results.

Table 4.3: Decision Time - All vs Groups

	All	Gamers	Non-Gamers	High Driving Freq.	Low Driving Freq.	High SoD	Low SoD
Mean	3.569s	2.881s	4.395s	4.589s	2.485s	4.186s	2.731s
Std.	3.058	1.807	4.008	3.886	1.185	3.624	1.876
p-value	-	0.4361	0.3794	0.2642	0.2451	0.2828	0.1886
Time dif.	0	-0.688s	+0.825s	+1.020s	-1.084s	+0.617s	-0.838s

Comparing the Decision Times by demographic factors using mixed-effects model shows High Driving Frequency as being the only significant predictor ($+2.05s, p = 0.048$), while Gaming trends toward faster ($-1.77s, p = 0.091$) and High Sense of Direction toward slower ($+1.79s, p = 0.088$), see Figure 4.10.

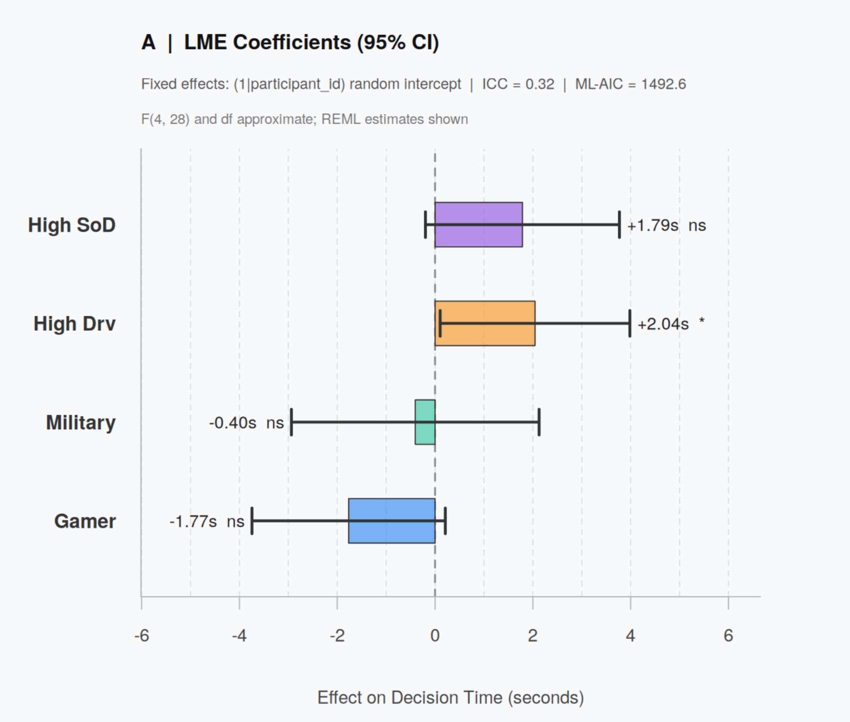


Figure 4.10: Decision Time - Group Effects

Further analysis comparing the probability of Take-Over between all participants and the demographic factors showed that no individual factor has a significant impact on Take-Over likelihood, however participants who reported a Low SoD seem to be more cautious of intervening in Conflict scenario ($p = 0.151$).

4.2.3 UAV Gaze Share

Gaze data from all participants was filtered by id, condition, and area of interest (AOI), and combined for subsequent analysis. The combined AOI gaze share is presented in a stacked area plot showing a steady AOI distribution for the first 20 seconds of the scenarios, which then become more UGV dominated as the truck or logs comes into view, followed by distinct UAV spikes in blue, a few seconds before the TOR, see Figure 4.11.

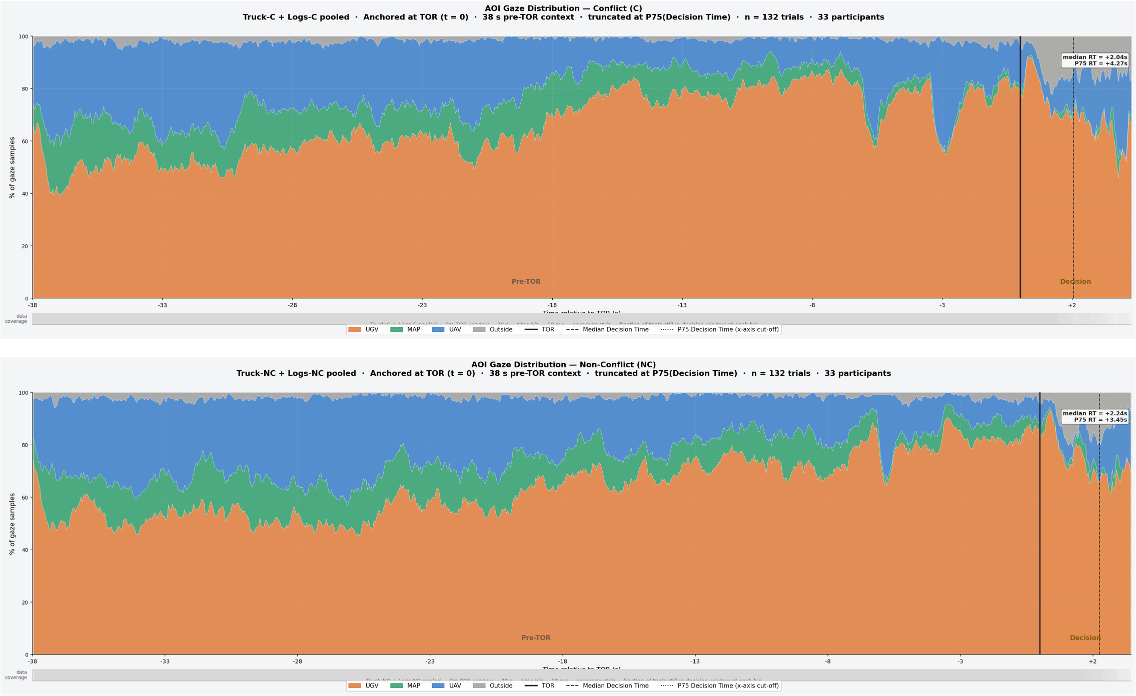


Figure 4.11: AOI Gaze Share - C/NC

Analyzing the UAV gaze share across the full videos of the two conditions using a linear mixed-effects model, shows that the average is 21.56% and 21.57% for C and NC respectively with outliers in both directions, shown in 4.12.

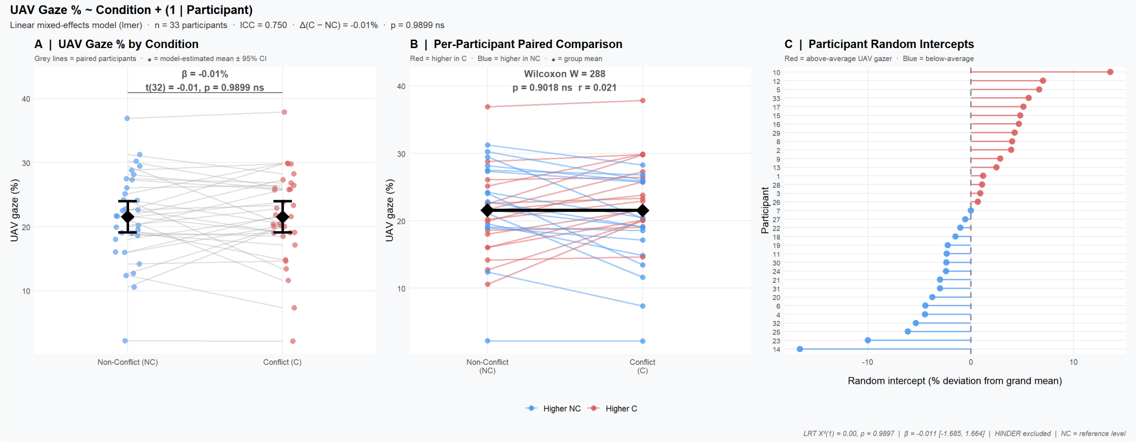
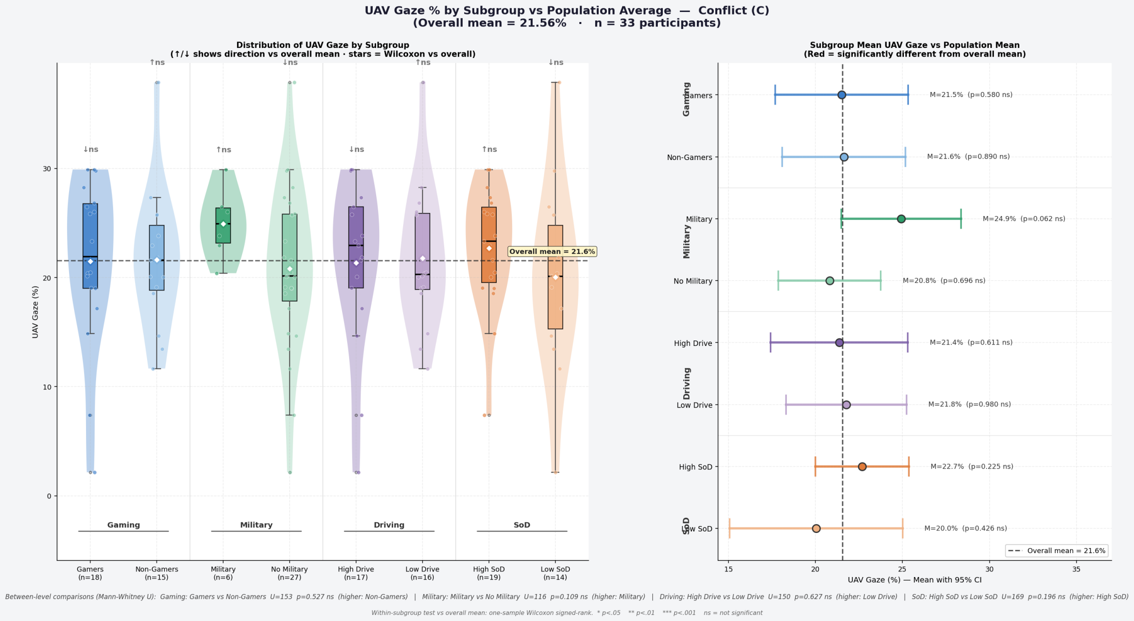


Figure 4.12: UAV Gaze Share

Comparing the groups UAV share to the average, as seen in Figure 4.13 and 4.14, is mostly non significant[1]. However, an interesting effect emerges in the Driving group where they share a similar UAV distribution during Conflict but diverge significantly during No-Conflict ($p = 0.050$) with the High Driving Frequency group using the UAV less, and the Low Driving Frequency using it more.


UAV Gaze % by Subgroup vs Population Average — Conflict (C)
(Overall mean = 21.56% · n = 33 participants)

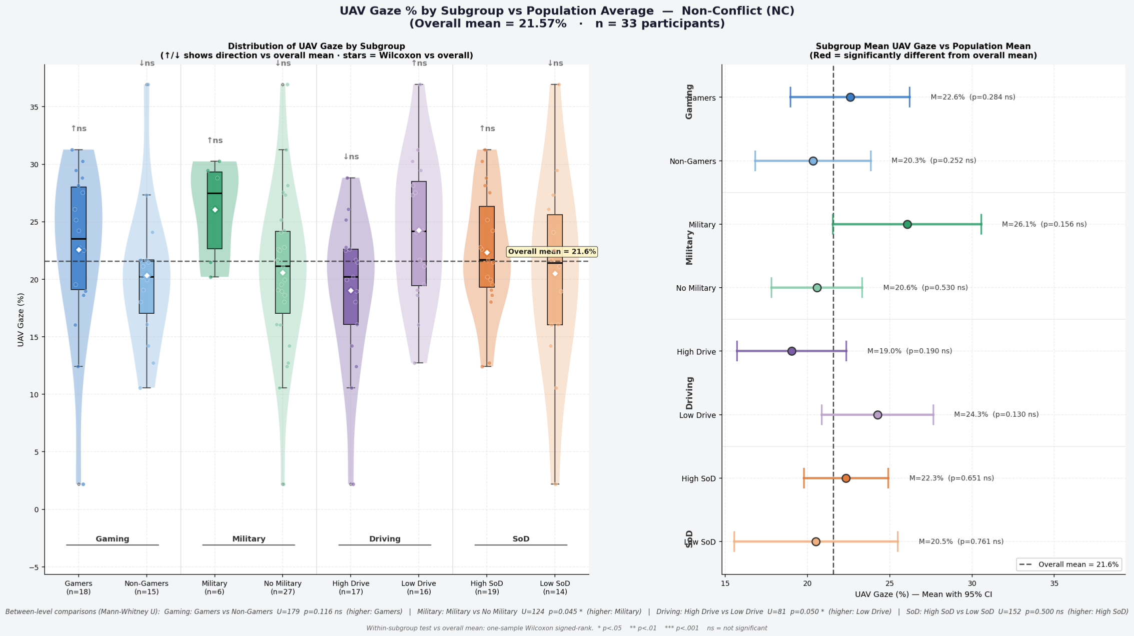
Distribution of UAV Gaze by Subgroup
(↑/↓ shows direction vs overall mean · stars = Wilcoxon vs overall)

Subgroup Mean UAV Gaze vs Population Mean
(Red = significantly different from overall mean)

Between-level comparisons (Mann-Whitney U): Gaming: Gamers vs Non-Gamers U=153 p=0.527 ns (higher: Non-Gamers) | Military: Military vs No Military U=116 p=0.109 ns (higher: Military) | Driving: High Drive vs Low Drive U=150 p=0.627 ns (higher: Low Drive) | SoD: High SoD vs Low SoD U=169 p=0.196 ns (higher: High SoD)

*Within-subgroup test vs overall mean: one-sample Wilcoxon signed-rank. * p<.05 ** p<.01 *** p<.001 ns = not significant*

Figure 4.13: UAV Share - Groups - C


UAV Gaze % by Subgroup vs Population Average — Non-Conflict (NC)
(Overall mean = 21.57% · n = 33 participants)

Distribution of UAV Gaze by Subgroup
(↑/↓ shows direction vs overall mean · stars = Wilcoxon vs overall)

Subgroup Mean UAV Gaze vs Population Mean
(Red = significantly different from overall mean)

*Between-level comparisons (Mann-Whitney U): Gaming: Gamers vs Non-Gamers U=179 p=0.116 ns (higher: Gamers) | Military: Military vs No Military U=124 p=0.045 * (higher: Military) | Driving: High Drive vs Low Drive U=81 p=0.050 * (higher: Low Drive) | SoD: High SoD vs Low SoD U=152 p=0.500 ns (higher: High SoD)*

*Within-subgroup test vs overall mean: one-sample Wilcoxon signed-rank. * p<.05 ** p<.01 *** p<.001 ns = not significant*

Figure 4.14: UAV Share - Groups - NC

---

[1] Note: Military, who despite being a small group (6) stand out in both conditions.

4.2.4 UAV Share in Decision-Making

Splitting the conditions by scenario reveals that the spikes are a result of the UAV-Recap (explained in Section 3.1) showing at different times before the TOR for each scenario, and being completely absent from the Truck-NC scenario, as seen in Figure 4.15. The potential effect of which is discussed in Section 6.2.2.

It can therefore be argued that the decision window is not simply from TOR → Decision. To account for this discrepancy in pre-TOR context, decision windows were extended separately for the two scenarios, Truck and Log, +4s and +7s respectively.

Figure 4.15 highlights the difference of UAV share within the No-Conflict condition between the Truck and Log scenario after the TOR. This is likely a result of Logs showing the UAV recap a few seconds before the TOR, and Truck not.

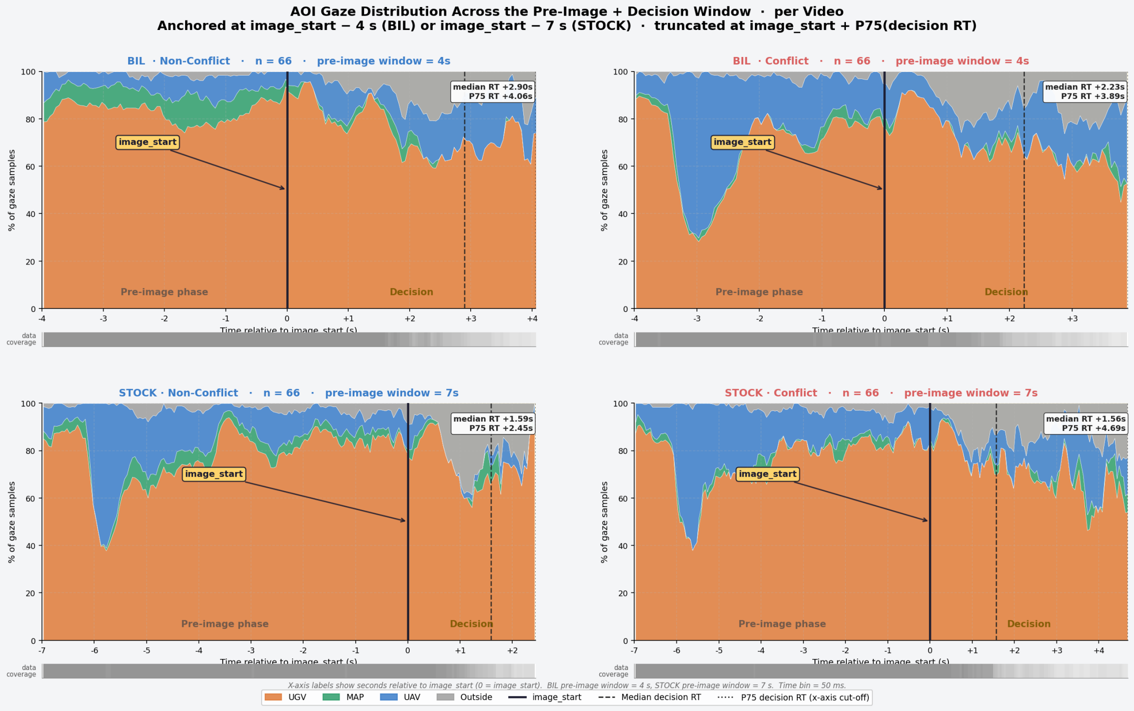


Figure 4.15: AOI Gaze Share - Extended Decision Windows

Focusing in on the TOR→ Decision window for each condition shows that the UAV share drops significantly after the TOR, as most operators rely on the UGV view, see Figure 4.16.[2] As expected, the UAV share is more prominent in the Conflict condition as operators deal with the conflicting information.

---

[2] Note: The Dot-Dash line represents 75th percentile of decisions, data to the right of that is considered noisy.

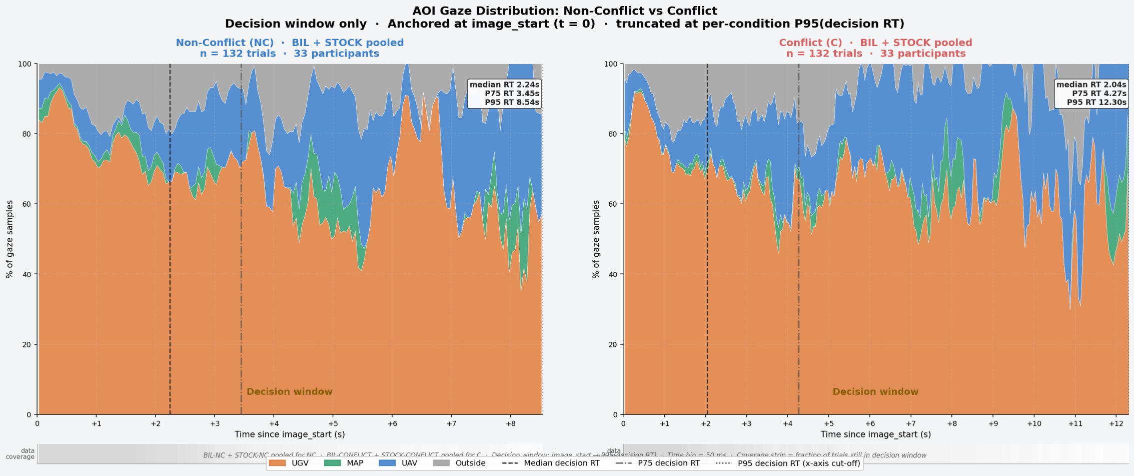


Figure 4.16: AOI Gaze Share - Decision Window - NC/C

Comparison of the UAV gaze share was done for both the TOR→ Decision window and the extended TOR + 7s → Decision window using Mann-Whitney U test[3]. Results demonstrate a significant difference between C and NC, only when including the pre-TOR context of the extended decision window, see Figure 4.17.

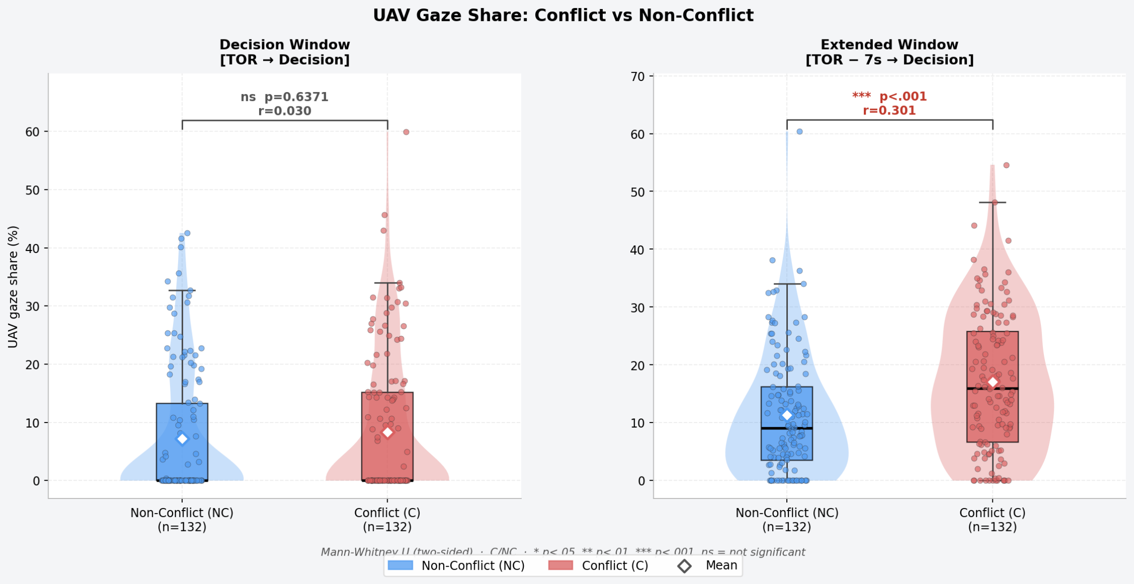


Figure 4.17: UAV Share in Decision - C/NC - Short/Extended Window

---

[3] Mann-Whitney U test assumes independent data, each participant has 4 data points for C and NC respectively.

The UAV usage during the extended decision windows (Truck/Logs = +4s/+7s) was therefore analyzed separately for each scenario[4]. Figure 4.18 clearly shows a difference in UAV usage in the Truck scenario as a result of the aforementioned UAV-recap discrepancy. However, it also shows a significant difference in the "cleaner" Logs scenario, supporting the idea that C condition increases UAV gaze share during decision-making[5].

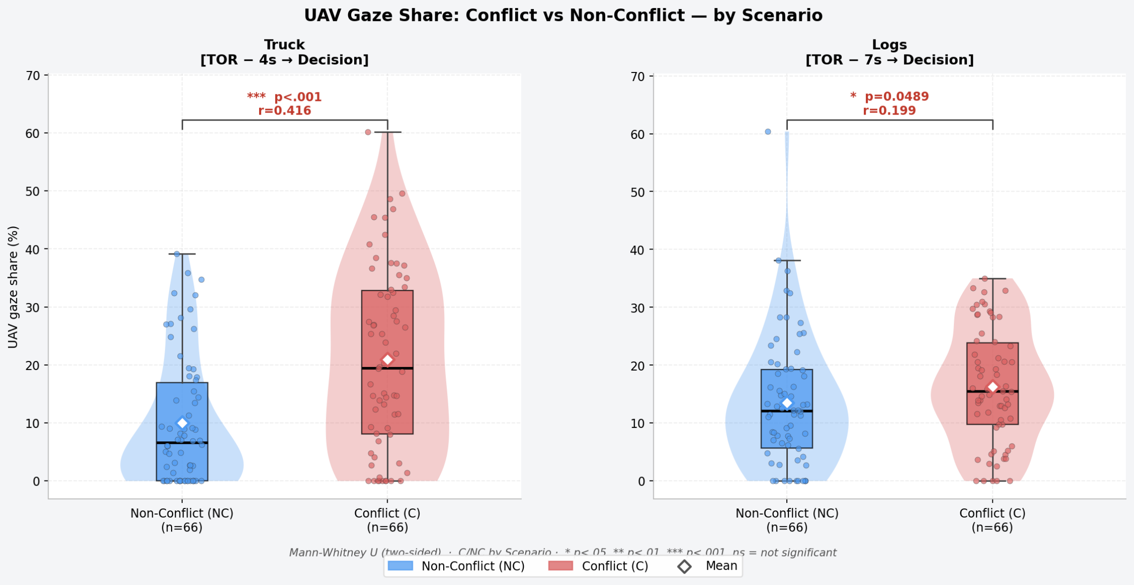


Figure 4.18: UAV Share in Decision - C/NC - by Scenario

Further analysis of UAV Share during Decison-Making yielded little result, see Appendix O.

---

[4] Mann-Whitney U test assumes independent data, each participant has 2 data points for C and NC respectively.

[5] Note: Splitting the conditions by scenarios weakens the power of the results.

4.2.5 Summary: UAV Effect on State & Performance

State

Confidence in decisions in C/NC differs between demographic factors. Specifically, High Driving Frequency and High SoD rate NC significantly higher, and C significantly lower, while the respective Low groups show little to no difference between C/NC.

Perceived Reliability indicates that operators consistently perceive A-LOA as more reliable then AA-LOA within each condition. As with Confidence, Low Driving Frequency and Low SoD groups rate both C and NC similarly high, in contrast to other groups who rate NC higher and C lower respectively.

Performance

Decision Time showed no significant increase in Conflict condition, however it points to certain demographic trends. Namely: Gaming, Low SoD and Low Driving Frequency trend toward faster decisions, while Non-Gamers, High SoD, and especially High Driving Frequency trend toward slower decisions.

Take-Over rate is substantially higher in C (31.8%) compared to NC (7.6%) and no demographic factor showed significant effect.

UAV Gaze Share

The gaze data reveals similar UAV gaze share in both conditions, (C: 21.56%, NC: 21.57%). Demographic factors show no effect in C, with exception to Military, who despite being a small group (6) stands out in both conditions. NC shows a significant demographic split, with High Driving Frequency decreasing UAV share, while Low Driving Frequency increased UAV share.

UAV share drops substantially after TOR, but less so in C where slower decision makers not only spend more time on the UAV, but a larger share of the decision time overall.

Decision-Making

Conflicting information correlates to the UAV receiving an increased share of attention during decision-making.

4.3 Human-Autonomy collaboration

This section addresses the following research question:
How can functions be appropriately allocated between the human operator, the UGV, and the UAV across different phases of the mission?

4.3.1 Situational awareness

Situation awareness addresses how well the participants perceive, understand, and project while monitoring and assisting the system.
The questions related to situational awareness were:
- How easy or difficult was it to perceive all relevant information in the environment?
- How easy or difficult was it to understand how the situation evolves?
- How easy or difficult was it to react appropriately in each situation?

The questions were aimed at measuring the participants' ability to maintain Situational Awareness during the experiment. Figure 4.19 illustrates the participants' subjective ratings on the 7-point Likert scales. The means and median scores remained high across all SA levels, see Table 4.4. However, it should be noted that all responses were self-reported.

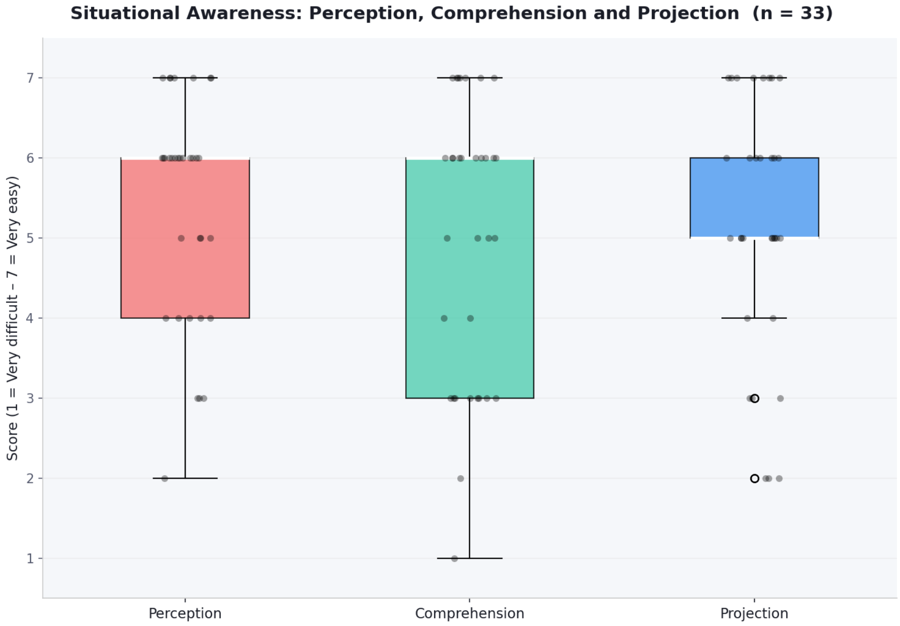


Figure 4.19: Situational awareness

Perception had the highest mean (5.39), indicating that the participants generally found it straightforward to perceive the relevant environmental information presented by the system. Additionally, the IQR was concentrated between 4 and 6.
Comprehension had the highest variability and the lowest mean (4.97), indicating that there was higher variability in participants' perception of understanding how the situations evolved.

Table 4.4: Descriptive statistics for the three levels of situational awareness.

SA Level	M	SD	Median	Min	Max
Perception	5.39	1.39	6	2	7
Comprehension	4.97	1.74	6	1	7
Projection	5.18	1.57	5	2	7

Projection scored a high mean of 5.18. The IQR starts at 5 and extends to 6, indicating that at least 75% of participants found it easy to make decisions and respond appropriately in the situations.

4.3.2 Perceived Reliability in UGV and UAV

The participants' Perceived Reliability of the UGV and UAV is visualized in Figure 4.20. The mean for the UGV was 5.73 and the mean for the UAV was 3.18. The information from the ground vehicle was perceived as more reliable than the information from the UAV and the Wilcoxon signed-rank test ($p = 0.0003$) confirms this. This was expected, as participants were informed that the UAV operates ahead of the UGV, meaning that the UGV consistently represents the current "live" situation.

Figure 4.20: Perceived Reliability in UGV and UAV

4.3.3 Qualitative analysis

The qualitative analysis was based on the KJ-method for the questions at the end of the experiment. The method was used to gather insights and opinions from the participants regarding the UAV use and interface, see Appendix L. Additionally, the notes from the experiment were analyzed and used for further understanding of the participant's thoughts of the system.

The following questions were investigated:
How helpful did you find the information provided by the UAV display in making your decision?

The responses were categorized based on how helpful the participants perceived the UAV-view to be, as shown in Table 4.5.

Table 4.5: Groupings and Distribution of Participants

Groupings	Not helpful	Not very helpful	Helpful sometimes	Helpful	Very helpful
Number of Participants	1	5	5	11	2

11 participants' answers were categorized under "Helpful". The reasons were primarily to be able to plan ahead, be more prepared, get an overview, and be alert and focused in the right moment. Examples are, *"It was helpful in identifying possible obstructions but not helpful at all in determining if intervention was needed."* and *"It was helpful in planning the route and guessing which situations would require intervention."*.

Two participants thought the UAV-view was "Very helpful", with reasons such as being able to scout for possible obstacles or danger, and to be given a heads-up. One example is *"Very helpful, especially the 'recap' the UAV provided before making a decision."*

Only one participant's response was categorized under the “Not helpful at all” group. The response was: *"I more solely trusted the ground vehicle, maybe because it reflects how I would perceive the situation.”*. Five participants' responses were categorized under the “Not very helpful” group. The reasons were either that the UAV-view was not very reliable, or that it was not always the same as the UGV, which they assumed was the correct one. Five participants' answers were categorized under the grouping "Helpful sometimes". An example of this is *"Helpful to an extent, but not for the final decision"*.

Would you have preferred a system without UAV-view? Why or why not?
In response to this question, almost all participants preferred a system with the UAV-view, see Table 4.6, as only three answered to the contrary. There were three main themes for preferring a UAV-view: Pre-planning, Comprehension and Safety.

Table 4.6: Groupings and Number of Participants

Grouping	**Number of Participants**
Yes	3
No (Pre planning)	13
No (Comprehension)	5
No (Safety)	5
No (Decision making)	1

The perceived usefulness for pre-planning was the most common reason why they would be opposed to a system without a UAV-view. Even when they noticed that the information was sometimes in conflict with the UGV information, an example being, *"No, because it is always nice to be prepared for what is coming, no surprises (unless info is wrong). But in this case it was only helpful if correct, if the info was wrong there was no big problems (life or death)"*. Another example is, *"No. Pre-notion was helpful. I could deal with the diverging, inconsistent information."* The advantage of having a UAV-view giving them a heads-up made them feel more prepared, which they considered more important overall.

Another grouping was Comprehension, which consisted of answers from participants who used the UAV for more context or information. An example is, *"I still think the UAV view was a good implementation, especially to get more context... I.e, how the truck moved. if someone had picked up the logs after they fell to know that they were unstable..."*. Other examples are, *"No, I liked to get an overview of the area."* and, *"No, more information is (almost) always better"*.

For some, Safety was the most important factor. For example, *"In a sense no, because it helped me become alert that a possible obstacle was being approached, which I think heightened my alertness and made me focus on the safety of the situation."*. Another example is, *"No. I do see the added benefit of viewing the situation from above and seeing it from two points of view does make it feel safer, in case one system malfunctions."*.

The three participants who expressed a desire for a system without the UAV wrote that the reasons were the misleading information and the resulting confusion, for example, *"Yes, although it provided some useful insights, it caused me to be distracted"*.

Did you look towards the UAV because you found it useful or because you felt unsure about the UAV information

The largest group was "Useful", where participants described how it was useful for having an overview or to be able to see ahead, similar to earlier answers. The "Useful Wary" group included answers where the participants were unsure which information to trust, or that they at times felt uncertain when receiving information that was incorrect. Examples are, *"Mostly useful but after the first time it got it wrong I was more wary of it."* and *"A bit of both. I looked to get an idea of what was up ahead, however once errors started occurring I looked more to see if the UAV provided correct information."*. Only one participant answered "Not Useful", stating, *"I looked to get more time to understand the situation but the information was wrong too often to be useful"*. Lastly, two participants wrote that they looked at the UAV-information because they felt unsure.

Table 4.7: Groupings and Distribution of Participants

Groupings	Useful	Useful wary	Unsure	Not useful
Number of Participants	15	9	2	1

4.3.4 Experiment Notes

During the experiment, notes were taken on the comments the participants made. The experiments were often followed by discussions, which were also noted. The following paragraphs describe the main takeaways.

A participant working professionally with drone development highlighted that allowing users to actively influence or control aspects of the system could help increase engagement and maintain concentration and focus. The participant compared this to surveillance monitoring, where interactive controls such as selecting a view or zooming in on multiple camera feeds can create a stronger sense of involvement than passively observing a fixed display. A suggestion was to have the option to switch between a north-up and heading-up view for the map. Another example is providing operators the option to intervene whenever, instead of having a system that has to initiate a TOR in order for the operator to take manual control.

Two participants expressed confusion regarding the size of the vehicle. One was worried that it would not fit through the stones in the Truck scenarios, and the other mentioned that they were not sure if the blue line was supposed to represent the width of the UGV, which was not the case in this experiment.

Several participants were confused about the time the UAV-recap was from. One participant wished to see the time difference displayed, and another participant stated that the UAV-view was a valuable support as long as the operator knew when the information had been obtained.

Worth noting is that two participants said after the experiment, that they were under the impression that the UAV was following *after* the UGV, which was the opposite from what was explained in the introduction. Another participant mentioned that they would not have trusted the system to stop in time if, for example, a moose had suddenly appeared from nowhere. Two participants felt some motion sickness.

4.3.5 Perception and Decision Allocation

Participants were asked to distribute 100% across the three agents to reflect how much each contributed to perceiving the situation. The human operator got a mean of 38.8% and the UGV a mean of 37.7%, resulting in nearly equal rating. The UAV received considerably less credit ($M = 23.5\%$) as expected, with two participants assigning it 0%, see Figure 4.21.

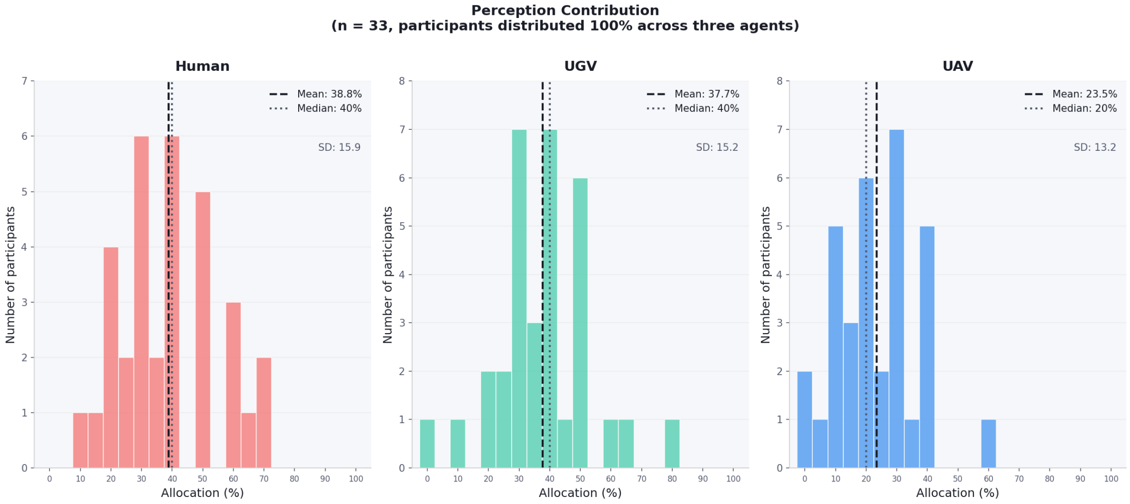


Figure 4.21: Perception Allocation

Secondly, the participants were asked to distribute 100% based on the agents perceived decision authority, see Figure 4.22. The human operator was rated as the dominant agent with a mean of 50.3%. Several participants assigned themselves full responsibility of 100%. The UGV received a mean of 37.6%, while the UAV was rated as contributing very little with a mean of 12.1%. Several participants assigned it 0%.

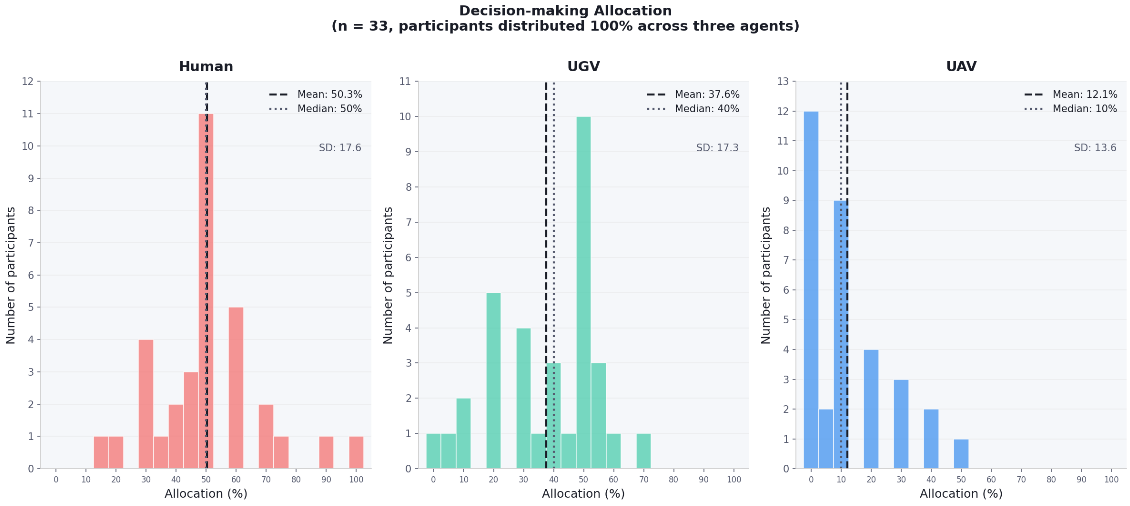


Figure 4.22: Decision Authority Allocation

5

Design Guidelines and Concept

This section presents the design guidelines and the concept design, based on the findings from the experiment and the literature review.

5.1 Design Guidelines

Fifteen design guidelines were derived from the study, representing design implications for human-autonomy systems, see Table 5.1. They were weighted from 1-5 and were subsequently organized, ranging from 5 (most important) to 1 (least important).

Table 5.1: Design guidelines derived from the study

Nr.	Design Guideline	Purpose	Src.	Weight
1	Assisted-Autonomous LOA	Facilitate human-in-the-loop vigilance and interaction	5	5
2	Show confidence level	System transparency to improve trust and mental model	5.3.3	5
3	Display UAV-view	Support operators perception, comprehension and planning	5.3.3	5
4	Display the time difference UAV-UGV	Chronological transparency to improve mental model	5.3.5	5
5	UAV time control	Facilitate timeline manipulation	5.3.4	5
6	UAV temporal-spatial connection	Anchor the UAV timeline to space through an icon on the map to support Situational Awareness and improve the mental model	5.3.4	5
7	UAV recap for TOR situations	Draw attention to the UAV and provide a birds eye view to support TOR decision-making	5.3.5	5
8	Emergency stop	Enable operator intervention	5.3.5	5
9	Display two lines corresponding to the front wheels of the UGV	Communicate width and projected path of the UGV	5.3.5	4
10	Provide a secondary screen	Increase screen real estate to facilitate improved cognitive ergonomics	5.3.5	4
11	Visual vehicle status	Provide immediate understanding of the vehicle's state	5.3.3	4
12	Additional options for the UAV-view and map	Facilitate operator workspace adjustments based on personal preference	5.3.5	4
13	Rotation toggle in map	Facilitate easy access to rotating the map	5.3.5	3
14	System log	Support machine-human communication through logging and error awareness	5.3.3	3
15	Information panel	Present relevant mission-related information (i.e. lists of vehicles, personnel, mission details, etc.)	5.3.2	3

In answer to the stated aim, *Assisted-Autonomous LOA* (1) is to be preferred. The results show that operators feel more responsible for the outcome of the mission when they have to make decisions, without an increase in Mental Workload. Additionally, it supports vigilance during monitoring phases. Regarding Perceived Reliability, A-LOA outscored AA-LOA, this is likely due to experimental factors and not LOA factors. Finally, AA-LOA somewhat negates the question of responsibility in case something goes wrong. To support smooth transitions and to provide transparency, research shows that the autonomous system should *Show confidence level* (2) to communicate system uncertainty (Endsley & Jones, 2025).

The data clearly shows that the UAV integration is beneficial and the system should therefore *Display UAV-view* (3). One important insight was that 90% of the participants preferred a system with the UAV-view, despite frequently displaying incorrect information. Almost all participants perceived the UAV-view as helpful, primarily for planning ahead, understanding the situation, and increasing their sense of safety.

The primary concern with the UAV-view was that it was unclear to participants how far ahead of the UGV it was. To mitigate this, the interface should *Display the time difference UAV-UGV* (4). This ensures chronological transparency and reduces critical gaps in Situational Awareness. Additional aid to the operator's temporal perception, is provided by including *UAV time control* (5), as it is important that the operator can control where in the UAV video feed they are. Further enhancing operators' Situational Awareness of where in the video and where in the world that corresponds to, the system should supplement the UAV time control with a *UAV temporal-spatial connection* (6). This can be achieved through a camera icon on the map which moves from the UAV icon, following the line which the UAV has flown, showing operators where in space the area currently shown in the UAV-view is, as well as when the footage was taken. This temporal-spacial connection should enable greater confidence in decision-making, assist information gathering activities, and further improve vigilance while the UGV is operating autonomously.

In instances where operators miss a potential hindrance and are presented with a TOR, gaze data and participant feedback made it clear that *UAV recap in TOR situations* (7) helped participants gather additional information by drawing attention to the UAV and providing a image of the potential scenario. In instances where the opposite is true, i.e. the operator spots something that the autonomous system missed, an *Emergency stop* (8) is crucial for ensuring mission safety and that operators always have the option to intervene. Supporting both the slower TOR decisions and rapid emergency stops, *Display two lines corresponding to the front wheels of the UGV* (9) visually communicates intent and ability to operators in a way akin to modern "self driving" cars.

Provide a secondary screen (10) facilitates greater flexibility in what, where and how information is provided to operators, allowing for increased control and more information while maintaining good cognitive ergonomics. This allows for clearer communication of mission critical information such as *Visual vehicle status* (11) for both the UGV and UAV without the obstruction of other important information. Additionally, the qualitative analysis established that operators wanted more ways to interact and engage with specifically the UAV and map, which can be achieved by *Provide more options for the UAV-view and map* (12), such as *Rotation toggle in map* (13). Further improving the sense of operators being in the loop, a *System log* (14) should provide updates and warnings from the machine-system at a micro scale, supported at the macro scale by an *Information panel* (15) providing an up to date overview of the mission status.

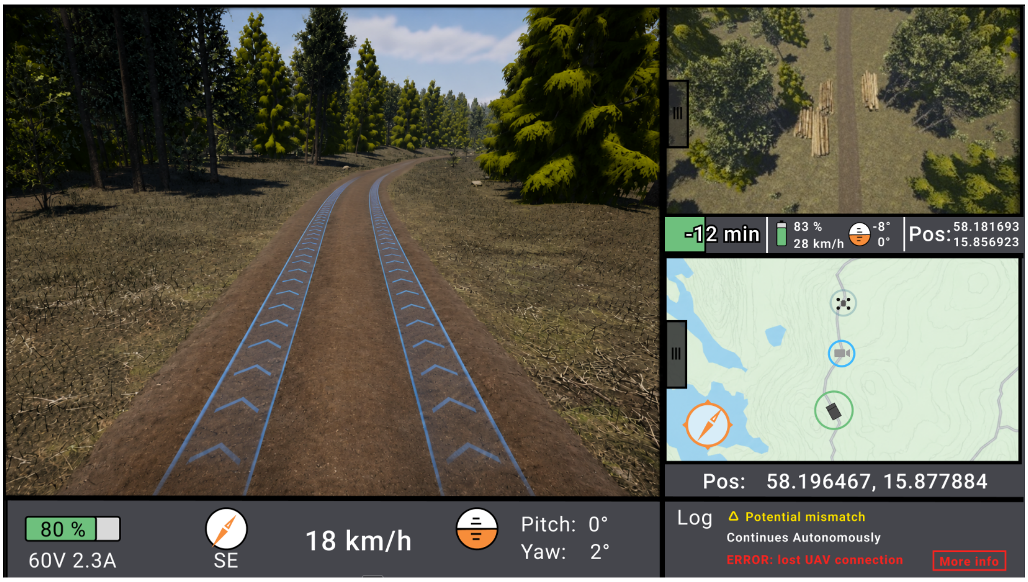
5.2 Design Concept

Based on the guidelines, a design concept was developed. The design concept consists of a dual-screen operator interface with a physical time control device, see Figure 5.1. The system is designed to support collaboration between the human operator, the UGV, and the UAV throughout the mission. The interface aims to balance situational awareness, navigation support, and system transparency, while reducing operator workload. The time control device consists of a shuffle wheel used to navigate the UAV timeline, along with two dedicated buttons, see Figure 5.2. One button returns the UAV feed to the live view, while the other activates a UGV "pseudo-sync" function, which repositions the timeline to the point where the UAV estimates the UGV to currently be in real time. The design concept was developed through brainstorming on inputs, outputs, and additional features (see Appendix M), and subsequently refined according to the design guidelines. Sketches of the interface can be seen in Appendix N.

Figure 5.1: Concept design sketch

Figure 5.2: Physical time control

The right screen, see Figure 5.3, focuses on operational awareness and vehicle monitoring. The main view is a view from the UGV's perspective, allowing the operator to observe the environment around it in real time. To the right, there is a UAV-view showing the UAV's perspective, and a map below, highlighting where the UGV and UAV is at the moment. The interface also displays real-time vehicle data, including battery status, a compass, speed, pitch, and yaw, together with system logs and potential error messages.



Potential mismatch
Continues Autonomously
ERROR: lost UAV connection

Figure 5.3: Right Screen

There is a "more options" button available on both the UAV-view and the map, in addition to a toggle rotation compass on the map. UAV data is displayed below the UAV-view. This includes a UGV-UAV-time display visualizing the time difference in color and a number of minutes, battery status, speed, pitch and yaw, and position coordinates. Finally, positional coordinates of the UGV are displayed below the map. The left screen, see Figure 5.4, presents a large map with associated tools located beneath it, as well as an information panel on the left side. The map is primarily used for placing way points and providing an overview of the UGV and UAV positions. The tools allow the operator to, among other things, view the locations of other units and monitor which areas the UAV has already scanned. The information panel functions as a central menu containing relevant mission-related information, such as lists of vehicles, personnel, and mission details.

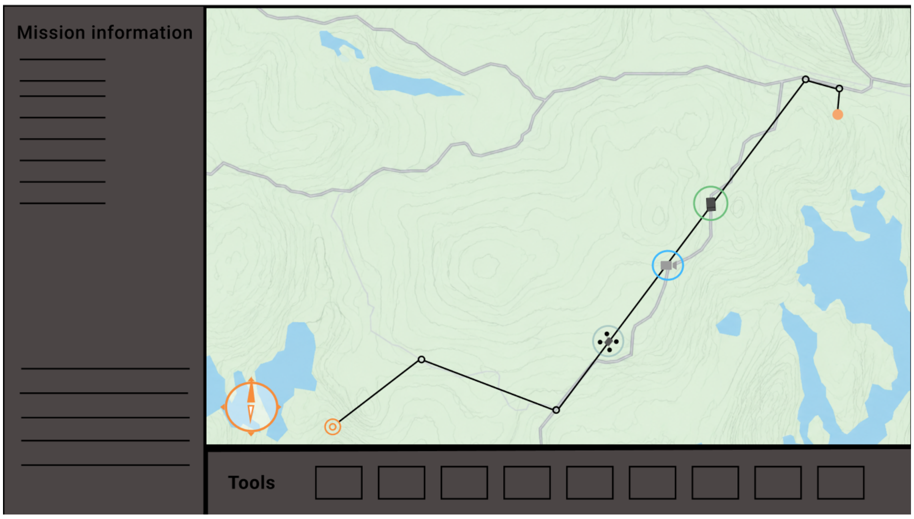


Figure 5.4: Left Screen

6

Discussion

This study investigated how UGV automation level and UAV information influence operator state, performance, and decision-making during control handovers. Overall, the results indicate that assisted autonomy increases perceived responsibility without increasing reported mental workload, while conflicting UAV information increases takeover rates but does not significantly increase decision time. These findings provide empirical support for the proposed design guidelines for human–UGV–UAV collaboration.

6.1 Aim

The stated aim was to:
1. Investigate how UGV automation level and integration of UAV information influence operator state, performance and decision-making when making control handover decisions in a human-UGV-UAV system.
2. Develop design guidelines for UAV integration, handover support, and function allocation in human–autonomy systems.

In regards to the first aim, the experiment successfully gathered quantitative (self reported rankings, logged times and decisions, and gaze data) and qualitative (questionnaires and notes) data from 33 participants with reasonably balanced group distributions. The number of participants combined with the amount of trials in the experiments provided a large enough data pool from which to draw fair conclusions. The analysis covered multiple factors within both operators state and performance in a semi structured manner, resulting in a few significant results from which we believe that we succinctly answer the research questions.

Secondly, 15 design guidelines have been synthesized, backed by analysis and literature studies. Finally, these guidelines are demonstrated through a design concept.

6.2 Method

This section discusses the choice and application of statistical models, as well as the design of the experiment.

6.2.1 Analytical Models

The models where chosen ad hoc based on the amount and type of data. Several analyses were revisited as the statistical approach was refined during the course of the project. Although some initial analyses used tests that assumed independent observations, subsequent analyses with more appropriate methods produced similar results. Some of the analysis using Wilcoxon and Mann-Whitney U test (which assume independent data points) are done on non-independent data, however on the ones that were re-done, the differences where marginal. Further analysis with stricter methods is justified but out of scope of this project.

6.2.2 Experiment Design

The choice of two LOA was done to ensure that the analysis of the human-in-the-loop, Assisted Autonomous system had a valid baseline from which to compare operators state while also providing participants with more experience. The participant groups where split into four groups as explained in Section 4.3 and analysis showed that the learning effects of having either LOA, and Conflict or Non-Conflict first, were adequately mitigated.

We selected Truck and Logs out of four initial ideas with the goal that a mismatch should be realistically feasible. However, many felt that the Logs were unrealistic since someone would have had to stack them back in just a couple of seconds. This was said by 3-5 participants, one of which claimed that this greatly effected his perception of the system as a whole.

All Conflicts were cases where the UAV showed an obstacle that didn't exist in the UGV, never the other way around. i.e. obstacle in UGV but not in UAV. Implementing additional Conflict conditions would however make it more difficult to draw conclusions from Take-Over decisions. We might see longer decision times, but that is not something we saw in the current configuration which should technically be "harder", since the majority trust the UGV. If the UAV was to show an obstacle but the UGV doesn't, participants would most likely be quick to take action regardless of what the UAV showed.

The UAV recap that came to light in Figure 4.15 was not the intended design and was a mishap on the part of the company delivering the videos. This made it harder to isolate the UAV use during the "decision window" as 3/4 videos provided a UAV stimulus in the form of a recap image warning the participants of an impending TOR, drawing their attention and directly affecting their gaze pattern.

Figure 6.1 below reflects the difference between the intended UAV behavior and the resulting one. The initial prototype had both the UAV and UGV moving faster, with the UAV at a steeper viewing angle in combination with a longer total video length. This resulted in the truck or log only showing for a few seconds during the drive, and then being shown again with the TOR. Meanwhile, the experiment video was shorter, with slower moving vehicles, so the truck or log was visible from the start, and then the recap showed 4-7 seconds ahead of the TOR. This discrepancy greatly affected the intended research as participants had more time to scan the UAV-view, which might explain the increased UAV gaze share in the first few seconds of each video. Further research with the initially planned UAV behavior could yield different results to the ones presented here.

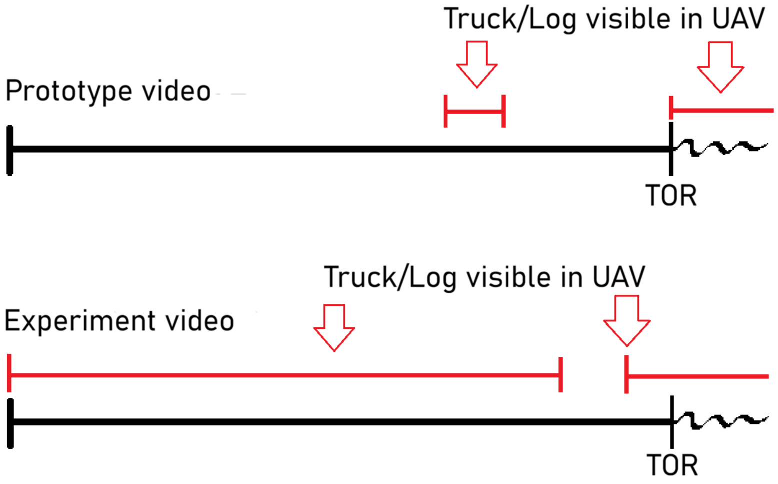


Figure 6.1: Prototype vs Experiment

The UI layout was technically out of scope and deserves a project all of its own, so the choices were made ad hoc based on video game layouts and basic SA principles early in the process. The results of the UI were deemed satisfactory and provided a good base on which to apply the subsequent design principles in the creation of the design concept.

Participant selection was primarily convenience-based, though we attempted to (and largely succeeded in) achieving balanced group distributions, with the exception of the military experience group. Age was skewed toward younger participants (20-25yr). Educational level was not collected, though it would have likely been skewed toward university level, given the recruitment context. Worth noting is that some participants were known to the experiment leaders prior to the study, and all participants were compensated with cinema tickets.

6.3 Experiment Results

6.3.1 Operator State and Performance

Mental Workload showed no significant difference between LOA, which was unexpected. We believe this to be the result of three key factors. Firstly, the manual driving was not yet implemented in the simulator, meaning that operators chose to take over manual control without actually performing the task of driving manually. Secondly, the two missions were only ≈10 min long, whereas a more realistic mission would be in the range of hours. Lastly, no external stress factors or parallel tasks were applied to participants, where a realistic situation would most likely contain a multitude. Taking this into account, care should be taken when comparing the Mental Workload of the LOA.

Regardless, the gaming effect still shows with statistical significance in an even split of participants (17 vs 16) that Gamers rank their mental workload 2-3 points lower on average in both LOA vs Non-Gamers. This supports further research into beneficial prerequisites for UGV-UAV operator training, such as map reading or multitasking from previous gaming experience.

With respects to Perceived Responsibility most participant (regardless of demographic factors) scored the Assisted Automation level higher. This was to be expected as they had a direct hand in the outcome of the mission. A risk with fully autonomous systems, aside from decreased vigilance - which multiple researchers, such as Endsley and Jones (2025) have stated before, is a loss of responsibility and "attachment" to the UGV in the real world. This might enforce the so-called "video game effect", specially if the UGV isn't sending audio, as this can lead to a lower sense of immersion, compounding the effect.

Confidence in combination with situational awareness are important factor in regards to decision making as it affects the probability of an operator acting or choosing to gather more information, see Figure 6.2 below, from Endsley and Jones (2025). It would therefore be interesting to analyze the correlation between participants reported confidence vs their decision time and probability of Take-over to see if there is a pattern in this data. Over all, participant rated their Confidence similarly for both LOA, but an interesting pattern emerged when comparing C and NC conditions. In general, participant scored decisions made Autonomously higher than their own decisions in the NC condition. This is likely due to several factors pertaining to the experiment design. Namely, in the Autonomous missions, participants get to see the UGV come to a stop, wait, and then drive on, completing the scenario and showing how the situation resolves. This in contrast to the Assisted Autonomous missions where after the UGV stops and initiates a TOR, it completes the scenario and participants don't see how the situation resolves. Additionally, since participants are unfamiliar with the system and are therefore expected to feel a bit uncertain, they might not understand why the UGV stopped in the first place since there is seemingly no reason for it communicated through neither the UGV or UAV.

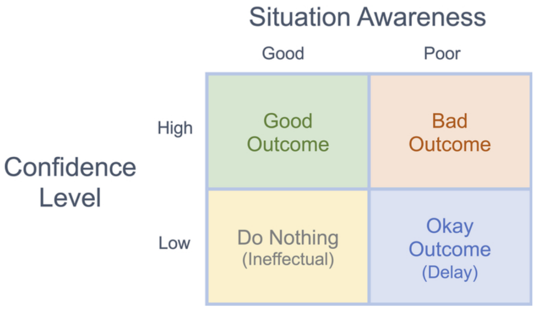


Figure 6.2: Confidence and Situational Awareness - Endsley and Jones, 2025

What is interesting is that this pattern reverses in the C condition, and that it does so differently for the different demographic groups. Notably, participants of the High Drive Frequency and High SoD rated their own choices significantly higher over those of the Autonomous system. This could be interpreted as people who have more experience on the road and who rate themselves as having a high SoD trusting their ability to read the situation and make a good choice in the face of misinformation from the UAV, higher than those who feel less experienced. Worth noting is that gaming was not a significant factor and both Gamers and Non-Gamers showed no statistical significance, though the pattern was the same.

Perceived Reliability is one of the most statistically significant results ($p < 0.01$), second to Perceived Responsibility ($p < 0.001$). Here participants clearly felt that the AA-LOA was less reliable even though the scenarios and conditions were identical in both mission. We believe that this is in part because of the previously mentioned "not seeing the resolution of the situation", but also because participants now had to try to integrate the incongruent UAV-view and felt more strongly that the system was not supporting them the way they expect it to. Additionally, the AA-LOA asking for help (per its design) could be interpreted as the system being less reliable. This is interesting when contrasting it to the Confidence scores, where some groups saw significant change between automation levels within at least the C condition. Here, Perceived Reliability only shows a significant difference for the High Drive, and that is in the NC condition, where once again they rank the autonomous system higher. Additionally, Low Driving Frequency and Low SoD show no significant difference between LOA while both High Driving frequency and High SoD show strong significance ($p < 0.01$), which in combination with the findings in Confidence helps build a case that people with less experience (or less faith in their abilities) are less critical of the systems flaws, regardless of LOA.

One of the most surprising results is that decision time does not significantly increase in the C condition. This is most likely due to the previously mentioned "UAV recap mishap" which gave participants additional time to make their decision ahead of the TOR, but we can't be sure without further research. What can be surmised though, is that the only statistically significant demographic predictor was High Driving Frequency with High SoD close behind, with both tending to take longer, which might help explain the increased Confidence in their decisions compared to other groups. Gamers, Low Driving Frequency and Low SoD trended toward making faster decisions, and seem to share a lot of patterns. Further analysis into their correlation would be interesting but is out of scope for this study.

As expected, Take-Over rate was significantly higher in the C condition which goes to show that participants were incorporating the UAV information into their decisions, often favoring to be "on the safe side." This was consistent across all demographic factors and only Low SoD showed a noticeable decrease in likelihood to takeover ($p = 0.151$) which is consistent with the hypothesis that they tend to trust the systems abilities more (or their own less) and are less critical of the system, in comparison to other groups.

6.3.2 Operator reliance on UAV

The most interesting result in the UAV gaze share analysis of the full length window videos, is how even the distribution among the groups is in the C condition, compared to the NC condition. Low SoD trends below the average in both conditions. The stand out feature is the split in NC condition that occurs within Gaming and Driving Frequency, for which we currently lack an explanation. Worth noting is that even though Military Experience has been largely excluded due to the low number of participants (6/32) there seems to be indications that military experience might result in increased UAV gaze share, which is supported through the qualitative analysis part of the experiment.

Analysis of gaze fixations search patterns were considered out of scope but preliminary studies yielded similar results.

6.3.3 Operator-Autonomy collaboration

Regarding Situational Awareness, it was expected that Perception would achieve the highest score, as it is generally considered the easiest SA level to achieve. Interestingly, Projection got a higher score than Comprehension. Since the third level of situational awareness (projection of future status) is generally considered the most difficult level to achieve, the results would be expected to reflect this more clearly, which they currently do not. An important note is that the definition of “projection” does not fully align with the question being asked, namely, *“How easy was it to react appropriately in each situation?”*. This is probably a reason why projection received relatively high scores.

In the qualitative analysis, the second question, *"Would you have preferred a system without UAV-view? Why or why not?"* should have been formulated as a less suggestive question. Same with the last question, *"Did you look towards the UAV because you found it useful or because you felt unsure about the UAV information"*. The answers would have been more useful if the question was formulated in a less suggestive way, for example, *"Why did you look at the UAV-view?"*. In the current version, all participants except three selected either “useful” or “unsure” from the provided options. With only these two suggestions in the question, it is likely that participants were primed to think in terms of usefulness.

The experiment notes provided many important insights by recording spontaneous ideas or thoughts that participants had, which would have likely been missed if they had only been collected afterwards. Something that could be important for a researcher, for example, an expressions of confusion, can contribute to the development of design guidelines aimed at reducing operator confusion. Additionally, the experiments often led to discussions in which participants reflected on their overall experience of the experiment or shared ideas for potential improvements they had thought of.

The perception and decision authority allocation was only assessed at the end of the experiment. It would have been interesting to compare the differences between the two LOA systems, however, to avoid influencing how much attention participants directed toward the UAV-view, these questions were not asked between the missions.

For perception allocation, the operator and UGV had similar mean values of approximately 40%, whereas the UAV had less, as expected. A comparison between the results in Figure 4.21 (Perception allocation) and Figure 4.13 and 4.14 (UAV gaze share in Conflict and No-conflict) showed an interesting similarity between how much the participants thought they used the UAV-view and how much they actually looked at it. The UAV gaze share had an overall mean of 21.6%, which is very close to their combined subjective answers of UAV perception with a mean of 23%. This suggests that, overall, they have a fairly good understanding of how much they use the UAV-view for perception.

For decision allocation, the human had a mean value of 50%, which is interesting since the operator did not make any decisions in the autonomous part. The UGV had a mean value of 38%, which is notably lower than the human got, even though these agents made the same amount of decisions throughout the experiment. As expected the UAV scored the lowest mean value of 12%.

6.3.4 Error Sources

The eye-tracker Pupil Invisible has an accuracy of 4.6° and is primarily designed for outdoor use where objects or areas of interest are further away from the subject, which partly negates the parallax effect according to Pupil, 2026. For the setup used in this thesis, the distance between participants and the monitor varied, in addition to the viewing angle. This introduces some noise in the gaze data, which partially explains the share of "Outside" in the AOI-plots. But this is also a result of participants looking off screen toward the keyboard when responding to the TOR. Preferably all participant would have seen the same calibration image at the start and end of each experiment, however as mentioned, all went through the same calibration process ahead of the experiment and the images were only used for post hoc adjustments in select cases where the video recording showed a mismatch. In lieu of calibration images, the input options and prolonged viewings were used to correct some positional data.

Time sync is always a source of error when using multiple systems, in this case we used the internal clocks to calculate the time difference and made sure to sync the two devices to the same time server ahead of each experiment. Additionally, four white flashes were shown throughout the experiment with corresponding UTC times logged. These were then used to verify that the clocks didn't drift throughout the $20 \approx$ minutes of each experiment in addition to post hoc time sync.

Misunderstandings are commonplace in every aspect of life, and this experiment was no different. The two main categories of misunderstanding were *system related* and *question related*. Regarding the system, several participants were confused when the autonomous level came to a stop and then proceeded without initiating a TOR. This confusion likely didn't affect the results in any substantial manner, but indicates that the pre-experiment explanation and practice was lacking. A autonomous example should have been included in the practice session. Additionally, two participants shared during the experiment that they assumed the UAV-view to be a "rendering of sensor data" and not a video feed. One explained that it's the only way that the Logs C scenario could make sense.

This could also be explained by the fact that English was used throughout the experiment, which was neither the participants' nor the experiment leaders' first language. This potential language "barrier" might have contributed some noise in the data as participants figured out what was being asked. Different interpretations of the same question may also have contributed to variability in the responses. This was partially mitigated by encouraging them to ask questions if they felt unsure about anything, with the caveat that we can't say too much about certain things in order to protect the integrity of the experiment.

6.4 Design Guidelines

The following section pertains to the discussion of the main Design Guidelines in regards to the aim of the project.

6.4.1 Level of Automation

Based on the findings, Design Guideline 1, *Assisted-Autonomous LOA* appears to be the preferable approach. Even though, as mentioned earlier in Section 6.3, the results of mental workload should be taken with a grain of salt. Based on the results of the experiment, literature review and communication with stakeholders throughout the project, the findings support recommending an assisted-autonomous approach. Additionally, fully autonomous UGV-UAV systems are still limited in their capabilities at the time of this writing and require further research and development. As this is true, Guideline 2, *Show confidence level*, would facilitate a more transparent system where the operator gets a sense of how confident the autonomous system is in its decisions or choices. This counteracts the typical "AI black-box", keeping the operator more in-the-loop while improving trust and perceived reliability.

6.4.2 UAV integration

The research clearly shows that a UAV is wanted by operators, but the quality is sensitive to the way in which it is implemented.

First, having a UAV conducting reconnaissance is for full transparency, a factor mentioned by the Swedish armed forces in a 2025 report requesting research and development of UGV systems for medevac and casevac, to which EIRA-systems is responding. From an autonomous systems perspective, the addition of a forward facing, aerial sensor provides information from which the UGV can plan further ahead, reducing the risk of it getting stuck or hitting a dead end. From an operators perspective however, the benefits are not as clear cut. The addition of a second vehicle with vastly different control characteristics puts great mental strain on operators according to Chen et al. (2008), and as such the UAV must have a robust autonomous system where operators only need to input way points to ensure that they keep their focus on the UGV actuating the mission. By simply adding a UAV-view window next to the UGV view, operators have greater perception of what is to come but the issues of a simple implementation such as this are twofold:

Firstly, when the system encounters a situation which it can not handle and requests a take-over, operators are left with only the UGV view which is often a low resolution, wide angle camera, low to the ground. Camera set-ups such as these prove difficult when judging downward slopes, holes and depth of foliage.

Secondly, the UAV is flying continuously towards its way point far ahead of the UGV (seconds, minutes or even hours) which means that any one spot in the terrain will only be seen for a few seconds. This leaves the operator at risk of missing hazards, without the ability to gather additional information, and to mentally map when and where this hazard might effect the mission.

To tackle these issues, the findings support that Guidelines 4-7 provide a solid starting point for most UGV-UAV (Actuator-Sensor) systems.

Regarding the former issue of information during TOR, Guideline 7, *UAV recap*, provides operators with a birds eye view over the area in which the UGV is currently, as it was when the UAV was there. The key issue, which this research has investigated, is whether or not the temporal mismatch of what is presented by the UAV and UGV is in anyway detrimental to the outcome of the mission.

Our findings were mixed, on the one hand, decision time did not increase significantly, on the other hand, the likelihood of manual intervention greatly increased. This is not a bad thing per se, but it does add to the operators workload in situations where the autonomous system could have handled it if the operator could confirm that the sensor mismatch was due to old and no longer relevant UAV information. Further, the decision time is affected by the UAV-recap mishap, presenting the mismatching information a few seconds ahead of the TOR, thereby providing participants with the chance to process the information and make a decision ahead of time. It can be argued that the similarity of decision time in both conditions is due to a lack of "stakes" in the experiment, i.e. nothing bad will happen if I'm wrong. But we would argue that the increase in manual intervention combined with the results of the qualitative portion of the experiment points towards a risk avoidance behavior, possibly explained by participants wanting to make the "correct" choice in a setting where they are watched and monitored. Additionally, participants where not instructed to make the fastest decisions possible, but to simply make a decision when they felt that they had come to one, and that we would keep track of the time. Further, the questionnaires strongly indicate that even when the UAV information mismatches that of the UGV 50% of the time (as was the case in this experiment), participants feel that the benefits outweigh the negatives.

Ergo, it is our belief that incongruent information in TOR is not detrimental to the mission outcome, that operators want the additional information, and that they can be further supported in their uncertainty management by implementation of Guideline 4, *UAV-UGV time difference*. This provides additional context at a glance, in the form of a time delta between when the UAV took the recap picture and when the TOR is initiated.

Regarding the latter issues pertaining to the UAV flying ahead of the UGV, we believe that enabling the operator to scan through the video provided by the UAV with Guideline 5, *UAV time control*, is a good place to start. By implementing a so called jog-wheel, (often used in surveillance stations and video editing for timeline manipulation) operators gain the ability to gather more information by manipulating where in the UAV feed they want to scan, without affecting the UAVs flight path. These controls should be supplemented with quick-action buttons to quickly return to the current live view, as well as "UGV pseudo-sync" which moves the video feed to correspond to the area in which the UGV is currently in.

This raises its own issues, namely the aforementioned time difference between information streams which could be solved with Guideline 4, however, time delta in this context would be an estimated time of arrival, not the age of the information provided. Because both the UGV and UAV are moving, and the UGV is constantly assessing and re-assessing its optimal path, estimates such as these have a great deal of uncertainty built in. The benefit of such a solution is that operators can at a quick glance get an idea of how long they have before they need to deal with a certain situation. The downside is that these estimates can vary quite a bit, and research shows that operators are better at dealing with uncertainty when they have concrete factors or sensor data than when presented with a compound approximation, Endsley and Jones, 2025. By instead providing hard data, such as distance between UGV and UAV, operators can take factors such as terrain, weather and payload into account and estimate the time of arrival themselves, thereby improving their mental model of the system and staying in-the-loop. To further support both the situational awareness and the mental model, Guideline 6, *UAV Spatial-Temporal connection*, provides a clear GUI icon on the map showing what area of the path the operator is currently viewing through the UAV. By supplementing the UAV feed control with this moving icon, operators can conduct more structured scans and build a greater understanding of the terrain ahead, while also training their "UGV $\rightarrow$ UAV-distance $\Rightarrow$ arrival time"-muscle, further enhancing operator efficiency. This of course comes with its own drawbacks, mainly the question of missed information. Say that the operator stops the feed when the UAV flies over something noteworthy; while the operator is taking a closer look, the UAV is still flying. If the operator now pushes the "UAV Live"-button, all of the area that the UAV has flown over since then will be missed. Therefore, further research should be conducted into how operators can gather more information, without missing out on whats happening live.

6.5 Ethical, Societal, and Ecological Aspects

This section will discuss societal, ethical, and ecological aspects of the Autonomous systems and the experiment in this report.

Autonomous systems have some inherent ethical questions, which are discussed in this section. Mainly, the question of who is responsible for the consequences of the systems' mistakes or accidents. The UGV's stated and intended purpose is transport and logistics, however it would be naive to consider this the UGV's only potential use case. The use of autonomous platforms with mounted weaponry, both remote controlled and autonomous, have been well documented in wars around the world which raises additional concerns from an ethical standpoint. On one hand, when used in a defensive capacity the UGVs protect soldiers and civilians alike while minimizing friendly casualties. On the other hand, when used offensively, especially in asymmetric warfare, the attacker can inflict casualties without committing troops to an area. As this paper and project is focused on the use of Operator-UGV-UAV collaboration in transport and logistics, the possibility of re-purposing the platform is disregarded.

From a societal aspect, increased automation in search and rescue operations could lead to safer and more efficient working conditions for operators and medics alike, resulting in more people being saved and less medics hurt. Automation could therefore lighten the burden on logistics- and medical–professionals, freeing up resources on a societal level. On the other hand, automation could lead to a decrease in some traditional work-fields, while employing more people in the high-tech sector, shifting the bar for employability towards higher educated societal groups.

From an ecological perspective, autonomous electric vehicles require substantial amounts of heavy metals and rare earth minerals, affecting environmental impacts related to resource extraction. Nevertheless, advancements in recycling technologies and the development of alternative materials may mitigate some of these effects over time.

Since participants have been part of this study, research ethics have been considered carefully. Each participant read and signed two consent forms, one for the researchers and one to keep with the participation ID. They were told that they could ask to delete their data at any time by the participation ID, and that they were allowed to leave at any time. Personal data has been collected, stored, and processed in accordance with the General Data Protection Regulation (GDPR).

6.6 AI Use

The LLMs ChatGPT by OpenAI and Claude by Anthropic were used to generate Python and R code for analyzing and visualizing the quantitative data, as well as HTML and JavaScript code for the experiment implementation. All code was parsed through and verified before use. No conclusions or answers generated in conjunction with the code have been used in this report.

ChatGPT and Grammarly have been used for translating specific words, grammar support and to find expressions that communicate the intended meaning from Swedish to English. Additionally, Claude was used to transcribe and digitize handwritten notes which were checked and corrected, in order to support an analog workflow.

6.7 Future Work

As mentioned in the project introduction, one of the main limitations of this project was the fact that the simulator was being developed in parallel to the research. This meant that key features, such as the ability for participants to manually drive the UGV were missing. Future studies would ideally employ a more advanced simulator were participants need to manually control the UGV after a Take-Over decision in order to measure the change in state and performance. Further, different control methods such as game controllers, keyboards and steering wheels should be compared and analyzed with the goal of providing design guidelines to the development of an operator station.

Continuing on the theme of limitations, further work would span a longer time frame, allowing for research into the effects of longer mission times (> 30 min) on operators state and performance, mainly vigilance and mental workload. Additionally, outside stressors such as time pressure, cramped environments or multi modal inputs (i.e. sounds, smells, vibrations, etc.) could provide more realistic data pertaining to operator state and performance, specifically regarding the integration of conflicting information during decision-making.

As discussed in previous sections, the main drawback of this experiment was due to the misaligned UAV recaps not being implemented as initially intended. As such, further work should ensure that all recaps are synced for a more robust result. Additional studies would analyze the potential effects of how far ahead of the UGV, the UAV is flying, with the aim of providing operational guidelines for future use of the system.

As simulators evolve, further studies into benefits of certain demographic factors could be a fruitful avenue to explore in greater depth. As this study lacked participants with a military or SAR background, that would be a good place to start since the results show that even small groups stand out in the data. Further work would also aim to find out whether gamers really are faster decision-makers, and if drivers are more confident or make safer decisions.

7

Conclusion

The study investigated how UGV automation levels and UAV-view integration influence operators decision-making during control handovers by gathering quantitative and qualitative data through simulated transport missions.

The findings show that Assisted-Automation results in higher perceived responsibility without a significant increase to mental workload, and for experienced operators it yields greater confidence in decisions based on incongruent data. In addition, the UAV is an important tool for maintaining good SA during decision making, regardless of the automation level. The results show quantitatively, through scores, times, decisions and gaze data that operators integrate the UAV-view throughout the mission and that they are aware of the degree to which they rely on it. Qualitative analysis shows that even though operators are more likely to intervene and perceive the system to be less reliable when presented with conflicting data, they are willing to deal with even a high (50%) miss-rate for the value that a UAV-view provides in overview, context, planning, and the increased sense of safety.

The findings support the synthesis of design guidelines such as, that by integrating temporal transparency measures, the benefits can be maintained while the drawbacks are mitigated - providing operators with more information and a greater sense of security and confidence in the system. The discussion provides nuance and underscores that further evaluation is necessary for some metrics such as decision-time and the effects of demographic factors.

Demographic factors had a significant effect in some areas:

Firstly, Gamers consistently scored lower on mental workload over Non-Gamers, as well trending faster decision times with a strong correlation between longer times and an increased probability of intervention.

Secondly, High Driving Frequency and High SoD were the lead predictors in longer decision times and where strongly associated with an increased confidence in their own decisions during Conflict scenarios over that of the Autonomous system.

Lastly, Low Driving Frequency and Low SoD showed no significant difference in their perceived reliability of the system between the different conditions, seemed less critical of the system, and were more cautious of intervening.

In summary, the benefits of assisted-autonomy in conjunction with a time-transparent integration of the UAV provides operators with the best conditions for situational awareness during decision-making in heterogeneous human-autonomy systems.

Bibliography

Bligård, L.-O. (2015). *ACD^3 - Utvecklingsprocessen ur ett människa-maskinperspektiv* (2nd ed., tech. rep.). Chalmers tekniska högskola. http://acd3.com/assets/files/ACD3%20-%20Utvecklingsprocessen%20ur%20ett%20manniska-maskinperspektiv%202.1.pdf

Chen, J. Y. C., Durlach, P. J., Sloan, J. A., & Bowens, L. D. (2008). Human-robot interaction in the context of simulated route reconnaissance missions. *Military Psychology*, *3*, 175–187. https://doi.org/10.1080/08995600802115904

EIRA Systems. (2026). *Eira systems.* Retrieved July 3, 2026, from https://eirasystems.com/

Endsley, M. R., & Jones, D. G. (2025). *Designing for situation awareness: An approach to user-centered design* (3rd). CRC Press. https://bookshelf.vitalsource.com/books/9781040272831

Gold, C., Happee, R., & Bengler, K. (2018). Modeling take-over performance in level 3 conditionally automated vehicles. *Accident Analysis & Prevention*, *116*, 3–13. https://doi.org/10.1016/j.aap.2017.11.009

Grocholsky, B., Keller, J., Kumar, V., & Pappas, G. (2006). Cooperative air and ground surveillance. *Robotics Automation Magazine, IEEE*, *13*, 16–25. https://doi.org/10.1109/MRA.2006.1678135

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland.

Liu, C., Zhao, J., & Sun, N. (2022). A review of collaborative air-ground robots research. *Journal of Intelligent & Robotic Systems*, *106*(3). https://doi.org/10.1007/s10846-022-01756-4

Munasinghe, I., Perera, A., & Deo, R. C. (2024). A comprehensive review of uav-ugv collaboration: Advancements and challenges. *Journal of Sensor and Actuator Networks*, *13*(6), 81. https://doi.org/10.3390/jsan13060081

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

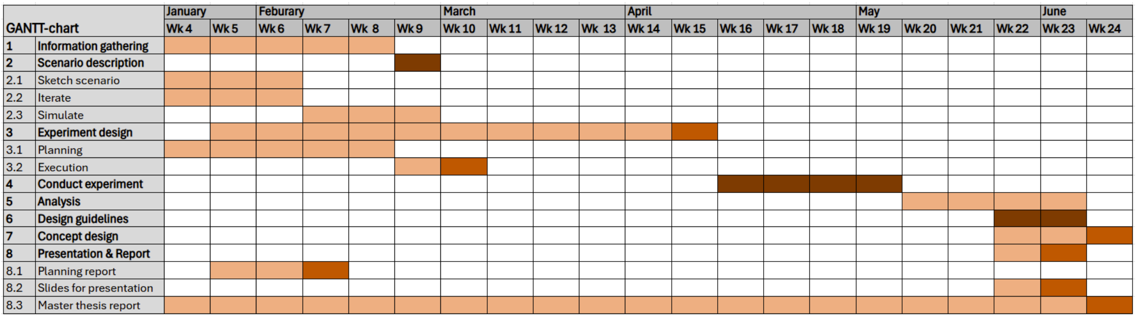
Pupil, P. L. (2026). *Pupil invisible* [Accessed: 2026-06-24]. Retrieved June 24, 2026, from https://pupil-labs.com/products/invisible/tech-specs

Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 285–307. https://doi.org/10.1016/j.trf.2019.04.020

A

Appendix

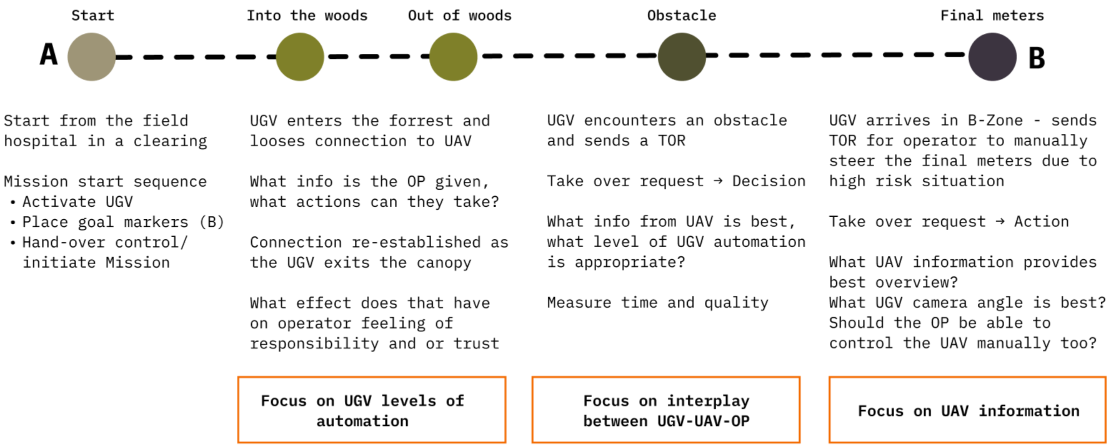
Appendix A - GANTT



		January		Feburary				March				April					May				June	
GANTT-chart		**Wk 4**	**Wk 5**	**Wk 6**	**Wk 7**	**Wk 8**	**Wk 9**	**Wk 10**	**Wk 11**	**Wk 12**	**Wk 13**	**Wk 14**	**Wk 15**	**Wk 16**	**Wk 17**	**Wk 18**	**Wk 19**	**Wk 20**	**Wk 21**	**Wk 22**	**Wk 23**	**Wk 24**
1	**Information gathering**	▨	▨	▨	▨	▨																
2	**Scenario description**						🟫															
2.1	Sketch scenario	▨	▨	▨																		
2.2	Iterate	▨	▨	▨																		
2.3	Simulate				▨	▨	▨															
3	**Experiment design**		▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	🟧									
3.1	Planning	▨	▨	▨	▨	▨																
3.2	Execution						▨	🟧														
4	**Conduct experiment**													🟫	🟫	🟫	🟫					
5	**Analysis**																	▨	▨	▨	▨	
6	**Design guidelines**																			🟫	🟫	
7	**Concept design**																			▨	▨	🟧
8	**Presentation & Report**																			▨	🟧	
8.1	Planning report		▨	▨	🟧																	
8.2	Slides for presentation																			▨	🟧	
8.3	Master thesis report	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	🟧

Figure A.1: GANTT

Appendix B - Scenario descriptions



Start from the field hospital in a clearing

Mission start sequence
- Activate UGV
- Place goal markers (B)
- Hand-over control/ initiate Mission

UGV enters the forrest and looses connection to UAV

What info is the OP given, what actions can they take?

Connection re-established as the UGV exits the canopy

What effect does that have on operator feeling of responsibility and or trust

UGV encounters an obstacle and sends a TOR

Take over request → Decision

What info from UAV is best, what level of UGV automation is appropriate?

Measure time and quality

UGV arrives in B-Zone - sends TOR for operator to manually steer the final meters due to high risk situation

Take over request → Action

What UAV information provides best overview?
What UGV camera angle is best?
Should the OP be able to control the UAV manually too?

Focus on UGV levels of automation

Focus on interplay between UGV-UAV-OP

Focus on UAV information

Figure A.2: Scenario description

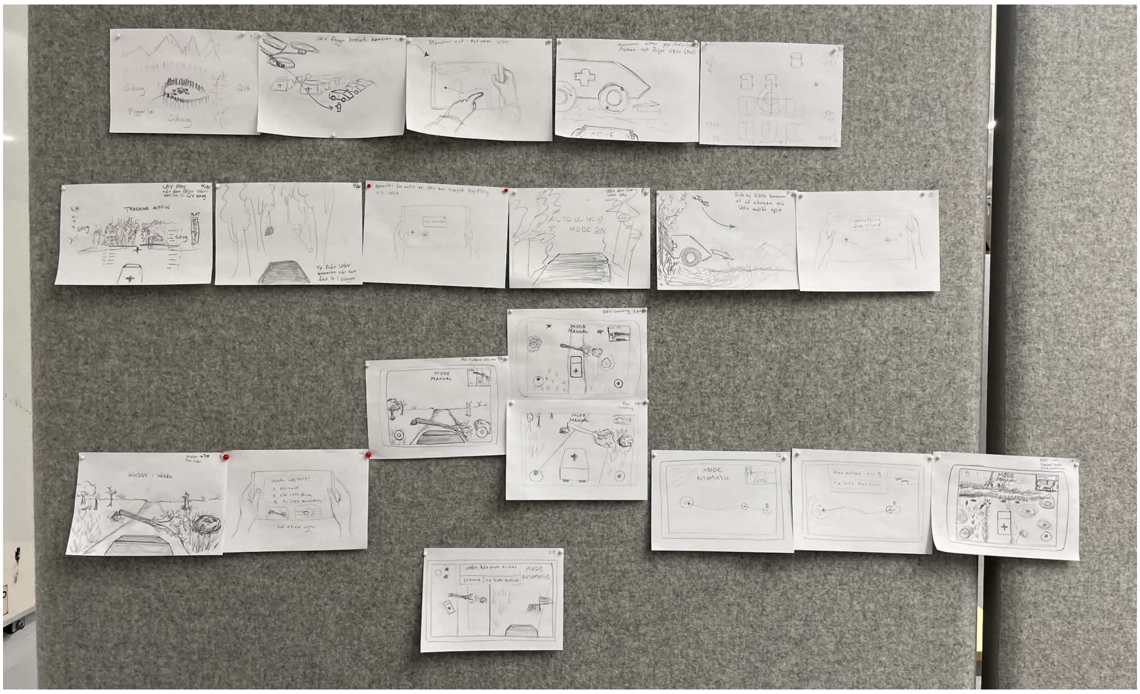


Figure A.3: Scenario sketches

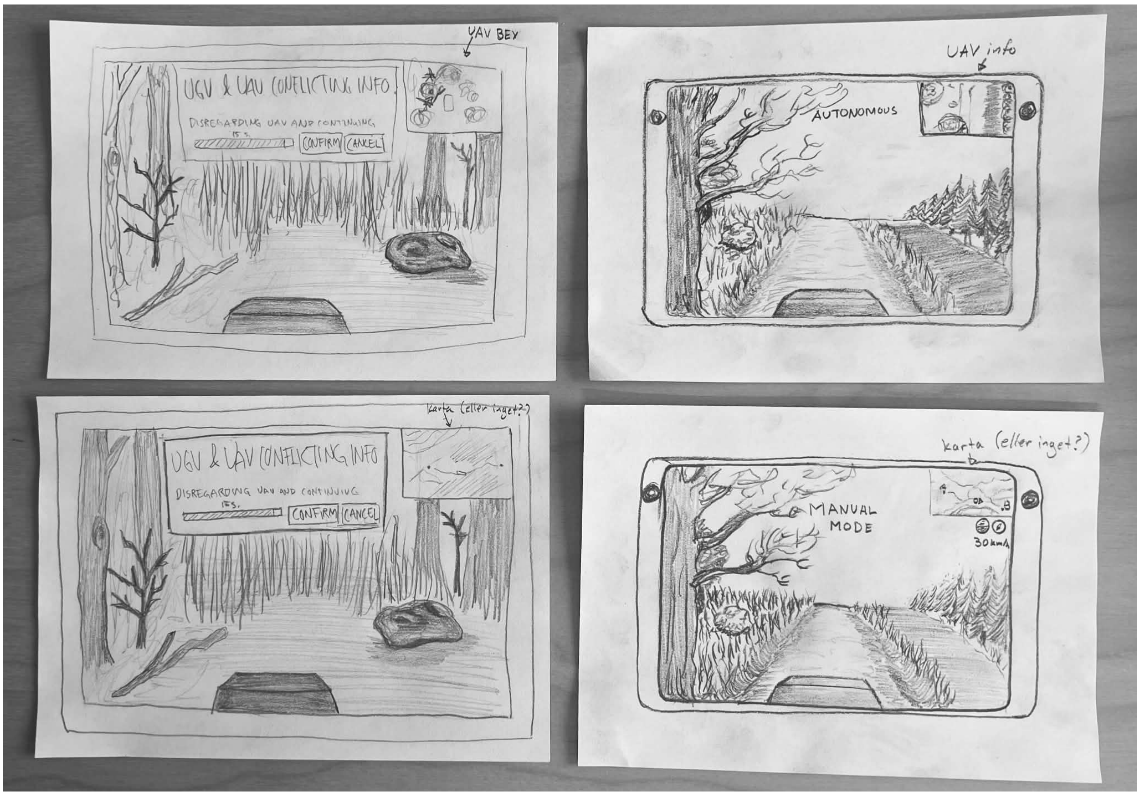
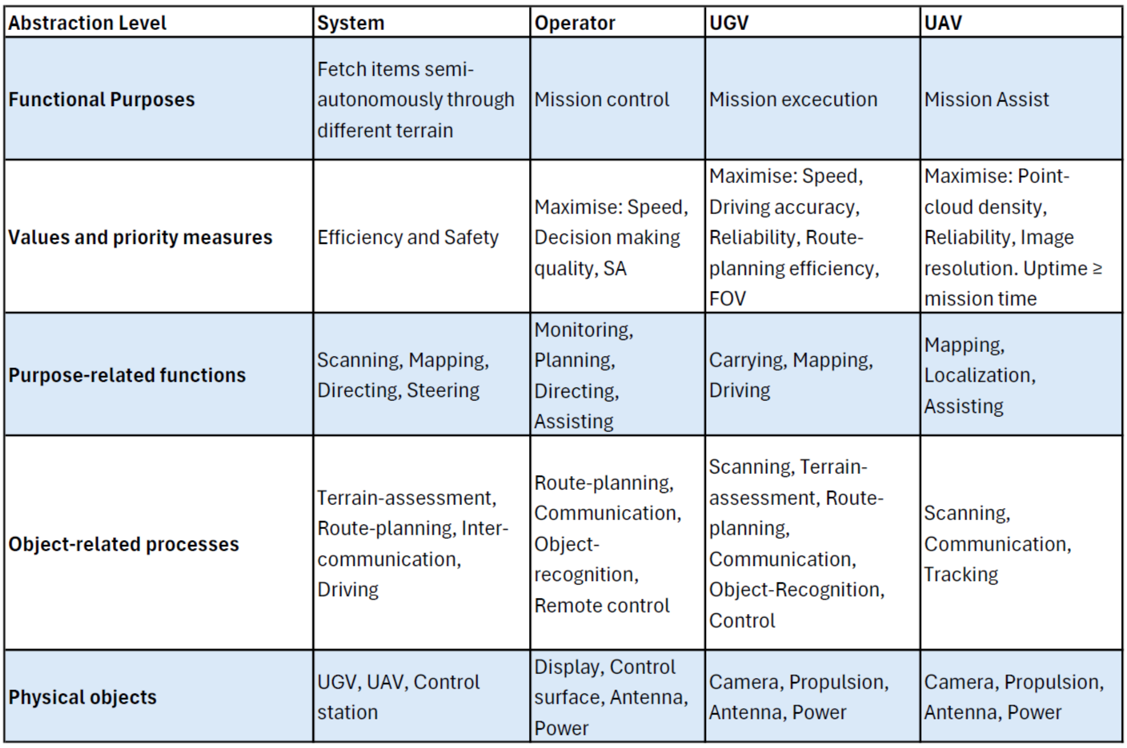


Figure A.4: Scenario sketches

Appendix C - Function allocation

Table A.1: Function allocation



Abstraction Level	System	Operator	UGV	UAV
Functional Purposes	Fetch items semi-autonomously through different terrain	Mission control	Mission excecution	Mission Assist
Values and priority measures	Efficiency and Safety	Maximise: Speed, Decision making quality, SA	Maximise: Speed, Driving accuracy, Reliability, Route-planning efficiency, FOV	Maximise: Point-cloud density, Reliability, Image resolution. Uptime ≥ mission time
Purpose-related functions	Scanning, Mapping, Directing, Steering	Monitoring, Planning, Directing, Assisting	Carrying, Mapping, Driving	Mapping, Localization, Assisting
Object-related processes	Terrain-assessment, Route-planning, Inter-communication, Driving	Route-planning, Communication, Object-recognition, Remote control	Scanning, Terrain-assessment, Route-planning, Communication, Object-Recognition, Control	Scanning, Communication, Tracking
Physical objects	UGV, UAV, Control station	Display, Control surface, Antenna, Power	Camera, Propulsion, Antenna, Power	Camera, Propulsion, Antenna, Power

Appendix D - Hindering condition TOR for A-LOA

Figure A.5: Truck Hindering

Figure A.6: Logs Hindering

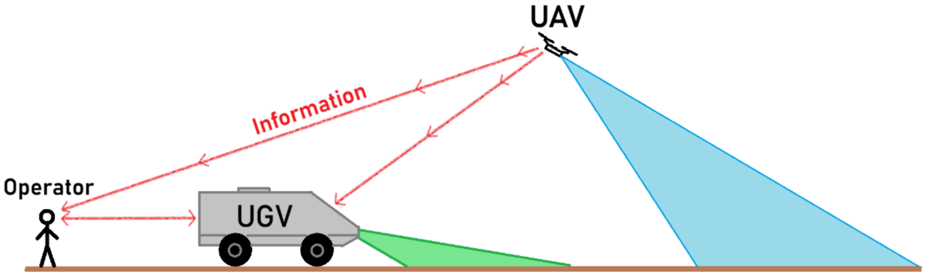
Appendix E – Consent form

Participant Information

Thank you for taking the time to participate in this experiment!

This study is part of a master's thesis in Industrial Design Engineering, within the EIRA-X project, funded by Vinnova. The aim is to investigate how operators make decisions in different situations using videos, images, and questions.

Your task is to be a remote operator that **monitors** a system consisting of a semi-autonomous ground vehicle (UGV), and an autonomous flying drone (UAV). The UGV is a large vehicle that can drive autonomously but may need assistance with certain decisions, while the UAV provides information to support the UGV's navigation by flying ahead (at least 5–10 seconds) of the UGV. You will complete two *missons* in which the system will navigate from a starting location to a destination to pick up a package. It is important that you monitor the system continuously, as it may require your assistance.



As part of this study, data on your responses and the response time will be collected, along with eye-tracking data to calculate gaze data. The study may involve mild fatigue or motion sickness. As the study also contains brief visual light transitions between video clips, we do not recommend you to participate if you have photosensitive epilepsy or is sensitivity to flashing lights. Participation is voluntary. You may stop at any time without a reason.

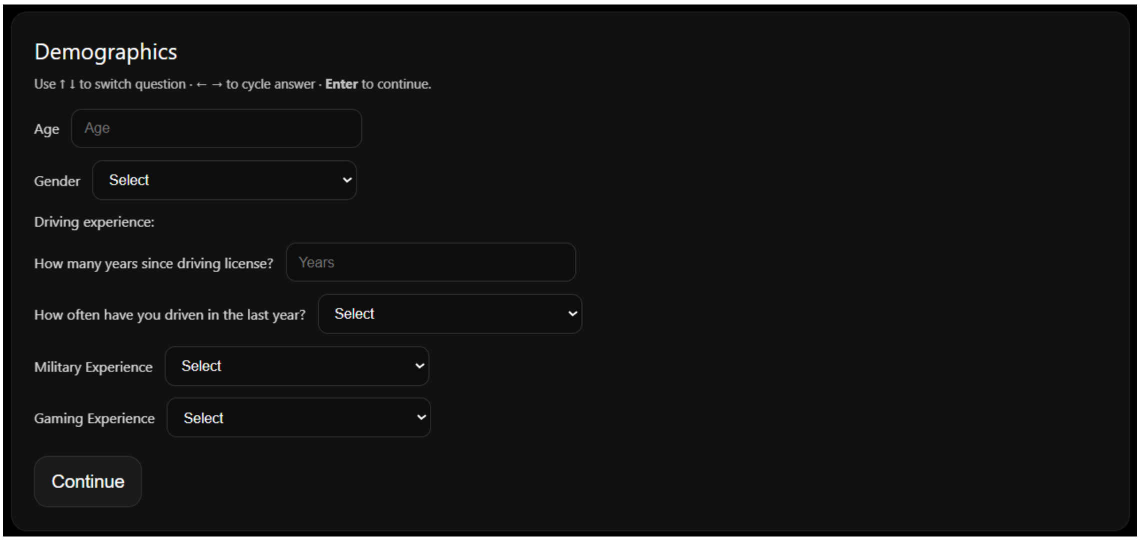
Your data is processed based on your consent (GDPR Art. 6 (1) a). No identifying personal data will be stored within the dataset used for analysis. Your data is linked only to a Participant ID. You have the right to ask us to delete your data (Note: You will need to know your "Participant ID"). Data is stored securely and only accessed by the research team. Raw recordings will be deleted or anonymized within 1 month. Anonymized data may be kept for research purposes.

By signing this form, I confirm that I have read and understood the information provided above. I agree to participate in this study and consent the processing of my data as described.

Location & Date	Signature
____________________	____________________

Contact
Linda Pipkorn, linda.pipkorn@chalmers.se
Chalmers University of Technology, Department of Mechanical Engineering

Appendix F - Initial questionnaire



Demographics

Use ↑ ↓ to switch question · ← → to cycle answer · **Enter** to continue.

Age: Age

Gender: Select

Driving experience:

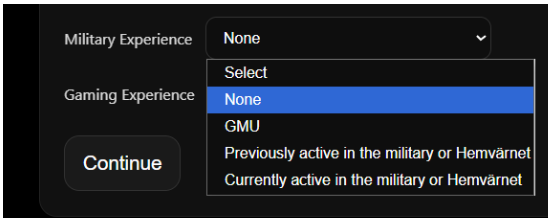
How many years since driving license?: Years

How often have you driven in the last year?: Select

Military Experience: Select

Gaming Experience: Select

Figure A.7: Demograpghics questions

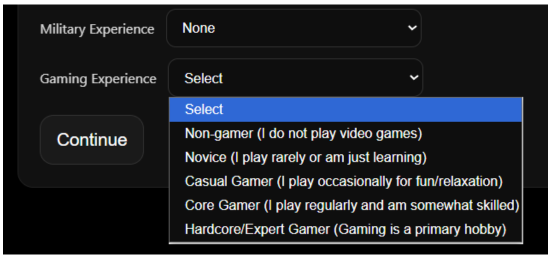


Military Experience: None

Gaming Experience

- Select
- None
- GMU
- Previously active in the military or Hemvärnet
- Currently active in the military or Hemvärnet

Figure A.8: Options for military experience

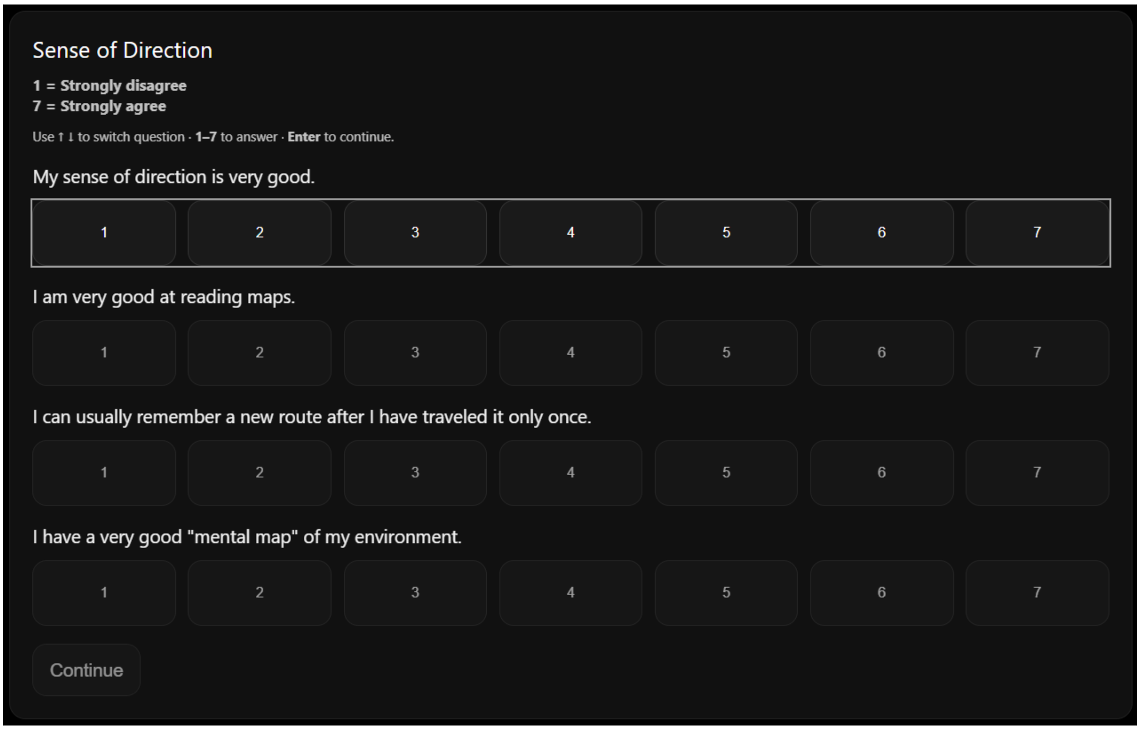


Military Experience: None

Gaming Experience: Select

- Select
- Non-gamer (I do not play video games)
- Novice (I play rarely or am just learning)
- Casual Gamer (I play occasionally for fun/relaxation)
- Core Gamer (I play regularly and am somewhat skilled)
- Hardcore/Expert Gamer (Gaming is a primary hobby)

Figure A.9: Options for gaming experience



Sense of Direction

1 = Strongly disagree
7 = Strongly agree

Use ↑ ↓ to switch question · **1–7** to answer · **Enter** to continue.

My sense of direction is very good.

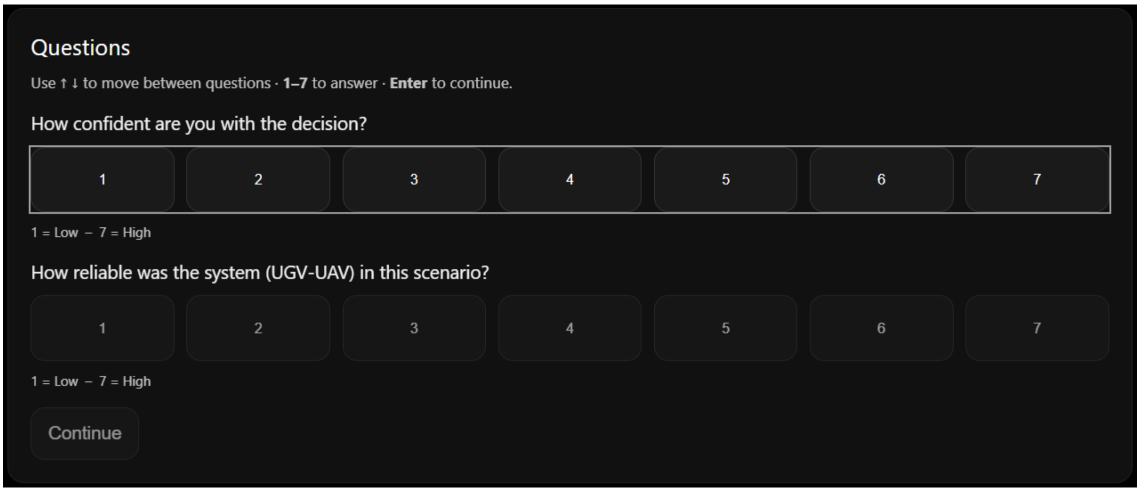
I am very good at reading maps.

I can usually remember a new route after I have traveled it only once.

I have a very good "mental map" of my environment.

Figure A.10: Questions related to sense of direction

Appendix G - Decision questions



Questions

Use ↑ ↓ to move between questions · **1–7** to answer · **Enter** to continue.

How confident are you with the decision?

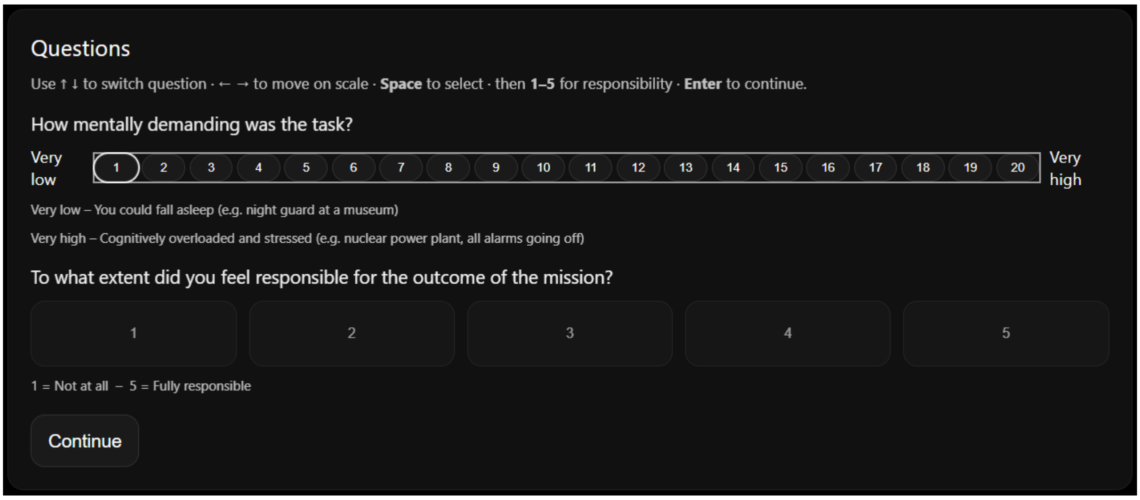
1 = Low – 7 = High

How reliable was the system (UGV-UAV) in this scenario?

1 = Low – 7 = High

Figure A.11: Questions related to decision confidence and reliability

Appendix H - Intermediate questionnaire



Questions

Use ↑ ↓ to switch question · ← → to move on scale · **Space** to select · then **1–5** for responsibility · **Enter** to continue.

How mentally demanding was the task?

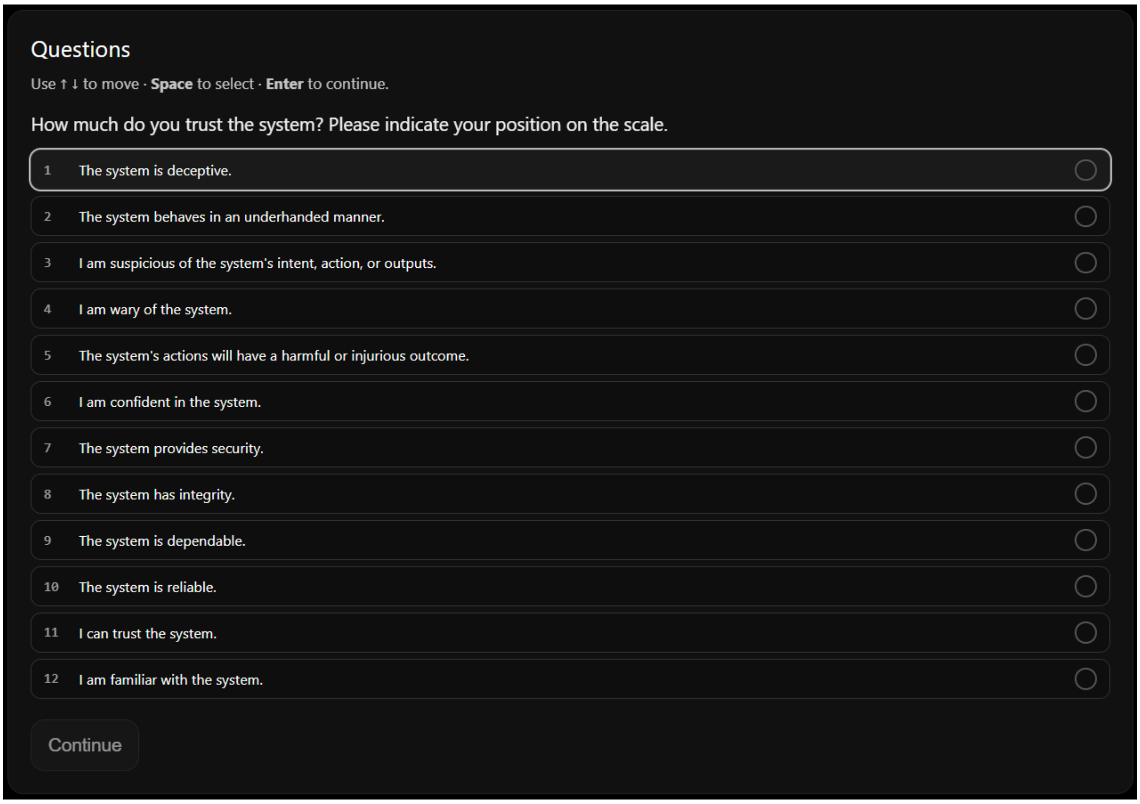
Very low – You could fall asleep (e.g. night guard at a museum)

Very high – Cognitively overloaded and stressed (e.g. nuclear power plant, all alarms going off)

To what extent did you feel responsible for the outcome of the mission?

1 = Not at all – 5 = Fully responsible

Figure A.12: Questions related to mental workload and responsibility



Questions

Use ↑ ↓ to move · **Space** to select · **Enter** to continue.

How much do you trust the system? Please indicate your position on the scale.

1 The system is deceptive. ○

2 The system behaves in an underhanded manner. ○

3 I am suspicious of the system's intent, action, or outputs. ○

4 I am wary of the system. ○

5 The system's actions will have a harmful or injurious outcome. ○

6 I am confident in the system. ○

7 The system provides security. ○

8 The system has integrity. ○

9 The system is dependable. ○

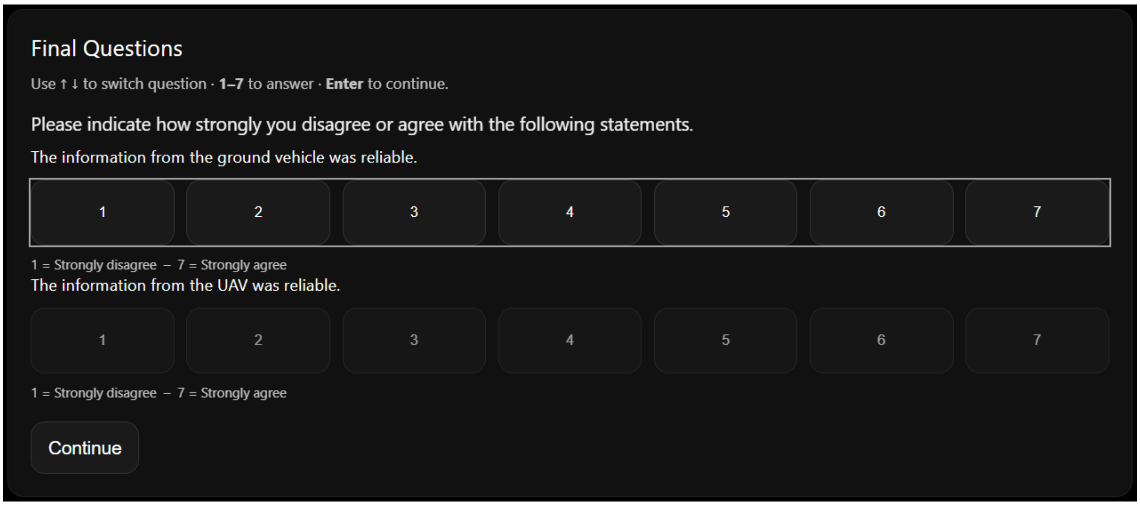
10 The system is reliable. ○

11 I can trust the system. ○

12 I am familiar with the system. ○

Figure A.13: Trust scale

Appendix I - Final questionnaire



Final Questions

Use ↑ ↓ to switch question · **1–7** to answer · **Enter** to continue.

Please indicate how strongly you disagree or agree with the following statements.

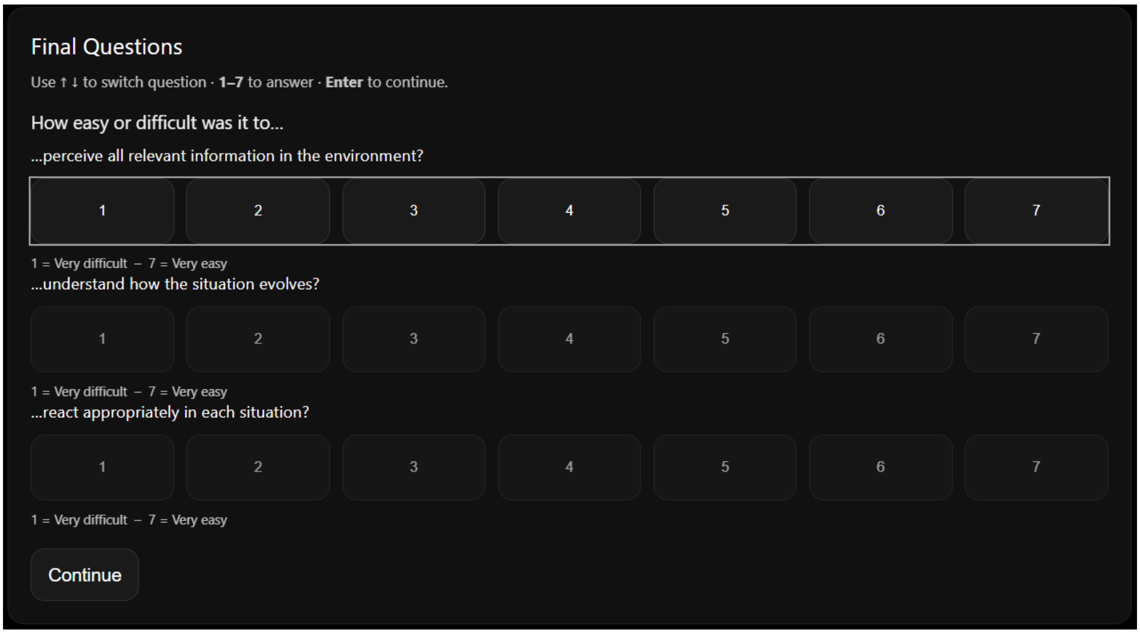
The information from the ground vehicle was reliable.

1 = Strongly disagree – 7 = Strongly agree

The information from the UAV was reliable.

1 = Strongly disagree – 7 = Strongly agree

Figure A.14: Questions related to reliability in the UGV and UAV information



Final Questions

Use ↑ ↓ to switch question · **1–7** to answer · **Enter** to continue.

How easy or difficult was it to...

...perceive all relevant information in the environment?

1 = Very difficult – 7 = Very easy

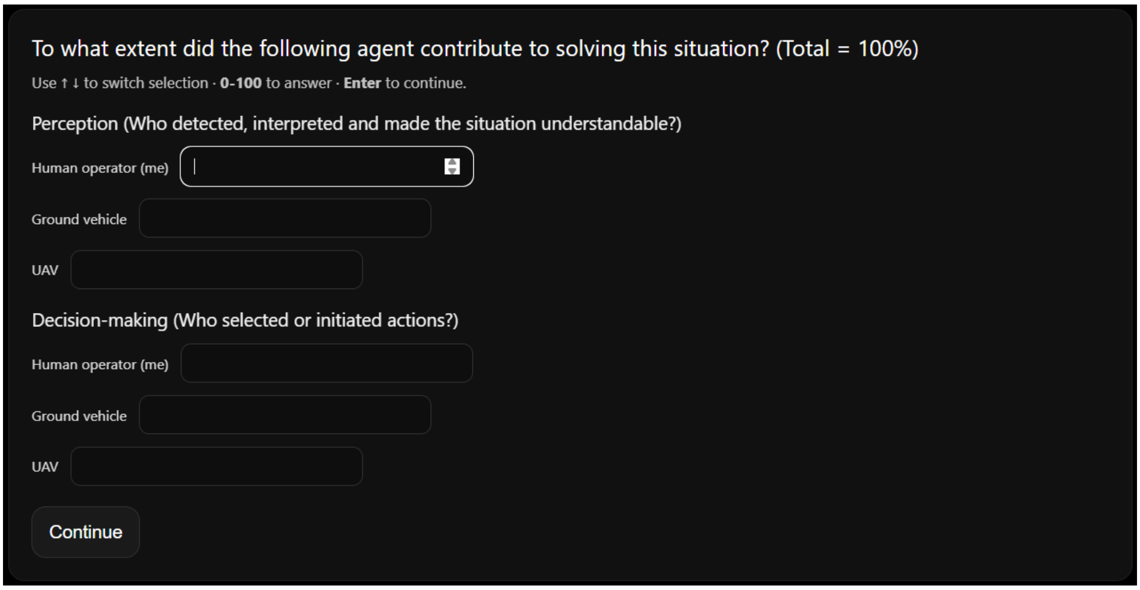
...understand how the situation evolves?

1 = Very difficult – 7 = Very easy

...react appropriately in each situation?

1 = Very difficult – 7 = Very easy

Figure A.15: Questions about situational awareness (perception, comprehension, and projection)



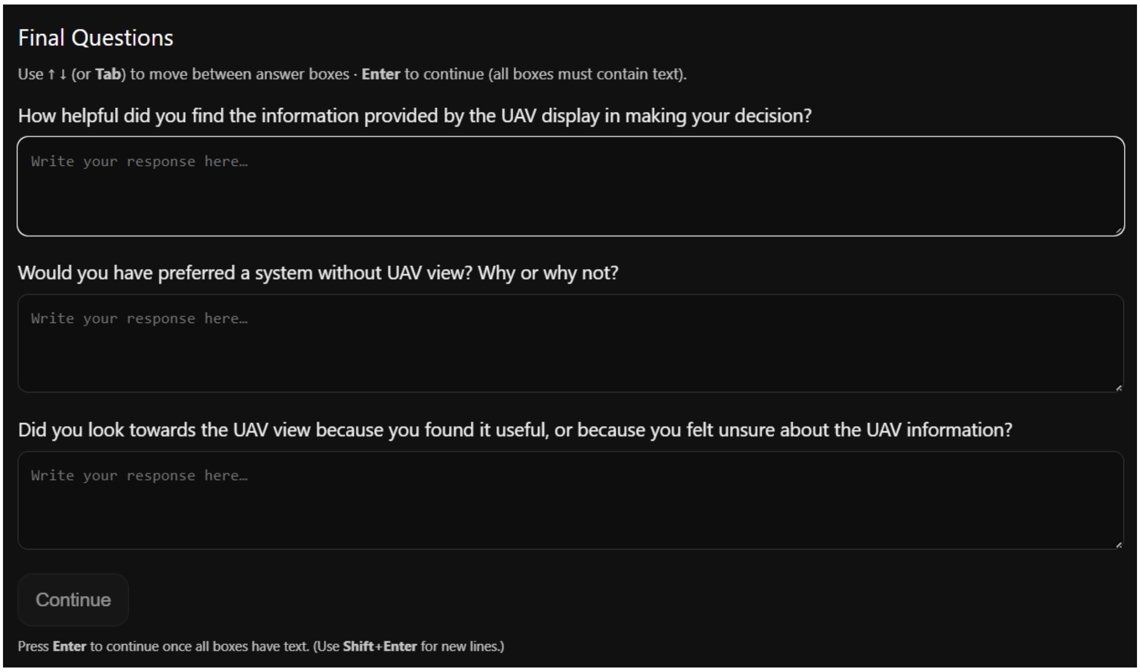
To what extent did the following agent contribute to solving this situation? (Total = 100%)

Use ↑ ↓ to switch selection · **0-100** to answer · **Enter** to continue.

Perception (Who detected, interpreted and made the situation understandable?)

Decision-making (Who selected or initiated actions?)

Figure A.16: Question about perception- and decision-making allocation



Final Questions

Use ↑ ↓ (or **Tab**) to move between answer boxes · **Enter** to continue (all boxes must contain text).

How helpful did you find the information provided by the UAV display in making your decision?

Would you have preferred a system without UAV view? Why or why not?

Did you look towards the UAV view because you found it useful, or because you felt unsure about the UAV information?

Press **Enter** to continue once all boxes have text. (Use **Shift**+**Enter** for new lines.)

Figure A.17: Qualitative questions regarding the UAV

Appendix J - Eye tracking validation

Figure A.18: First version of number validation

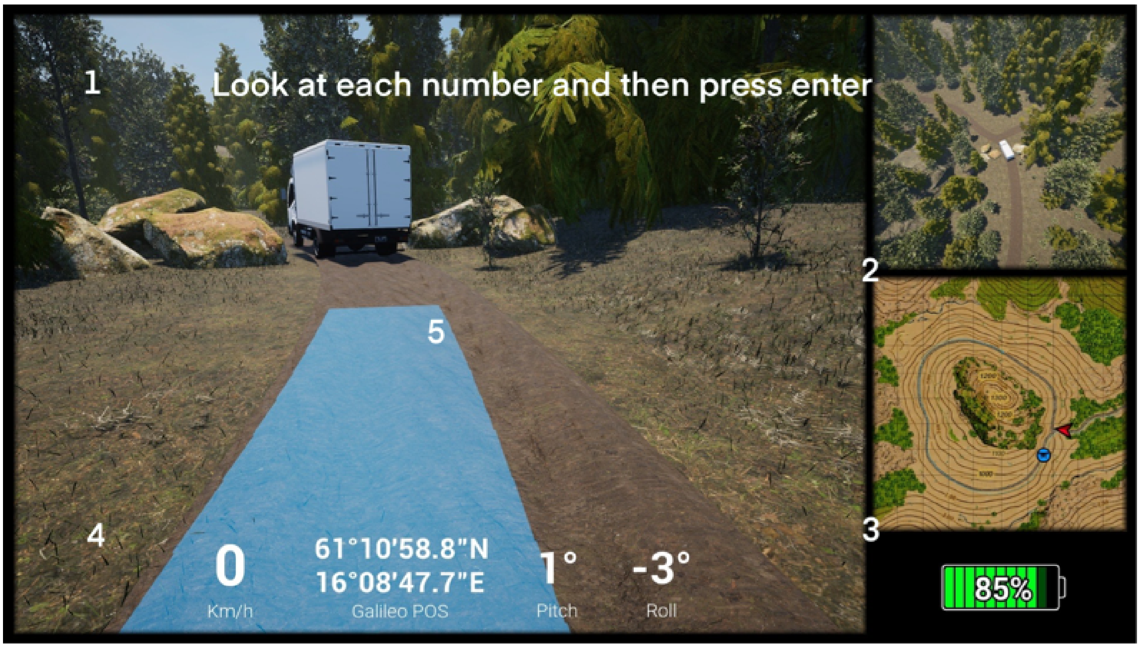
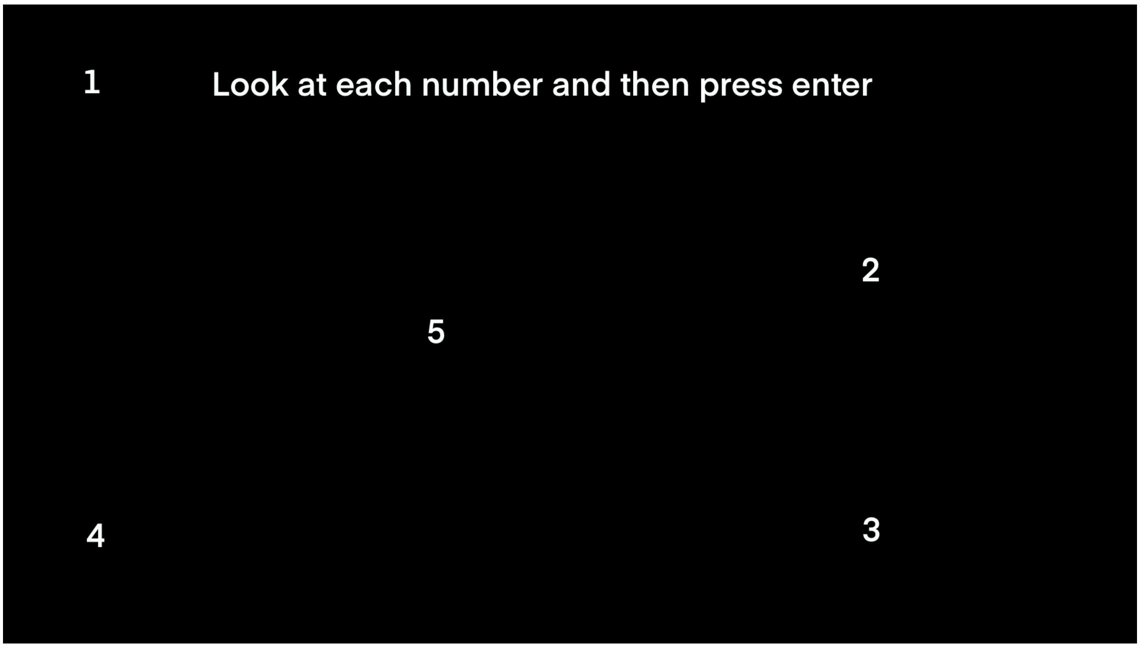

Look at each number and then press enter

Figure A.19: Positions of the new number validation



Look at each number and then press enter

Figure A.20: Final version of number validation

Appendix K - Mission start and outro

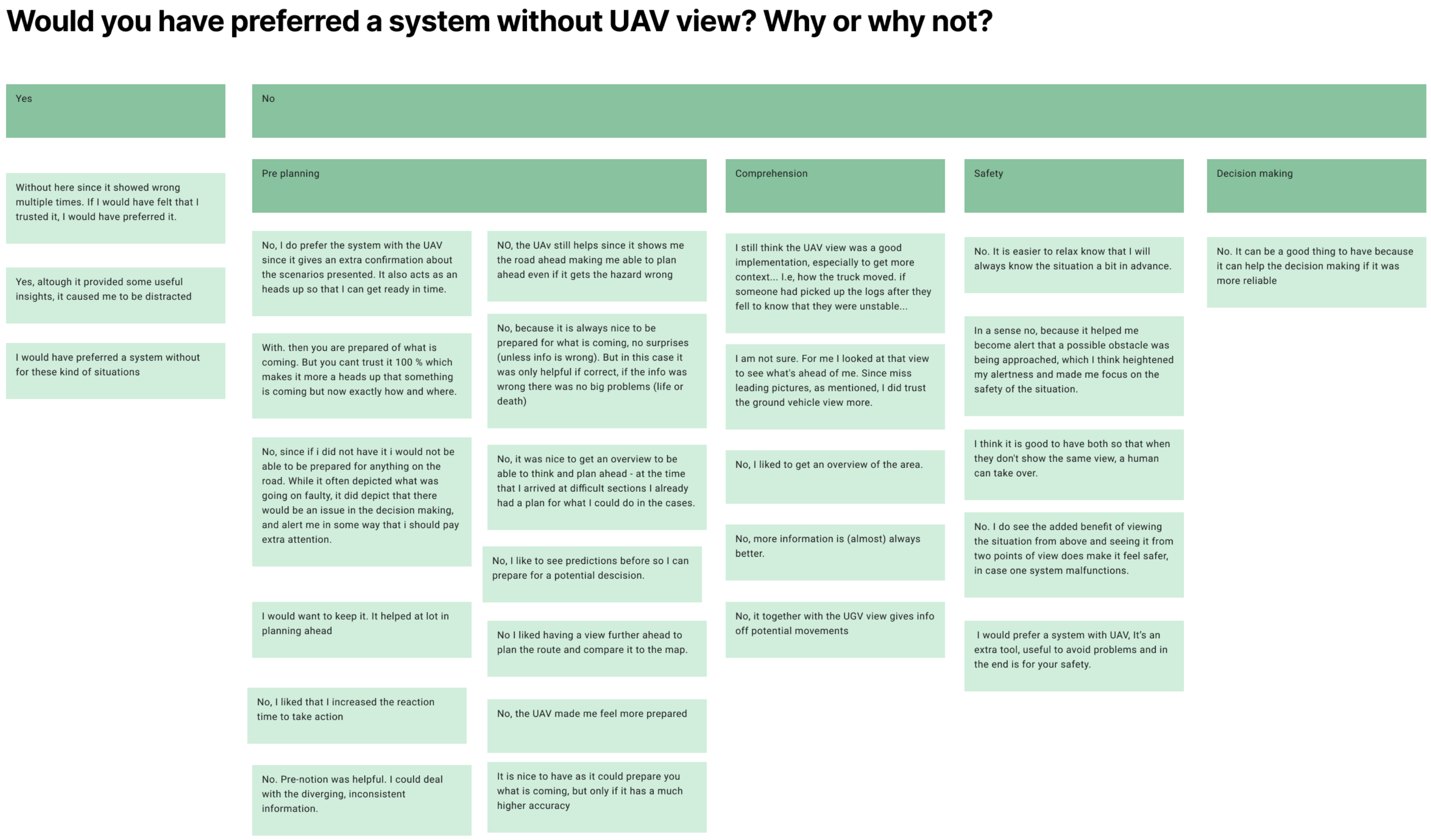
MISSION STARTED

Figure A.21: Mission started

MISSION COMPLETED!

Figure A.22: Outro with finding package

Appendix L - KJ-method



Would you have preferred a system without UAV view? Why or why not?

Yes

Without here since it showed wrong multiple times. If I would have felt that I trusted it, I would have preferred it.

Yes, altough it provided some useful insights, it caused me to be distracted

I would have preferred a system without for these kind of situations

No

Pre planning

No, I do prefer the system with the UAV since it gives an extra confirmation about the scenarios presented. It also acts as an heads up so that I can get ready in time.

With. then you are prepared of what is coming. But you cant trust it 100 % which makes it more a heads up that something is coming but now exactly how and where.

No, since if i did not have it i would not be able to be prepared for anything on the road. While it often depicted what was going on faulty, it did depict that there would be an issue in the decision making, and alert me in some way that i should pay extra attention.

I would want to keep it. It helped at lot in planning ahead

No, I liked that I increased the reaction time to take action

No. Pre-notion was helpful. I could deal with the diverging, inconsistent information.

NO, the UAv still helps since it shows me the road ahead making me able to plan ahead even if it gets the hazard wrong

No, because it is always nice to be prepared for what is coming, no surprises (unless info is wrong). But in this case it was only helpful if correct, if the info was wrong there was no big problems (life or death)

No, it was nice to get an overview to be able to think and plan ahead - at the time that I arrived at difficult sections I already had a plan for what I could do in the cases.

No, I like to see predictions before so I can prepare for a potential descision.

No I liked having a view further ahead to plan the route and compare it to the map.

No, the UAV made me feel more prepared

It is nice to have as it could prepare you what is coming, but only if it has a much higher accuracy

Comprehension

I still think the UAV view was a good implementation, especially to get more context... I.e, how the truck moved. if someone had picked up the logs after they fell to know that they were unstable...

I am not sure. For me I looked at that view to see what's ahead of me. Since miss leading pictures, as mentioned, I did trust the ground vehicle view more.

No, I liked to get an overview of the area.

No, more information is (almost) always better.

No, it together with the UGV view gives info off potential movements

Safety

No. It is easier to relax know that I will always know the situation a bit in advance.

In a sense no, because it helped me become alert that a possible obstacle was being approached, which I think heightened my alertness and made me focus on the safety of the situation.

I think it is good to have both so that when they don't show the same view, a human can take over.

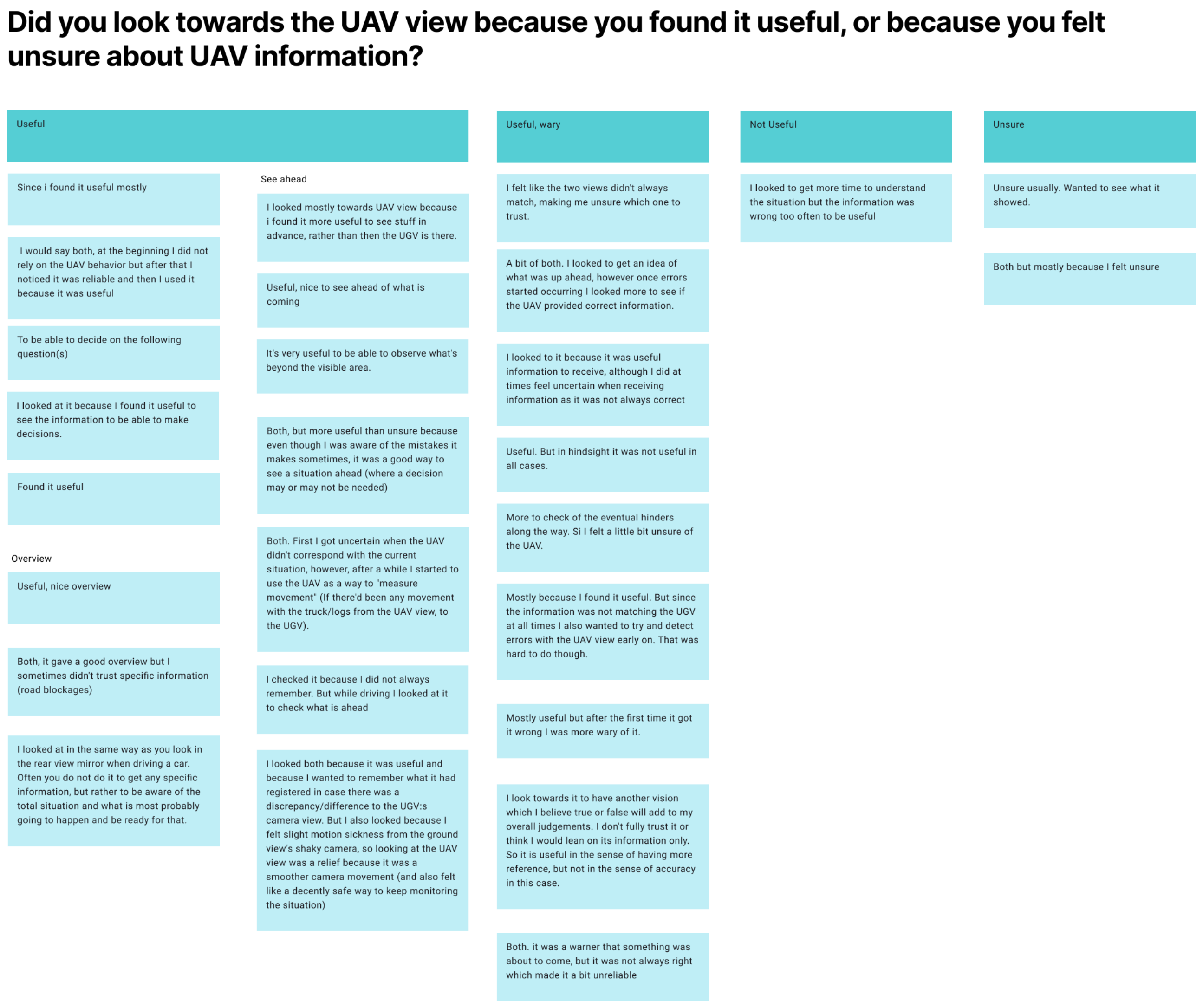
No. I do see the added benefit of viewing the situation from above and seeing it from two points of view does make it feel safer, in case one system malfunctions.

I would prefer a system with UAV, It's an extra tool, useful to avoid problems and in the end is for your safety.

Decision making

No. It can be a good thing to have because it can help the decision making if it was more reliable

Figure A.23: KJ-method Q1



Did you look towards the UAV view because you found it useful, or because you felt unsure about UAV information?

Useful

Since i found it useful mostly

I would say both, at the beginning I did not rely on the UAV behavior but after that I noticed it was reliable and then I used it because it was useful

To be able to decide on the following question(s)

I looked at it because I found it useful to see the information to be able to make decisions.

Found it useful

Overview

Useful, nice overview

Both, it gave a good overview but I sometimes didn't trust specific information (road blockages)

I looked at in the same way as you look in the rear view mirror when driving a car. Often you do not do it to get any specific information, but rather to be aware of the total situation and what is most probably going to happen and be ready for that.

See ahead

I looked mostly towards UAV view because i found it more useful to see stuff in advance, rather than then the UGV is there.

Useful, nice to see ahead of what is coming

It's very useful to be able to observe what's beyond the visible area.

Both, but more useful than unsure because even though I was aware of the mistakes it makes sometimes, it was a good way to see a situation ahead (where a decision may or may not be needed)

Both. First I got uncertain when the UAV didn't correspond with the current situation, however, after a while I started to use the UAV as a way to "measure movement" (If there'd been any movement with the truck/logs from the UAV view, to the UGV).

I checked it because I did not always remember. But while driving I looked at it to check what is ahead

I looked both because it was useful and because I wanted to remember what it had registered in case there was a discrepancy/difference to the UGV:s camera view. But I also looked because I felt slight motion sickness from the ground view's shaky camera, so looking at the UAV view was a relief because it was a smoother camera movement (and also felt like a decently safe way to keep monitoring the situation)

Useful, wary

I felt like the two views didn't always match, making me unsure which one to trust.

A bit of both. I looked to get an idea of what was up ahead, however once errors started occurring I looked more to see if the UAV provided correct information.

I looked to it because it was useful information to receive, although I did at times feel uncertain when receiving information as it was not always correct

Useful. But in hindsight it was not useful in all cases.

More to check of the eventual hinders along the way. Si I felt a little bit unsure of the UAV.

Mostly because I found it useful. But since the information was not matching the UGV at all times I also wanted to try and detect errors with the UAV view early on. That was hard to do though.

Mostly useful but after the first time it got it wrong I was more wary of it.

I look towards it to have another vision which I believe true or false will add to my overall judgements. I don't fully trust it or think I would lean on its information only. So it is useful in the sense of having more reference, but not in the sense of accuracy in this case.

Both. it was a warner that something was about to come, but it was not always right which made it a bit unreliable

Not Useful

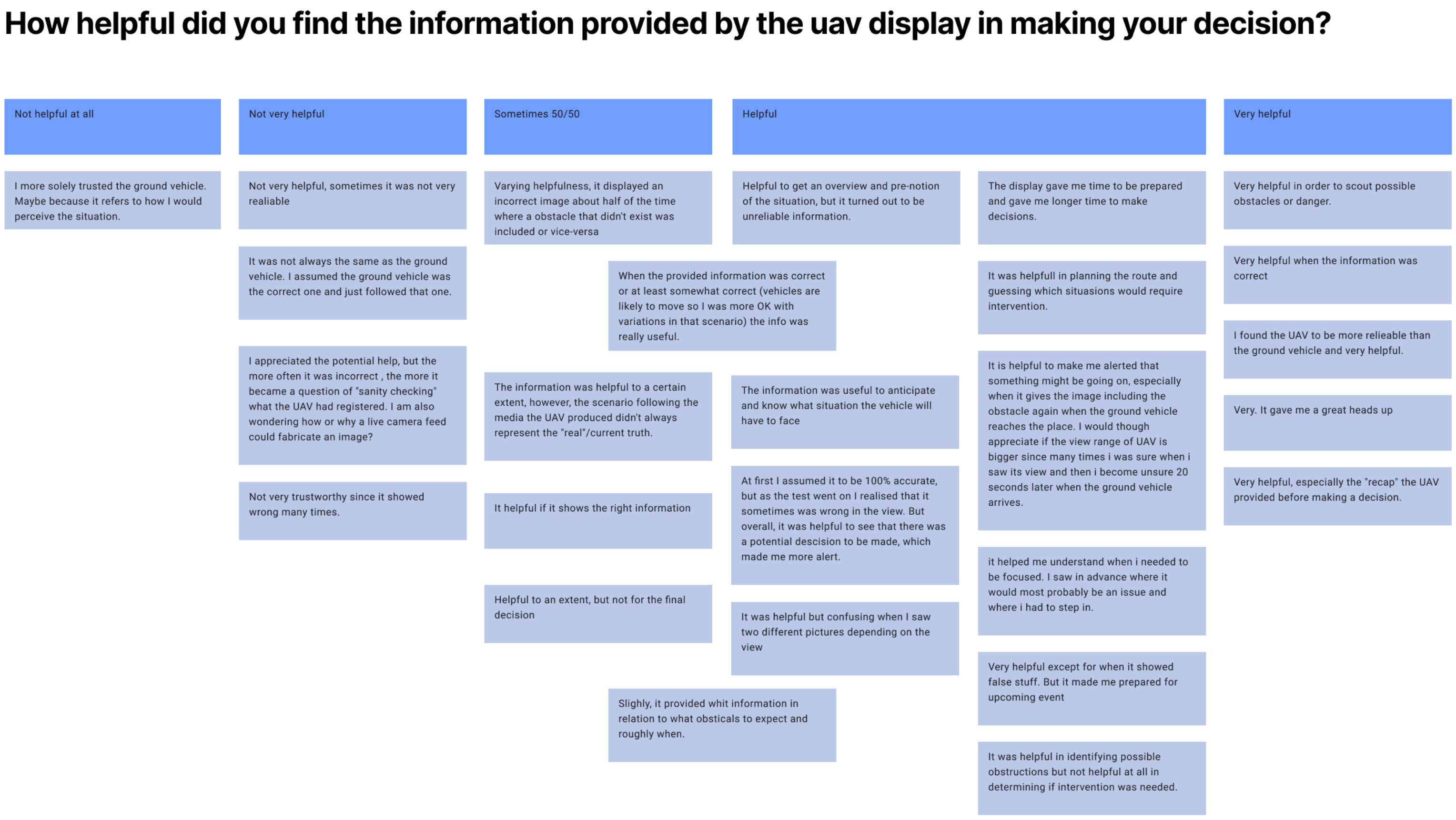
I looked to get more time to understand the situation but the information was wrong too often to be useful

Unsure

Unsure usually. Wanted to see what it showed.

Both but mostly because I felt unsure

Figure A.24: KJ-method Q2



How helpful did you find the information provided by the uav display in making your decision?

Not helpful at all

I more solely trusted the ground vehicle. Maybe because it refers to how I would perceive the situation.

Not very helpful

Not very helpful, sometimes it was not very realiable

It was not always the same as the ground vehicle. I assumed the ground vehicle was the correct one and just followed that one.

I appreciated the potential help, but the more often it was incorrect , the more it became a question of "sanity checking" what the UAV had registered. I am also wondering how or why a live camera feed could fabricate an image?

Not very trustworthy since it showed wrong many times.

Sometimes 50/50

Varying helpfulness, it displayed an incorrect image about half of the time where a obstacle that didn't exist was included or vice-versa

When the provided information was correct or at least somewhat correct (vehicles are likely to move so I was more OK with variations in that scenario) the info was really useful.

The information was helpful to a certain extent, however, the scenario following the media the UAV produced didn't always represent the "real"/current truth.

It helpful if it shows the right information

Helpful to an extent, but not for the final decision

Slighly, it provided whit information in relation to what obsticals to expect and roughly when.

Helpful

Helpful to get an overview and pre-notion of the situation, but it turned out to be unreliable information.

The information was useful to anticipate and know what situation the vehicle will have to face

At first I assumed it to be 100% accurate, but as the test went on I realised that it sometimes was wrong in the view. But overall, it was helpful to see that there was a potential descision to be made, which made me more alert.

It was helpful but confusing when I saw two different pictures depending on the view

The display gave me time to be prepared and gave me longer time to make decisions.

It was helpfull in planning the route and guessing which situasions would require intervention.

It is helpful to make me alerted that something might be going on, especially when it gives the image including the obstacle again when the ground vehicle reaches the place. I would though appreciate if the view range of UAV is bigger since many times i was sure when i saw its view and then i become unsure 20 seconds later when the ground vehicle arrives.

it helped me understand when i needed to be focused. I saw in advance where it would most probably be an issue and where i had to step in.

Very helpful except for when it showed false stuff. But it made me prepared for upcoming event

It was helpful in identifying possible obstructions but not helpful at all in determining if intervention was needed.

Very helpful

Very helpful in order to scout possible obstacles or danger.

Very helpful when the information was correct

I found the UAV to be more relieable than the ground vehicle and very helpful.

Very. It gave me a great heads up

Very helpful, especially the "recap" the UAV provided before making a decision.

Figure A.25: KJ-method Q3

Appendix M - Brainstorming

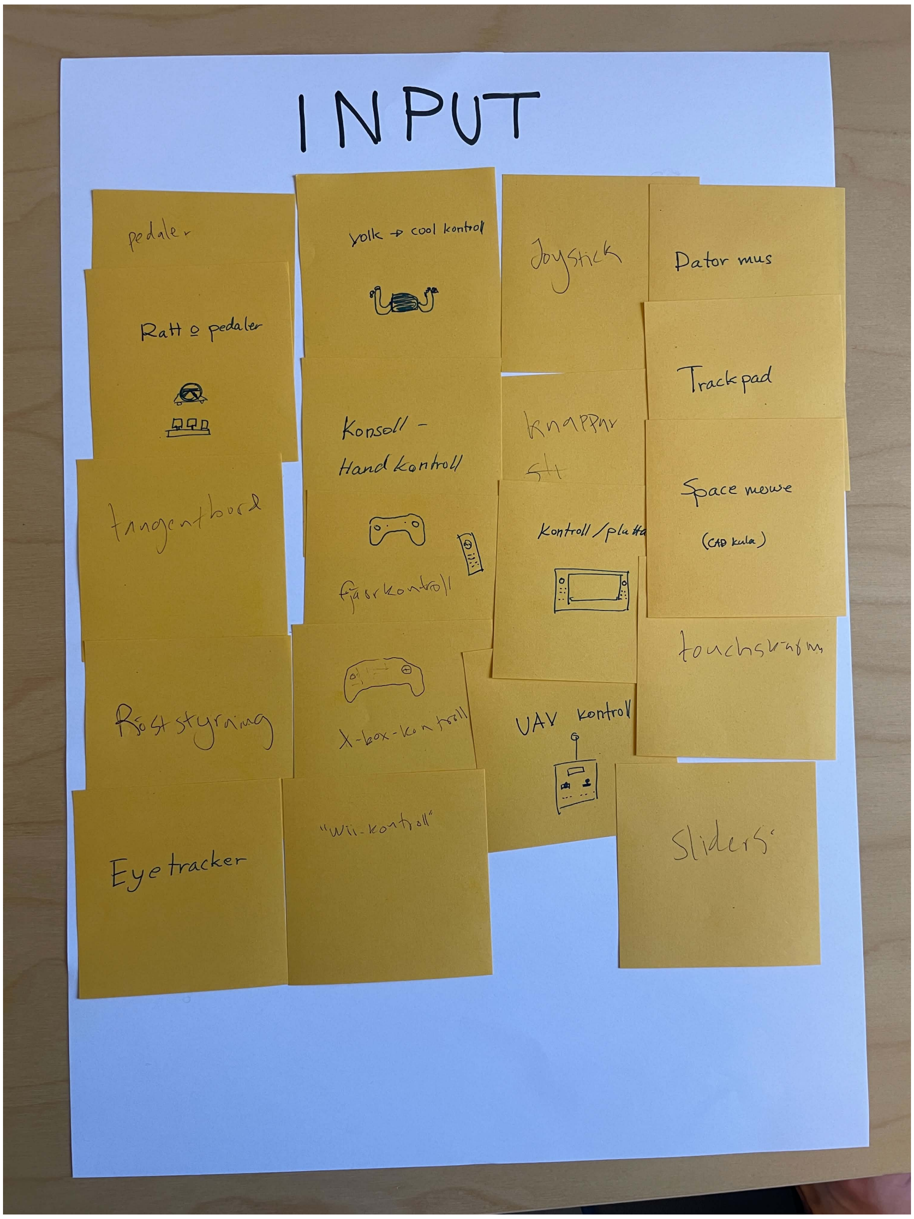


Figure A.26: Brainstorming ideas - Input

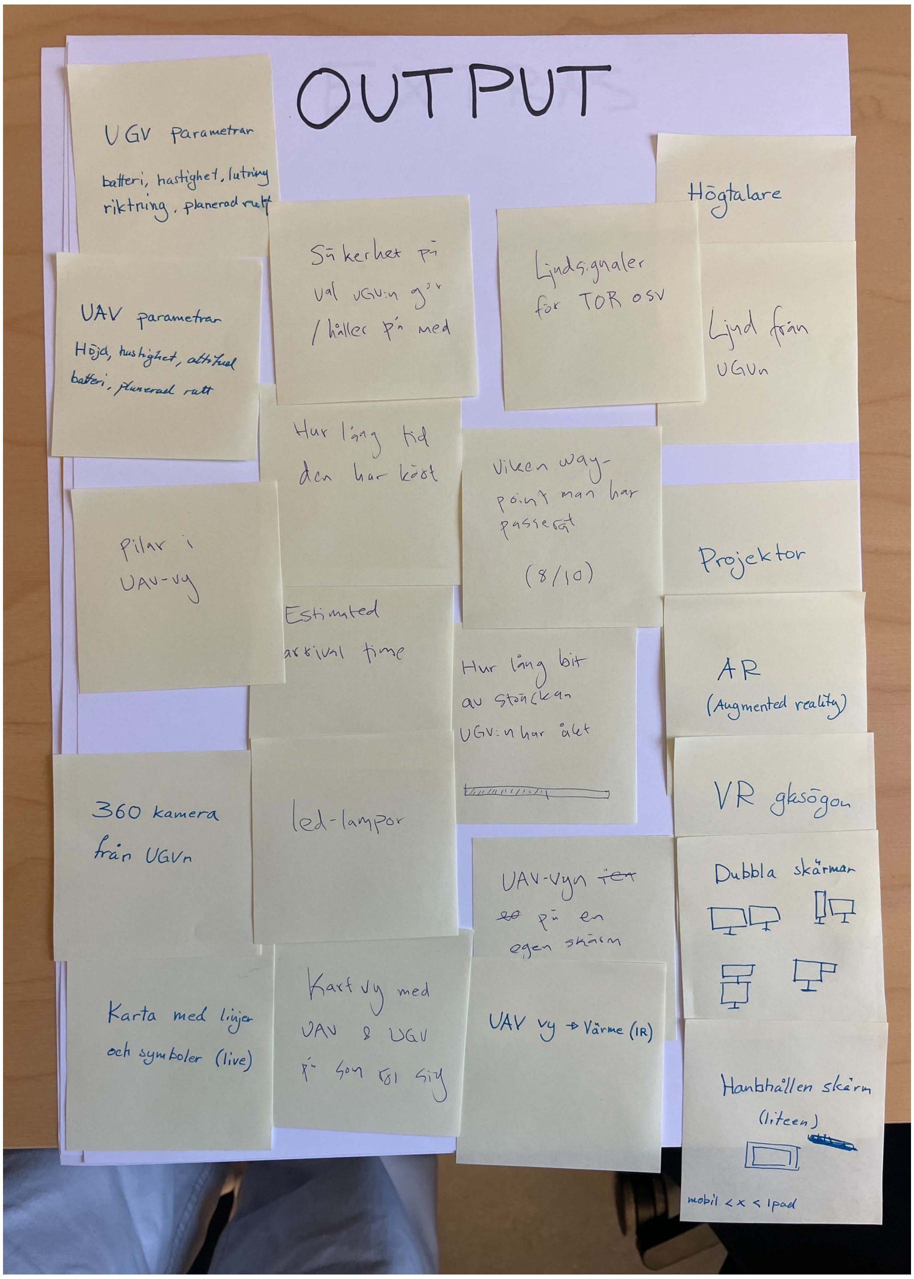


Figure A.27: Brainstorming ideas - Output

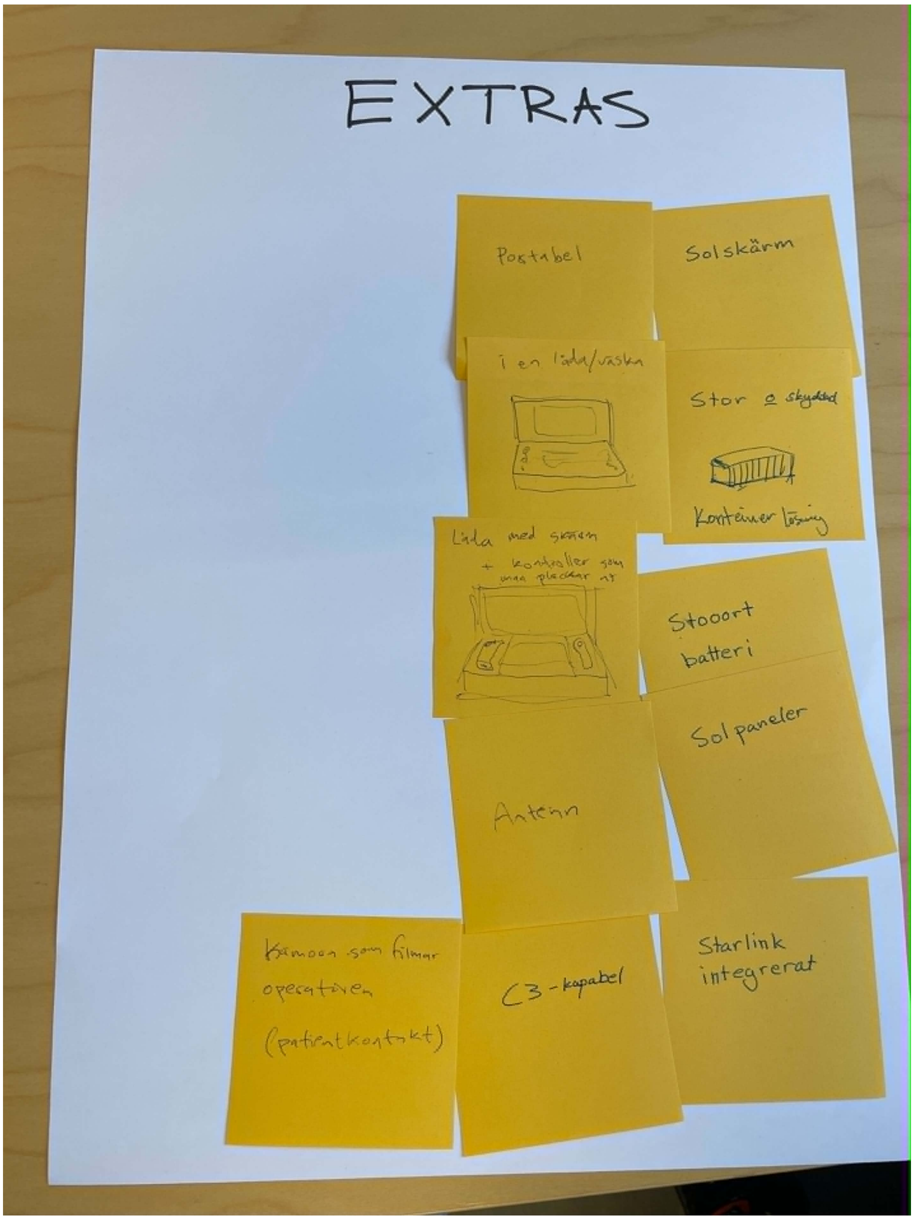


Figure A.28: Brainstorming ideas - Extras

Appendix N - Sketch design concept

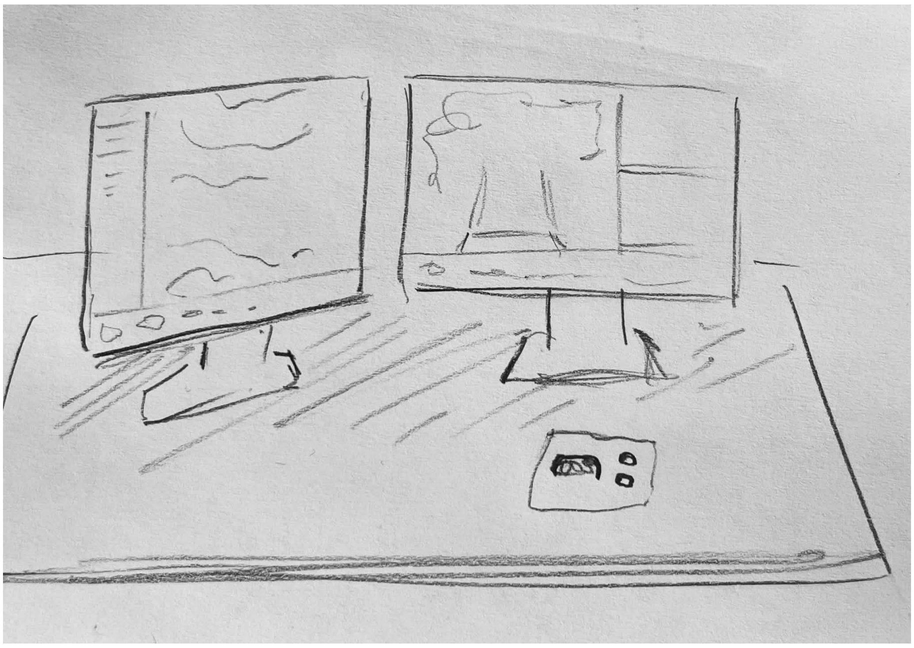


Figure A.29: Concept sketch

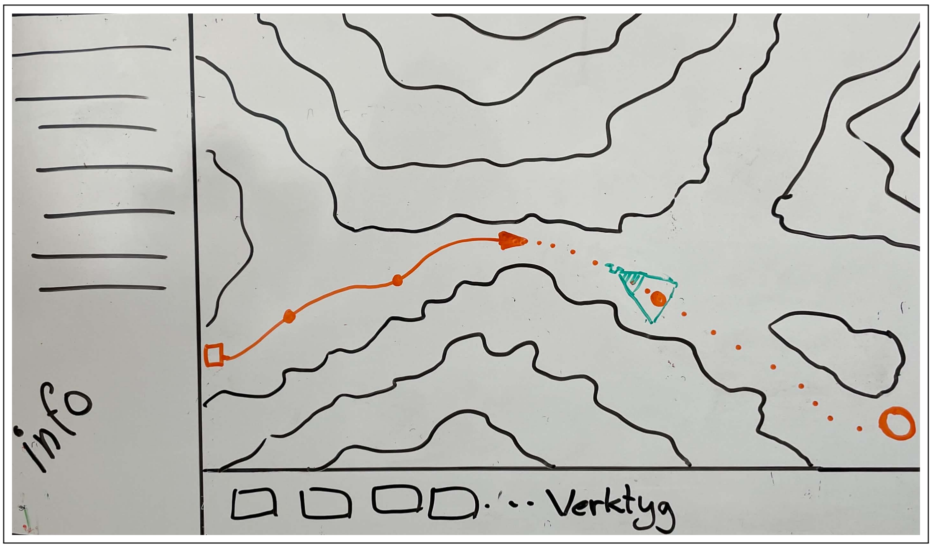


Figure A.30: Concept sketch - Left screen

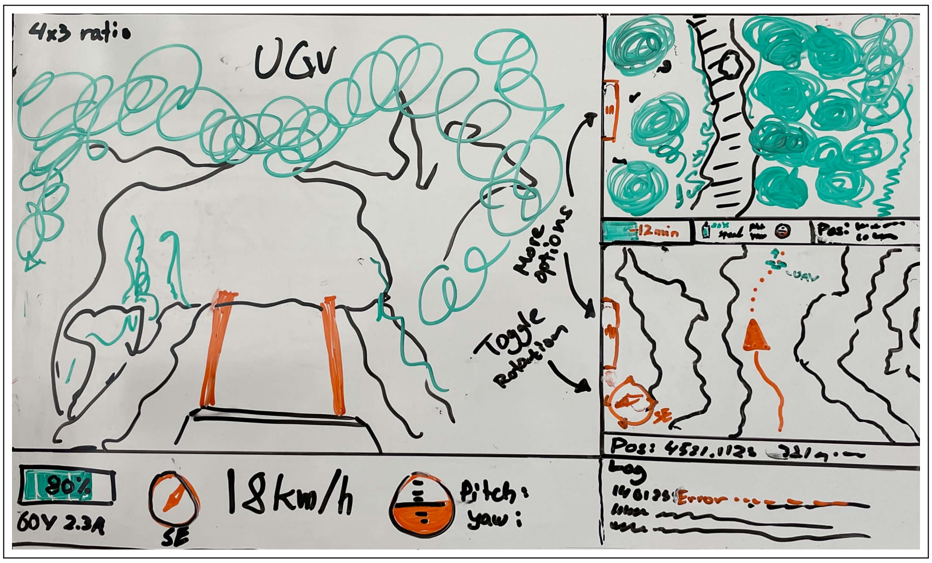
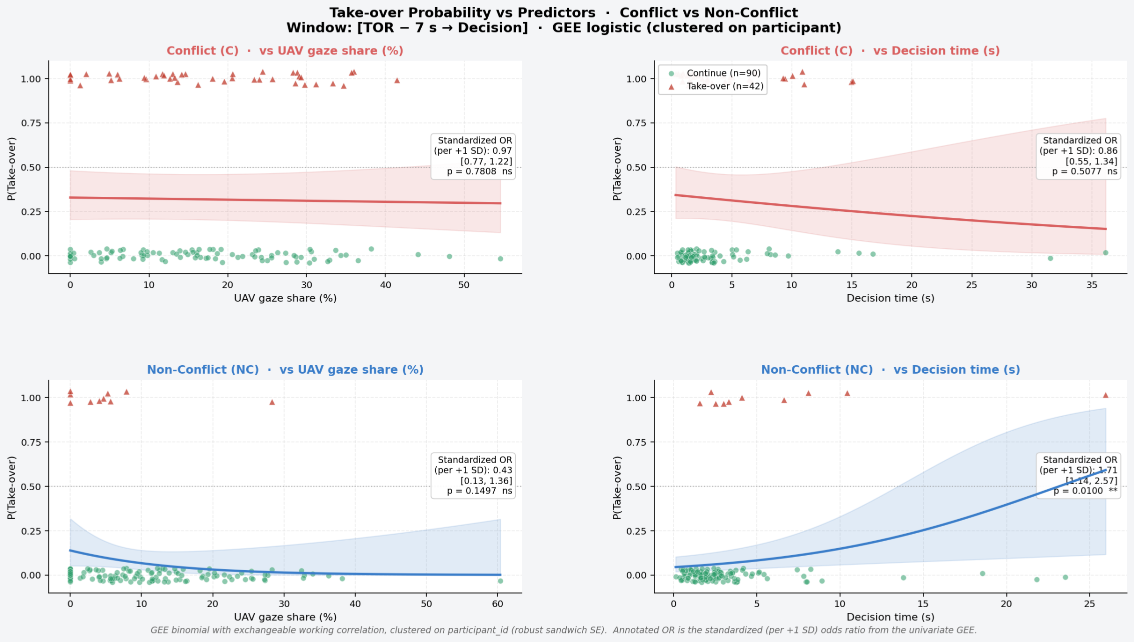


Figure A.31: Concept sketch - Right screen

Appendix O - Take-Over Probability - Further Analysis

To further isolate the effect of UAV gaze share in decision-making, gaze share vs decision times' effect on take over probability was analyzed using uni- and multivariate GEE logistic regression, accounting for the non-independent observations with a Bayesian binominal GLMM as a sensitivity check.

The univariate GEE shows a strong correlation in NC between increased Decision Time and likelihood of Take-over and indicates that an increase in UAV gaze share correlates to a lower likelihood of takeover, which is considered the *"correct"* decision[1], see Figure A.32.



GEE binomial with exchangeable working correlation, clustered on participant_id (robust sandwich SE). Annotated OR is the standardized (per +1 SD) odds ratio from the univariate GEE.

Figure A.32: Decision x UAV Share x Time - Forest Plot

The multivariate analysis shown in Figure A.33 supports the trend presented above, but to a lesser degree. This is confirmed by the sensitivity check in Table A.2.

Table A.2: Bayesian binomial GLMM - Random Intercept Per Participant

	C UAV %	C Time	NC UAV %	NC Time
β	+0.005	-0.261	-1.246	+0.871
OR	1.005	0.770	0.288	2.388
95% CI	[0.58, 1.73]	[0.45, 1.32]	[0.13, 0.65]	[1.49, 3.82]
p post	0.9862	0.3431	0.0028	0.0003

---

[1] Note: NC model only has 10 Take-overs so it's statistically thin.

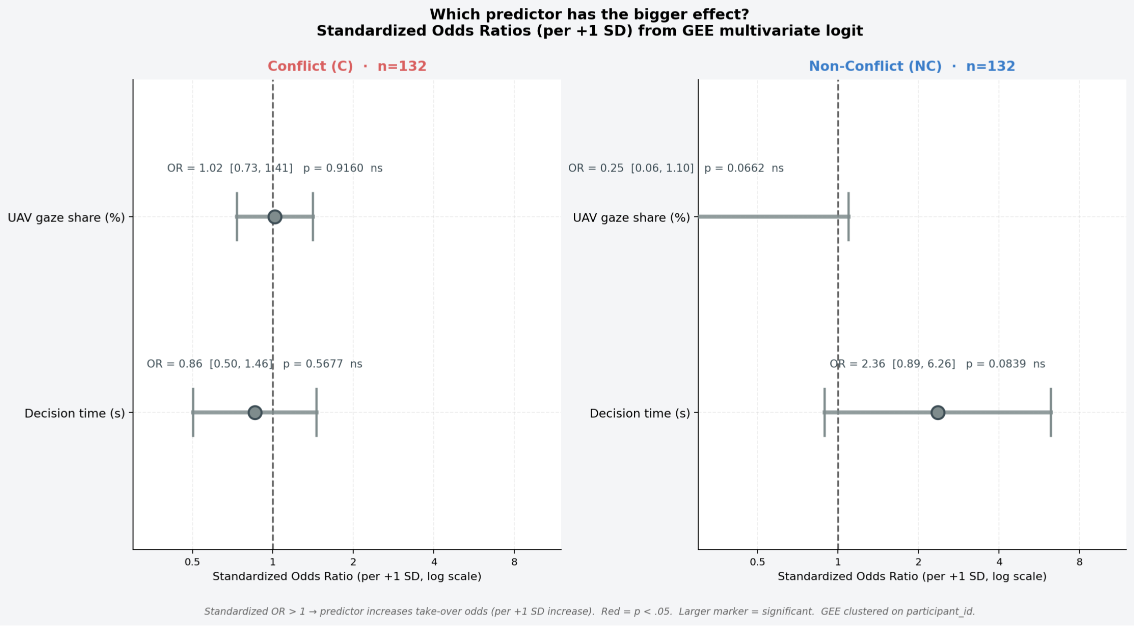


Figure A.33: Decision x UAV Share x Time - Forest Plot

Splitting the analysis into Truck/Log scenario and NC/C conditions while using the expanded decision window results in two interesting cases:

Truck NC in Figure A.34 shows an increased UAV share correlating ($p = 0.092$) to a decrease in time. This is interesting since, as previously mentioned, Truck NC does not show the UAV image before the TOR. This could be the result of participants who scanned the UAV on their own accord having a better SA and therefore making faster decisions, but this needs further research.

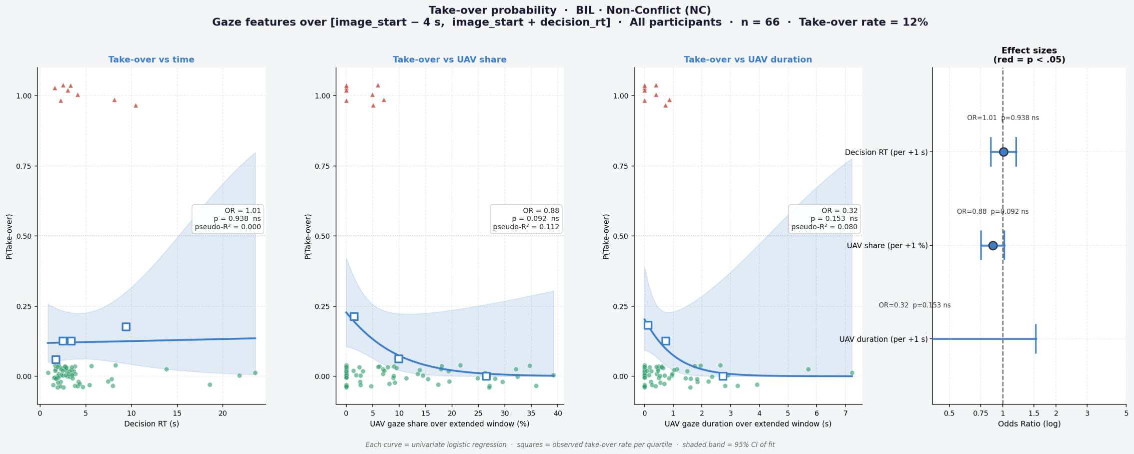


Figure A.34: UAV Share x Time x Decision - Truck NC

Logs C in Figure A.35 shows a break from the trend as decision time clearly dominates ($p = 0.016$) UAV share in take-over probability, in stark contrast to the Conflict trend during the short decision window presented in Figure A.33.

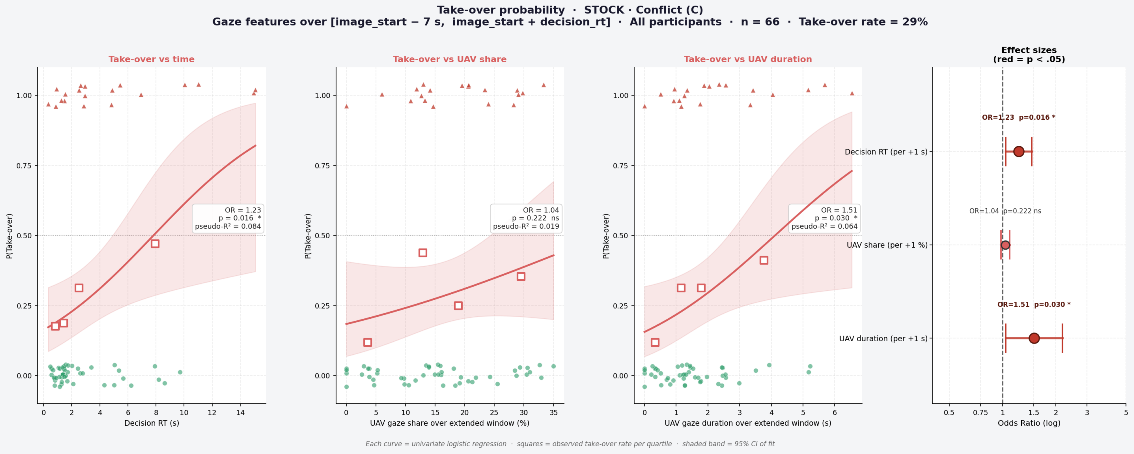


Figure A.35: UAV Share x Time x Decision - Log C

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY