Supported scheduling of unmanned vehicles for military operations

A.S.C. Schutter

TU Delft

ALTEN

TNO innovation for life

Supported scheduling of unmanned vehicles for military operations

by

A.S.C. Schutter

to obtain the degree of Master of Science
at the Delft University of Technology,
Faculty of Electrical Engineering, Mathematics and Computer Science.
Specialization: Computer Science

Student number:	4260724	
Defence in public on:	November 1, 2016	
Responsible Professor:	Prof. Dr. M.A. Neerincx,	TU Delft, TNO
External Supervisors:	Dr. ir. V.F. van Ravesteijn,	Alten Nederland
	Dr. ir. H.J.L.M. Vullings,	TNO
Thesis committee:	Prof. Dr. M. A. Neerincx,	TU Delft, TNO
	Dr. ir. W-P. Brinkman,	TU Delft
	Dr. S. G. Lukosch,	TU Delft

Abstract

The role of military operators is changing towards supervision of multiple unmanned vehicles. This places new demands on the operators that require supportive systems. This study focused on the design of a system to support the user to allocate high-level tasks to unmanned vehicles for military operations and that respects the objectives and restrictions of the user. The domain analysis resulted in an abstraction decomposition space for the domain and nine proposals of patterns for interaction design. This includes a method to align the user and system objectives and to provide additional transparency of the system about a proposed schedule's compliance with the user objectives. The effect of the additional transparency was investigated in a user study using a prototype of the system on user performance, situational awareness (SA), trust and cognitive task load. However, there was insufficient statistical power to conclude that the additional transparency, compared to a condition without additional transparency, nor the order of these conditions, had a direct or interaction effect on the participants, except for a learning effect on SA. Besides other interesting effects, the experiment showed that the order of the conditions with and without additional transparency had an effect on the use of it and a positive attitude of participant towards the system resulted in automation-induced compliance.

Keywords: Military domain, system transparency, aligning objectives, unmanned vehicles, human factors, situational awareness, scheduling

Foreword

This study was done in cooperation with Alten Nederland and TNO. In this report, I combined my military and technological experience and knowledge to lay a foundation for future development of supportive systems in the military domain. The choice for the military domain was obvious for me. Before I started this study, I graduated from the Royal Military Academy and served as a platoon commander of the combat engineers in a mechanize brigade. After several years, I decided to make a step towards development of technology, and here is my first result.

This report describes the analysis for the design of supportive systems in the military domain. The report proposes nine interaction-patterns for the design of a system to support a military operator with the allocation of high-level tasks to multiple unmanned vehicles. The user values, objectives and constraints for the operation are key in the design. Besides, the report lays a theoretical foundation that could be a basis for future development of supportive systems in the military domain. The results might also be helpful for other domains such as urban search and rescue (USAR).

The project required over a year of work and resulted in a full prototype of a supportive system called eSupport, build upon a Battlefield Management System, the CommonSense Framework, developed by TNO. The software is open for use and development by third parties. I am glad to present this report to you, I hope that the findings will proof to be useful as a basis for future development.

Acknowledgement
First, I would like to thank those involved of Alten Nederland for their trust and their patience to await the completion of this project, despite my pursuit of expansion of the project. Also many thanks for those involved of TNO Waalsdorperweg, TU Delft and the military forces (JISTARC and DMO) for their interest, support and the useful suggestions.

At last, I would like to thank everybody who supported me, who has been a worthwhile sounding board and who reminded me that this project is not my life's work. So many thanks, Stephanie, Richard, Paula, Steffan, Emma, Linda, fellow students of the 12th floor, Chateau Burgwal, fellow Slopend athletes and those who I did not mention.

Disclaimer
Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of TNO or Alten Nederland.

A.S.C. Schutter
Delft, October 2016

Contents

List of Figures

List of Tables

List of Abbreviations

AC	Attentional control
BML	Battlefield management language
BMS	Battlefield management system
CAT	Combined arms team
CTL	Cognitive task load
DMO	Defensie materieel organisatie (Defence materiel organisation)
EOD	Explosive ordnance disposal
FAC	Forward air controller
ISTAR	Intelligence, Surveillance, Target acquisition and Reconnaissance
IVES	INSITU video exploitation system
JISTARC	Joint ISTAR commando
JTAC	Joint terminal attack controller
LDP	Landmacht doctrine publicatie (Army doctrine publication)
PACT	People, activities, context, technology (and human factors)
RTLX	Raw task load index
SA	Situational awareness
SAGAT	Situational awareness global assessment technique
SART	Situational awareness rating technique
sCE	Situated cognitive engineering
SME	Subject matter expert
STANAG	Standardization agreement
STUV	Small tactical unmanned vehicle
TLX	Task load index (NASA)
TNO	Toegepast-natuurwetenschappelijk onderzoek (company)
UAS	Unmanned aerial system
UAV	Unmanned aerial vehicle
UGV	Unmanned ground vehicle
USAR	Urban search and rescue
UV	Unmanned vehicle
VGE	Video game experience
VSD	Value sensitive design

1 Introduction

"Intelligent agents and unmanned systems with greater autonomy will change the military landscape. However, it will not make the Soldier's task easier, it will only change the type of tasks and problems he or she will face" (Barnes et al., 2014).

Unmanned systems are currently used for many military applications, supporting soldiers in tasks such as load-carrying, hazardous material detection, casualty extraction, firefighting, building mapping/clearing and counter explosive devices (Chen and Barnes, 2012b). Therefore they are becoming an essential part of the battlefield (Chen et al., 2011). In the future the battlefield may be the theater of dozens of unmanned vehicles (UVs) performing their tasks concurrently. This changes the role of UVs in military operations. However, current UVs are mostly controlled remotely (teleoperated) by a single or more operators, so the increasing number of UVs requires a lot of man power to control these systems. Although the use of more UVs keeps more military personnel out of harm's way, it increases operating and training cost and it challenges the demand for more UV operations (Clare, 2013). Therefore, a lot of research focuses already on enabling a single operator to control multiple heterogeneous (air, sea, land) vehicles and thus on decreasing the operator-to-vehicle ratio. This requires the UVs to operate more autonomously.

Autonomous UVs have been developed rapidly over the last years. The UVs therefore require less direct control of the human operator, allowing the operators to shift their role from teleoperation towards supervision. The supervisory tasks can be compared with the tasks of an executive and include setting higher level goals, intervention in cases of emergencies or malfunctions and decisions when lives are at risk (Barnes et al., 2014; Chen et al., 2007). The changing role of the operator towards supervision of multiple vehicles places new demands on the operators.

One of the responsibilities of the supervising operator is scheduling tasks (Chen et al., 2011), for this task the operator has to allocate tasks to the UVs while satisfying goals and constraints. Various algorithms are already available to execute these scheduling tasks autonomously. However, fully autonomous scheduling algorithms do not always perform well in highly unpredictable environments. This is because they depend on an algorithm defined during design time and are therefore often not able to deal with unforeseen circumstances appropriately. Besides, these systems leave the operator out of the loop of the scheduling process and this negatively effects the operator's understanding about for example the situation and the reasoning behind the schedule (Linegang et al., 2006). A potential approach to deal with these shortcomings is to let the operator cooperate with the automated scheduler during the scheduling process (Clare et al., 2015). This concept of collaboration between a human and an optimization algorithm to solve a complex problem is referred to as human-automation collaboration, but also known by other names such as human-computer collaboration, human guided algorithms and mixed-initiative planning (Clare et al., 2015). Collaborative tools typically use the strengths of both humans and autonomy, for example the experience and pattern recognition skills of the human and the strength of the autonomy to solve optimization problems (Cummings et al., 2007; Clare et al., 2012c).

Previous studies showed that this kind of collaboration can actually improve the performance of both the human and the automation, however they also showed that supportive tools have certain effects on the user which are not always predictable (Cummings et al., 2007).

1.1. Background

Several human-automation collaboration tools for unmanned vehicle scheduling have been developed in previous studies and the effects on the users were tested. A tool to support users for mission planning and execution for the US Navy' Autonomous Operations future Naval Capability - Intelligent Autonomy program was investigated by Linegang et al. (2006). The evaluation of the system showed that subject matter experts (SMEs) often have difficulties with understanding the reasoning of the automation behind a mission plan and therefore question its accuracy and effectiveness (Chen et al., 2014). The authors proposed a method to give more insight to the user about the automation's reasoning, so to be more transparent (Chen et al., 2014). The authors therefore predefined three objectives of the system (see 1.1.1) and designed the user interface to show the impact of the environmental threats to these objectives. However, the results were not promising, mainly because the participants had different interpretations of the (importance of) these objectives and the impact of the threats on them. Linegang et al. (2006) concluded that the key challenge was to align the human's conceptualization of the mission planning problem with that of the automation system. In other words, to align the goals of the user with those of the system.

From these results, we hypothesize that the combination of goal alignment and system transparency is key to improve the user's understanding of the system's behavior. Various studies about these topics are discussed in the following sections.

1.1.1. Alignment of human and system goals

One of the key research areas within human-automation interaction that has not been investigated in detail, is understanding how users should and could express their desires to an automated planner to ensure alignment of the objective functions of the human and automation (Clare et al., 2012b). Human specified-goals and constraints can be used to align the user's and system's goals (objectives), resulting in a goal-driven mission planning system (Linegang et al., 2006). In such a system, the user specifies their goals and constraints for a specific mission and the automation computes an optimal solution that satisfies these goals and constraints.

It is important not only to consider the explicit user goals (so the explicit tasks), since human and automation typically also have implicit goals. Human implicit goals are for example the goals about risk (e.g. risk is accepted due to time pressure) and fuel consumption (e.g. optimize fuel consumption). The implicit goals of automation are typically introduced during design time, a scheduling agent for example might always attempt to finish a schedule as soon as possible. The automation's goal to finish as soon as possible is considered implicit, because this goal was never explicitly stated and does not necessarily comply with the user's goals. So the implicit goals of the user and the automation do not necessarily overlap, this means that the automation can compute an optimal schedule with its own set of goals which does not fit the user's set of goals. This can result in confusion and decreased trust of the user in the system (Clare et al., 2012b,c).

Linegang et al. (2006) propose a system to align the human and system's goals with three predefined (explicit and implicit) purposes of the system: 1. Intelligence, 2. Asset preservation and 3. Secrecy. The first purpose characterizes the (*intelligence*) objectives in a mission (the actual reconnaissance tasks) relative to the other two more abstract purposes (so *asset preservation* and *secrecy*). In the study, the automation considered these three objectives to find an optimal solution to complete the intelligence tasks.

However, according to Cummings and Bruni (2010); Clare et al. (2012c), a better solution, so a solution that fits better to the user's objectives, should not be expected using predefined goals, since these goals are identified and set by the system designers during design time. In a complex and changing environment, an "optimal" solution to satisfy explicit and implicit pre-programmed goals might not hold once the dynamics of the environment change (Cummings and Bruni, 2010). An automated system has to be able to adjust to the changing dynamics of the environment and the user's objectives in real time. This requires interaction between the system and the user during the process and a collaborative system should facilitate this (Clare et al., 2012c).

To allow users to indicate their goals during the process, Clare et al. (2012b) evaluated a method in which the user was able to indicate for five (implicit) objective parameters whether they are important or not. The parameters they used were:

- Area coverage: It is more important to cover as much area as possible for search tasks than executing search tasks set by the user;
- Search/loiter tasks: It is important to execute the tasks set by the user;
- Target tracking: It is important to track targets;
- Hostile destruction: It is important to destroy hostiles;
- Fuel efficiency: It is important to use fuel efficiently by moving at cruise velocity.

The method did not lead to improvements in performance. The authors concluded that the five parameters might not have been the best possibilities and the binary choice between important or not important might not have been sufficient. The current study elaborates on these finding for the design of a new system to better align the human' and system' goals.

1.1.2. System transparency

A transparent system supports the user's understanding of the system's behavior. Chen et al. (2014) defined transparency as "the descriptive quality of an interface pertaining to its abilities to afford an operator's comprehension about an intelligent agent's intent, performance, future plans, and reasoning process". Chen et al. presented a model in which system's transparency is defined at three levels, the authors based this on the model of Endsley (2011) for situational awareness. Providing transparency at the right levels should improve the user's comprehension of the agent behavior and their estimation of the system's reliability (i.e. the user's mental model of the system and its behavior). The first level comprises the basic information about the system, so the purpose of the system (the current goals), the process, the current performance and the status. Second level information describes the why of the system, and helps the user to understand the reasoning of the system for a suggestion and the constraints that exist (such as limitations, given constraints, risk, trade-offs, etc.) related to the planning process. Level three information helps the user to understand expected future states and limitations.

Agent transparency promotes an appropriate level of trust of the user in the system, which typically results in better automation usage. When the user's trust is calibrated, it is likely that the user will accept (or not adjust) a correct solution or will reject (or correctly adjust) a wrong solution. Distrust of the user in the system would more likely lead to a rejection (or adjustment) of a correct solution, and overtrust more likely leads to acceptance of a wrong solution. It is also possible that the user makes an error by for example misinterpretation. The trust levels can also be expressed in terms of compliance of the user to the system's suggestions and solutions. **Table 1.1** shows, for all four levels of trust, the expected decision of the user for correct and incorrect suggestions of the automation. The last column about compliance provides the terminology for a correct or incorrect decision for a correct or incorrect suggestion of the system. The current report uses this terminology.

Better agent transparency ultimately leads to improved human-agent team performance (Chen et al., 2014). However, as the amount of information presented to the user increases, this might have implications and result in higher cognitive task load of the user (Chen et al., 2014). Cognitive task load might however also decrease due to more transparency since the user does not have to speculate about the reasoning of the system (Mercado et al., 2016).

In Mercado et al. (2016) the transparency level necessary to create an effective and transparent interface to support human-agent teaming was investigated. The task of the operator in the study was to

Table 1.1: Expected decision of users, based on the level of trust and the correctness of the automation suggestion, adapted from (Chen et al., 2014)

Automation's state of accuracy	**Trust**	**User's decision of automation usage (AUD)**	**Compliance**
Correct suggestion	Calibrated trust	Accepts	Appropriate compliance
Correct suggestion	Distrust	Overrides, incorrectly adjusts	Inappropriate noncompliance
Incorrect suggestion	Calibrated trust	Overrides, correctly adjusts	Appropriate noncompliance
Incorrect suggestion	Overtrust	Accepts	Inappropriate compliance
Incorrect suggestion	Error	Overrides, incorrectly adjusts	Erroneous noncompliance

cooperate with the system to make a schedule for multiple heterogeneous UVs, similar to the current study. Before each mission the intent and goals were presented to the participant prior to the scheduling and the system calculated the most optimal task allocation. The system presented two alternative courses of actions (schedules), one best and one second best solution according to the system. However, the "best" solution was actually worse than the second best for three out of eight missions (so less than 70% of the suggestions were correct, this is the reliability under which automation may not be beneficial to performance). The evaluation suggested that additional transparency improved overall human-agent team performance and usability and led to more appropriate compliance with the agent's recommendations, as a result of more effective trust calibration, without increasing cognitive task load or response time.

1.2. Research objectives and questions

Clare et al. (2012b) showed that automation can outperform humans in scheduling tasks due to their computational capacity . However, the use of full autonomy to create a schedule typically leaves users without understanding of the reasoning behind the schedule. This typically results in users that either overrule the system or that follow the suggestions of the system without checking for mistakes. The previous section describes two important reasons why user's do not understand the reasoning of the system for a solution. First, the objectives of the user and the system for the allocation of tasks to UVs are often not aligned and second, the user lacks understanding of the system's considerations for the suggestion.

The aim of the current study is to investigate how the transparency of a supportive tool for scheduling high-level tasks to multiple unmanned vehicles, from now on called eSupport, can be improved and thereby improve the overall performance of the user. In the study, a military operator, responsible for the supervision of multiple UVs, will be the user of the system. The performance is divided in several components. The first component of performance is appropriate (non-)compliance, this means that the user accepts correct solutions or rejects incorrect solutions of eSupport. The second component is efficiency, this is the time the user requires to create a schedule. The last component is situational awareness (SA), this is the level of knowledge and understanding of the user about the situation and the implications for the future.

Since the study focuses on improving the transparency of eSupport, the user's performance is measured relative to the system without improved transparency. Therefore, an improvement of performance is defined in terms of: 1. no increase of scheduling time (efficiency), 2. no decrease in appropriate (non-)compliance and 3. no decrease of SA. The main research question of the study is:

How can an automated supportive system (eSupport), used for scheduling high-level tasks to multiple heterogeneous unmanned vehicles, provide additional transparency about a proposed schedule's *compliance with the user's objectives, and thereby improve the user's performance?*

eSupport supports the user in order to achieve a schedule that satisfies the user's objectives and constraints. Therefore the objectives of eSupport have to be aligned with the user's objectives. The user's objectives however depend on the specific domain for which the system is used. The military domain will be the focus of the current study to answer the research question. This leads to the following research objective:

Identify the user's main objectives considering the allocation of high-level tasks to heterogeneous unmanned vehicles within a military operation and communicate the objectives to eSupport. Align the objectives of eSupport with the user's objectives and constraints in order to allow eSupport to reason about a schedule's compliance with these objectives. To achieve this objective, the following research questions have to be investigated:

Research question 1a: What are the user's main objectives considering the allocation of high-level tasks to heterogeneous unmanned vehicles within a military operation?

Research question 1b: How can the user's main objectives defined in research question 1a be communicated to eSupport in order to allow eSupport to reason about a schedule's compliance with these objectives?

Besides aligning the system's objectives with the user's objectives, eSupport should aim to provide more transparency about the compliance of a schedule to the user's objectives to ultimately calibrate the user's trust in the system's behavior better. The second research objective is:

eSupport provides transparency about the schedule's compliance with the user's objectives to increase user performance in terms of efficiency, appropriate (non-)compliance and situational awareness. This leads to the following research questions:

Research question 1c: How can eSupport provide more transparency to the user about a schedule's compliance with the user's objectives?

Research question 1d: What is the effect of the method to provide additional transparency found in research question 1c on the efficiency, appropriate (non-)compliance and situational awareness?

1.3. Methodology

The main research question requires a user study with a system to support the user with the allocation of high-level tasks to unmanned vehicles. Such a system is not available for the current study. Therefore a prototype of eSupport will be designed and build for the current study. The system will be designed to support a specific group of users, the military operators. Military operators have undergone a special training, have specific experiences and therefore might consider specific things important in life, these things are called their values (Friedman et al., 2013). To properly design a system that complies with the requirements of a specific group of users, it is important to focus on these values.

A value sensitive design (VSD) method takes the user values into account throughout the design process in a principled and comprehensive manner (Friedman et al., 2013). sCEthics (de Greef et al., 2013) is a design practice that incorporates VSD into another methodology that supports to design systems for cognitive processes in specific situations: the situated Cognitive Engineering methodology (sCE methodology) (Neerincx and Lindenberg, 2008). The sCE methodology supports to specify system requirements by a structured requirements elicitation methodology using three components: the foundation, the specification and the evaluation, such as shown in Figure 1.1.

The foundation phase comprises the analysis of the operational, human factors, and technological demands. The specification component comprises the specification of the requirements together with a design rationale to justify the requirements. The evaluation component comprises the evaluation of the results to improve the design rationale. The sCE Methodology is an iterative process of design, creation, evaluation and refinement. sCEthics integrates VSD in the sCE Methodology in three ways.

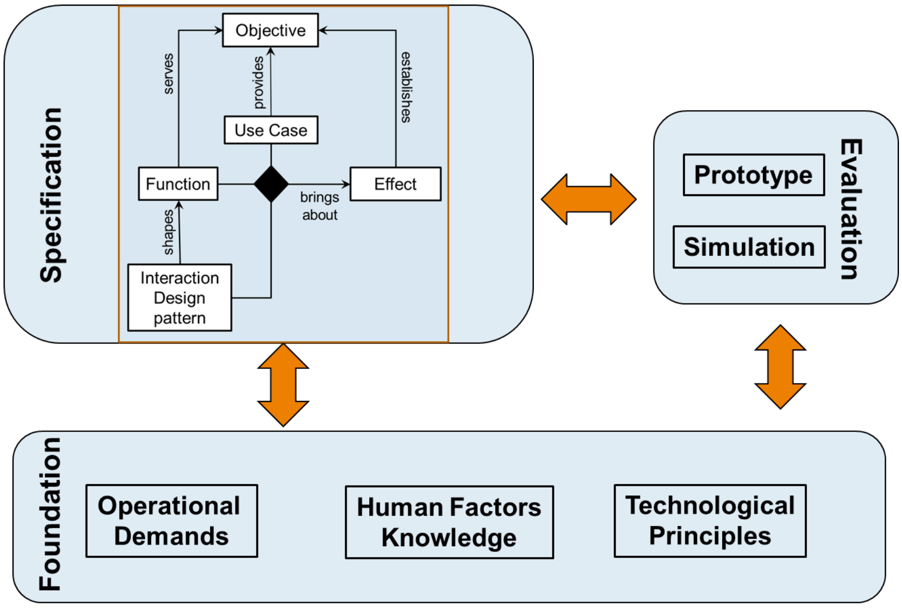


Figure 1.1: Overview of the sCE Methodology

First, it links the stakeholder analysis of VSD with the analysis of the operational demands. Second, it links the values of VSD with the human factor considerations. At last, it links the technical investigation of VSD with the technological demands of sCE (de Greef et al., 2013). The structure of sCEthics is followed for the design in this report.

1.3.1. Foundation

The Foundation comprises the analysis of the current situation for which the system is intended. To analyze the domain in a comprehensive matter for the design, the PACT framework is used (Benyon, 2010). PACT stands for People, Activities, Context and Technology. The analysis concludes with the Human Factors analysis. The analysis is done by literature research and three interviews with military subject matter experts.

The People section describes the analysis of the direct and indirect stakeholders together with their values. The direct stakeholders are those people who use the system directly, the indirect stakeholders do not use the system directly but are still affected by it. The second section describes the activities the UV-team will possibly be used for in future operations. The environment of the activities is described in the third section, the military structure for the analysis of the environment is maintained for this analysis. The fourth section focuses on the currently existing and relevant technology. In the final section the Human Factors are discussed, going more in depth about the known effects of the use of supportive systems on users and human-automation collaboration. The analysis provides a solution for research question 1a, 1b and 1c, together with a set of requirements for eSupport.

1.3.2. Specification and Design

The specification section first describes realistic scenarios in which the intended system can be used. More specific cases within these scenarios, called use cases, describe the interaction between the user and the system in more detail together with their individual actions. The use cases provide more insight in the functional requirements of the system. Subsequently, all user values are supported with one or more claims, these claims justify the requirements with the user values they support.

The requirements form the basis for the design of eSupport together with existing interaction design principles. For the design, we propose several interaction design patterns. Interaction design patterns are design principles to support the generation of consistent solutions for generic recurring human-computer interaction problems that have proven to be effective (TNO, 2012). These patterns enforce re-use of knowledge of former research about human-automation interaction. Since the patterns we propose have not been proven to be effective yet, we refer to these prototypes of interaction design patterns as proto-patterns.

A prototype of eSupport is build from the design, this prototype is suitable to do experiments with participants.

1.3.3. Evaluation

The prototype of eSupport will be used for a user study. Each participants has to work with eSupport to allocate high-level tasks to multiple unmanned vehicles for several scenarios. eSupport thereby takes the user objectives and the constraints into account to produce an optimal schedule. In some mission, eSupport will produce a wrong solution on purpose.

The results of the questionnaires and the actual user performance during the scenarios are used to evaluate the effect of additional transparency and the order of the conditions with and without additional transparency on the participants in different scenarios. We specifically focus on the effects on the performance, so the scheduling duration, appropriate (non-)compliance, the rating of compliance and SA, together with trust and CTL. Furthermore, the perceived usability of eSupport is measured for further development of eSupport and to check for possible side effects.

1.4. Document outline

Chapter 2 describes the foundation analysis for the design of eSupport and the specification is described in chapter 3. The evaluation of the prototype is thereafter covered in chapter 4 and concludes with several conclusions and recommendations for further research.

2 Domain Analysis

This chapter describes the analysis of the operational domain for eSupport. The analysis leads to requirements that will ultimately be used for the design of eSupport. The analysis follows the PACT principle: People, Activities, Context, Technology and concludes with Human Factors. The phases of the analysis are described in this order to encourage reuse of the results of the analysis for similar systems in the future. The required assumptions and constraints for the design are listed and validated in a later stadium. These assumptions can be validated once there is more experience and knowledge about the use of UVs in the military domain.

This chapter starts with narrowing down the scope of the domain in Section 2.1. Section 2.2 describes the relevant actors in the domain and the stakeholders of eSupport are identified together with their values. Section 2.3 goes more into depth about the activities within the domain that eSupport might have to deal with, described at different levels of abstraction. The context of the domain that eSupport might have to deal with is analyzed in Section 2.4 to gain more insight in other factors of importance. Section 2.5 provides an overview of relevant technology for the design. The chapter concludes with an analysis of the Human Factors in Section 2.6.

2.1. Scope

A detailed domain analysis and value sensitive design requires a specifically defined user group and domain. It is however not clear yet how unmanned vehicles (UVs) will be integrated in the military hierarchy in the future. To make well grounded assumptions about the application of UVs for military operations in the future, the current practices for the deployment of UVs in the Dutch Army are investigated. Two interviews with military experts (see appendix A) were done to gain more insight. One of the interviews was done with a captain of the Joint ISTAR commando, 106th intelligence squadron, who served in an intelligence unit (intelligence module) responsible for gathering intelligence during a deployment. The other interview was done with a captain of Joint ISTAR commando, 107th Aerial Systems battery, responsible for Unmanned Aerial Systems (UAS) flight operations, so for the deployment of a remote controlled UV called ScanEagle.

The first interview gave more insight in the current operational procedures for intelligence modules, e.g. for planning intelligence operations, managing the intelligence requirements and gathering and analyzing intelligence information. From this interview we conclude that intelligence operations typically have a long planning horizon up till days, involve multiple units with broad capabilities and cover a wide area of operations. Due to these characteristics, the challenge is not the complexity but rather the prioritization of tasks, therefore it is not likely that intelligence modules would use a supportive system for task allocation.

The second interview gave more insight in the capabilities of the ScanEagle and the procedures before and during an operation. To deploy the ScanEagle, a lot of conditions have to be created and a team of multiple soldiers is required. This team is among others composed of a vehicle operator in control of the vehicle, the payload (camera) operator and an IVES (INSITU Video Exploitation System; actors that analyze the videos in real time and make the intelligence report/product). Concluding, the vehicles require a lot of personnel, Murphy and Burke (2010) also concluded this for the USAR domain. Besides, since the ScanEagle has a larger area of operation and therefore has a high cruise altitude, a lot of conditions (for example for airspace deconfliction) have to be created before the system can actually be deployed. These conditions hinder a fast deployment of the UV and decrease the need for a supportive system for the allocation of the tasks.

The interviewees indicated that tactical units such as reconnaissance, manoeuvre and commando units already use smaller UVs such as the Raven. Up to date, these vehicles are remotely controlled, but they are smaller, easier to use and require a very short preparation time. These type of units have a much smaller area of responsibility than an intelligence module and require no long-term planning. These units can decide to use these vehicles last minute. Due to these characteristics, eSupport could be a helpful support for these types of units for the deployment of their smaller tactical unmanned vehicles.

The mentioned tactical units still vary much in their operating procedures and therefore the scope has to be narrowed down even more. While reconnaissance units and commando squads have very specific goals and operational procedures, a combined arms team (CAT) has less specific operational procedures since it is composed of various units. A CAT is composed of a manoeuvre unit that can provide firepower such as infantry, extended with additional units such as a tank troop (typically four tanks), Joint Terminal Attack Controllers (JTAC, formerly FAC, coordinator of Close Air Support), combat engineers, forward artillery observers, Explosive Ordnance Disposal (EOD), combat service support (signals, transportation and medical), etc. It depends on the characteristics of each operation such as the type of threat, specific required capabilities or other reasons if and what units are embedded to the CAT. An advantage of a CAT is that the different types of units can utilize each others knowledge and capabilities to successfully accomplish an operation. The CAT commander typically is in command of these embedded units and the manoeuvre units can protect them. Since a CAT is meant to combine different units to utilize each others capabilities, it seems logical to embed a team of unmanned vehicles to a CAT to deploy these vehicles. The current study assumes that the team of unmanned vehicles, from now on referred to as a Small Tactical Unmanned Vehicles (STUV)-team, is embedded to a CAT (DA0001).

The STUV-team is defined in the current study as a team composed of several heterogeneous (aerial and ground) vehicles under responsibility of one operator and likely several additional human enablers (DA0002). Up-to-date there are no such teams in the Dutch military forces (and the existence of operational teams is not reported so far) and thus there are no common practices nor knowledge about procedures or compositions. Section 2.5.2 goes into more detail about the types of vehicles that the STUV-team contains in the current study.

2.1.1. Conclusions

This section narrowed down the scope for the design. The assumptions for the design of eSupport that followed from this section are listed below. The current study focuses on the design of eSupport to support the operator of a STUV-team, the activities of other actors are not considered in the current study (DC0001).

- **DA0001**: A team of UVs can be embedded to a combined arms team (CAT) during future military operations
- **DA0002**: The UVs are combined in a team, an operator is responsible for the unmanned vehicles and likely several additional human enablers

- **DC0001**: The study focuses on the design of eSupport to support the commander of a UV-team with their related activities

2.2. People

This section starts with a description of the actors and the communication in a CAT. Thereafter, the direct and indirect stakeholders of eSupport are identified with their most important values.

2.2.1. Actors and information requirements

The manoeuvre component of the CAT in the current study is an organic infantry company. An infantry company is basically composed of a company staff and three infantry platoons. The captain is in command of the company and responsible for the operation. Subordinate units and embedded units however have their own responsibilities and specialties. These units are responsible for decisions at their own area of expertise. Therefore all levels of the hierarchy have specific information requirements, this makes it extremely important that information goes up and down rapidly through the chain of command. In current operations, a Battle Management System (BMS) is used to share information between soldiers about the environment and the available intelligence. More details about BMS are provided in Section 2.5.1. The current study assumes that intelligence requirements and information about the operation are communicated through and displayed on BMS and eSupport has full access to this information (DA0003).

2.2.2. Stakeholders

The stakeholders of a system in general are those actors that are directly or indirectly affected by the system. A direct stakeholder has a direct interaction with the system and can as such experience direct advantages or consequences of the system. Indirect stakeholders are not directly interacting with the system, but still experience the advantages or consequences. The direct and indirect stakeholders are identified in this section to determine each actor's respective interest in the system. The focus of the stakeholder analysis in the current study is on the interaction with eSupport for the allocation of tasks to unmanned vehicles. Figure 2.1 gives an overview of the interactions between the relevant actors and systems.

Figure 2.1 shows that eSupport only focuses on the operator to support their scheduling activities. Therefore, the only actors that directly interact with eSupport are the operators, these are the direct stakeholders of the system. The interaction with BMS nor the automation (controller agent) that controls the autonomous UVs are included in the analysis of the current study (DC0002). Actors that depend on the schedule, for example because they have intelligence requirements or the UVs are going to be in their area of operations, are considered indirect stakeholders. These actors are the company commander[1], the embedded commanders, the staff and the possible human enablers for the STUV-team. The indirect stakeholders may have specific preferences or restrictions for the schedule, for example to ensure an area is not entered before a certain time or to execute a task before another task, these should also be taken into account during the planning process. Other stakeholders such as the pop-

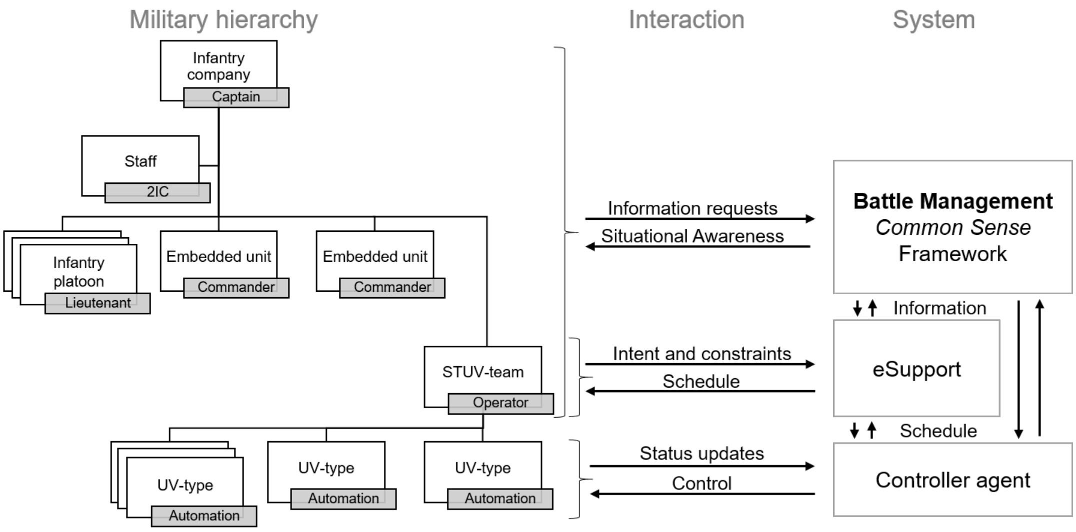


Figure 2.1: Interaction between stakeholders and system components

[1]Note that these tasks can be delegated to a subordinate

ulation, the enemy and air control are probably also affected by the output (the schedule) of eSupport because of the physical presence and behavior of UVs resulting from the execution of the schedule. These interactions with UVs however are not considered in the current study, the analysis focuses only on the creation of the schedule and the resulting schedule, not on the actual execution of the schedule by UVs. However, once eSupport is further developed with more focus on the execution and physical behavior of UVs, for example when a definitive scheduling and path finding algorithm is developed or when eSupport is coupled to the automation controlling the behavior of UVs, it is necessary to take the values of these stakeholders into account. The values of the selected direct and indirect stakeholders are analyzed in the next section.

2.2.3. Values

A study of Friedman et al. (2013) concludes that stakeholder values should not be fixed based on the opinion of the consulted stakeholders, but rather should be based on the most important values of different stakeholders. This allows system designers to evaluate alternatives for the design. In this light, a list of stakeholder values is created based on the most common of all values identified in various studies about the military domain, but also specifically for military and USAR UV operators. The values of the more experienced USAR UV operators are also considered, since the use of UVs in the USAR domain is similar to military operations in urban terrain (Murphy and Burke, 2010). The values in these studies had been identified with stakeholder interviews and with value assessment workshops (Harbers et al., 2015). The current study uses formerly identified values for military commanders (Streefkerk et al., 2014; Jenkins, 2012), unmanned robot operators for military (Parasuraman and Riley, 1997; Parasuraman et al., 2003; Chen et al., 2011; Chen and Barnes, 2012a,b; Barnes et al., 2014; Clare et al., 2012b; Cummings and Bruni, 2010) and USAR (Harbers et al., 2015) applications and general human values for automated systems (Friedman et al., 2013).

The most common identified values in these studies are shown in **Table 2.1** for the direct stakeholders and **Table 2.2** for the indirect stakeholders. The definitions of the values are listed in **Table 2.3**. Besides the identified most common values in prior studies, the value *deconfliction* is added because it was indicated as most crucial for the use of UVs in the first interview (Section A.1).

The user values are categorized as either System Interaction values or Purpose values. The System Interaction values are the values considering the interaction of the operator with eSupport, the purpose values are the values about the schedule itself. Based on the interviews with SMEs, described in Appendix A, the value *trustworthiness and validation of information* should be added to the purpose values, however this value was not included in the analysis. The value overview does not exceed the scheduling phase, so there are no values about the actual deployment of UVs included nor about the interaction of UVs with the environment. Note that the indirect stakeholders have no direct interaction with eSupport and therefore only have Purpose values.

2.2.4. Conclusions

The stakeholder analysis resulted in one assumption and one constraint for the design of eSupport.

- **DA0003**: The relevant information about the operation is communicated via a Battlefield Management System (BMS). Authorized stakeholders can insert, prioritize and manage intelligence requirements and can see the status through BMS at all times in a way that the user can easily process it. eSupport has access to all the information available in BMS.

- **DC0002**: The analysis and design of eSupport do not include the development and the interaction of the user with the Battlefield Management System, nor the interaction of eSupport with the agent(s) in control of the autonomous UVs

2.3. Activities

This section describes the military activities as such defined in the Dutch Doctrine for Land Operations (Dutch Royal Army Doctrine Commitee, 2014). All military operations can be defined in terms of these activities at all desired abstraction levels.

Table 2.1: The values of the direct stakeholders

Value level	Operator
System interaction:	• Effectiveness • Efficiency • Situational Awareness • Balanced cognitive task load • Decision authority • Calibrated trust
Purpose (schedule):	• Effective use of resources • Efficient use of resources • Deconfliction • Minimization of civilian casualties • Compliance with imposed constraints (RoE/Legal/Boundaries) • Safety • (Trustworthiness and validation of information)

Table 2.2: The values of the indirect stakeholders

Value level	Company commander, embedded commanders and staff	UV Enablers
Purpose (schedule):	• Effective use of resources • Efficient use of resources • Coordination • Deconfliction • Minimization of civilian casualties • Hearts and minds / media • Compliance with imposed constraints (RoE/Legal/Boundaries) • Safety • Conservation of combat effectiveness • (Trustworthiness and validation of information)	• Effectiveness of UV • Efficient use of UV • Coordination • Deconfliction

The analysis of activities is divided in three steps. This section start with a description of the different levels of military activities. Subsequently, the military activities a CAT can execute are described in more detail. The STUV-team should be able to contribute to these activities. At last, the actual activities of the STUV-team operator during the briefing, analysis and command phase are described in use cases in the specification chapter, see Section H.1.

2.3.1. Levels of military activities

Military operations can be categorized in four levels according to the Dutch Doctrine for Land Operations (Dutch Royal Army Doctrine Commitee, 2014). These levels are the strategic level (divided in political and military), the operational level, the tactical level and the technical level. The levels are often not

Table 2.3: Value descriptions

Value	Description
• Effectiveness (interaction)	The system is producing the intended or expected result
• Efficiency (interaction)	The system performs in the best possible manner with the least waste of time and effort
• Situational Awareness	The perception of environmental elements with respect to time or space
• Balanced cognitive task load	A balanced amount of mental processes that does not under- or overload the user
• Decision authority	The final right to decide for the user
• Calibrated trust	Appropriate confidence about the system's ability to perform the task satisfactory.
• Freedom from bias	Absence of preconceived or unreasoned opinions of the user about the systems fitness for its task
• Effectiveness (schedule)	A task allocation to assets that optimizes utility in terms of user's goals and intent
• Efficiency (schedule)	A task allocation to assets in the best possible manner in terms of least waste of time and effort
• Safety	The freedom from the occurrence of risk, injury, danger or loss for friendly troops and possibly the environment
• Deconfliction	The prevention of collision of various units and assets in the same area
• Coordination	The act of coordinating the activities within the same area or the same phase of the operation
• Compliance with imposed constraints (RoE/Legal/Boundaries)	Not offending given constraints such as the Rules of Engagement, legal policies and rules and given boundaries for the operation
• Conservation of combat effectiveness	Not hindering the effectiveness of the military operation
• Trustworthiness and validation of information	The extent to which data/information is reliable and the possibility to validate the reliability of the data/information

strictly separable and influence each other even more in modern operations. A short description of each level is given below.

Political strategic level The political strategic level comprises the coordination, systematic development and deployment of all instruments of power of a state, alliance or coalition to serve interests. The political desired end state is defined in the grand strategy. The political strategic level determines the required instruments to achieve the goals and the additional guidelines for the usage of these instruments.

Military strategic level Military strategy is the coordinated, systematic development and deployment of military instruments of power of a state, alliance or coalition to achieve the military components of the goal described in the grand strategy. The military strategic level determines a general mission statement in their strategic directive based on the military strategic goals. Subsequently this level assigns military resources and possibly determines guidelines for the deployment of these resources.

Operational level The operational level comprises the planning, command and execution of joint and/or combined campaigns to achieve the military goals determined in the strategic directive. The operational level is the link between the military strategic goals and the tactical deployment of units.

Tactical level Tactics is the way of disposing and maneuvering formations to execute military activities in ordered arrangement to accomplish a goal. The tactical level can directly deploy units for execution of military activities to realize the operational goals of a campaign or sometimes to realize military strategic goals.

Technical level The technical level concerns the way of disposing and maneuvering small units, possibly individual soldiers or systems, in ordered arrangement to achieve the (tactical) goal of a military activity. This level concerns the actual execution of (combat) techniques and tactics (skills and drills) and the low level tasks in support of these actions.

The activities that can physically be executed belong to either the tactical or technical level. The next section describes these tactical and technical activities in more detail.

2.3.2. Military activities

The military doctrine LDP-1 (Dutch Royal Army Doctrine Commitee, 2013) identifies four main military activities at tactical level, these activities are: offensive, defensive, stability and enabling activities. A complete list of activities within these four categories is provided in Appendix B. A fifth but different category is *information operations* that concern the coordination of activities in the information domain. These operations aim to influence knowledge, attitude and behavior. Information operations are planned and executed in close cooperation and coherence with the tactical activities, but are not a separate form of operation.

Offensive activities aim to eliminate the factors that keep a crisis going, in most cases by eliminating the opponent. The doctrine identifies eight offensive activities, such as attack and reconnaissance in force. The defensive activities are defined as the activities to resist enemy offensive activities. The purpose of these activities is to remain own power and involve the combination of essential information, personnel, materiel, infrastructure and terrain (Noël, 2009). There are two types of defensive activities, which are defense and delay.

Stability activities are executed in close collaboration with other actors in the environment to maintain, recover or create a situation to ensure proper functioning of the responsible governments. The doctrine identifies four types of stability activities, e.g. security and control. The enabling activities support and link the transition between offensive, defensive and/or stability activities and are therefore never independently executed. The doctrine identifies thirteen enabling activities, such as reconnaissance, security, obstacle breaching/crossing and march.

The military activities are the complete set of military possibilities without going into details about the execution by units. Tactical activities can typically only succeed by completing several technical activities. These technical activities describe the way of execution of a task by a specific unit. Some of them are also known as ”skills & drills”, since these activities are trained as such to individual soldiers and units. Some examples of technical activities are the skills and drills to move with a vehicle under a specific enemy threat level, to react during enemy contact and to reconnoiter an obstacle under a specific enemy threat level. It is to soon to specify the capabilities of UVs and decide their capabilities in terms of technical and tactical activities. This requires more knowledge about these systems.

The design of the system requires knowledge about the capabilities of UVs. The lack of knowledge therefore enforces to make assumptions about these future capabilities. The next section describes three activities with the assumption that UVs can execute these high-level activities autonomously (DA0004). The capabilities of the UVs do not need to be described in detail for the design of eSupport.

2.3.3. Activities in problem scenarios

The current study defines three types of high-level tactical activities performed by UVs. These are based on the activities already defined by TNO in their CommonSense Framework (see Section 2.5). These activities are: person reconnaissance, vehicle reconnaissance and building reconnaissance, they all belong to the tactical activity *reconnaissance*.

Person reconnaissance

The activity *person reconnaissance* is a specification of the tactical activity *reconnaissance*. The activity can be executed by autonomous UVs through a combination of several technical activities, like *move* and *identify*. The activity comprises the search for a specific person, the presence of humans or humans that satisfy a certain condition (e.g. carrying a weapon) in a certain area.

Vehicle reconnaissance

The activity *vehicle reconnaissance* is a specification of the tactical activity *reconnaissance*. The activity can be executed by autonomous UVs through a combination of several technical activities, like *move* and *identify*. The activity comprises the search for a specific vehicle, the presence of vehicles or vehicles that satisfy a certain condition (e.g. a gun mounted on the vehicle) in a certain area.

Building reconnaissance

The activity *building reconnaissance* is a specification of the tactical activity *reconnaissance* and specifies the location of the activity to be a building. The reconnaissance will be carried out in and in the direct environment of the building. The activity comprises various possibilities, e.g. a person reconnaissance, making a 3D model of the building and the search for specific objects.

2.3.4. Conclusions

A CAT can execute a broad range of tactical activities. Tactical activities require one or more lower level technical activities to be accomplished. The design of eSupport requires an assumption about the tactical activities the STUV-team can execute.

- **DA0004**: The UVs are able to autonomously execute the high-level activities person reconnaissance, vehicle reconnaissance and building reconnaissance

2.4. Context

The context of an operation has a strong influence on the course of the operation. Therefore it is important for a system to have sufficient and detailed information about the operational context. In an operations order a commander gives information to its subordinates about the operation and its context. The NATO defines a standard form for operation orders in STANAG 2014, structuring orders in 5 standard paragraphs: *Situation, Mission, Execution, Service Support* and *Command and Signal*.

The *Situation* section comprises the situation of the enemy forces, the situation of the friendly forces and civil and/or terrain considerations. The *Mission* section comprises the assignment and the commander's intent. The commander gives more details about the execution of the mission in the *Execution* section in its concept of operations. In this section also more detail about the operation is provided for each subordinate unit, e.g. the dedicated tasks. In the resulting two sections more specific information about service support (logistics), the command hierarchy and signals are provided. The next section describes those sections of the operation order with relevant contextual information for eSupport in more detail. These are the sections *Situation, Mission* and *Execution*. The identification of relevant information leads to requirements for eSupport, these are described in the conclusion.

2.4.1. Situation of enemy forces

During the briefing, all available relevant information about the enemy is provided including the enemy's composition, disposition and strength together with an estimation of the enemy's capabilities, limitations. The briefing is sometimes accompanied with an estimation about the enemy's most likely course of action and the most dangerous course of action. A subordinate commander can use this information to analyze and decide about their own course of action. The current study assumes that BMS provides an overview of the known threats and the characteristics (DA0003).

To indicate the probability of an event or threat, the Dutch Intelligence Doctrine (Brouwer and Scholten, 2012) defines five confidence levels to quantify the probability. This level is often defined by an analyst. The levels are shown in **Table 2.4** together with a more explanatory synonym and the quantified probability in percentages.

Table 2.4: Confidence intervals from the Dutch Intelligence Doctrine (Brouwer and Scholten, 2012)

Description of Probability or confidence	Synonym	Percentage
Highly Likely	Highly Probable	> 90%
Likely	Probable	60 - 90%
Even Chance	Chances are slightly greater (or less) than even	40 - 60%
Unlikely	Probably not	10 - 40%
Highly Unlikely	Highly Improbable	< 10%

The information about the type of threat and the probability can be used to categorize the risk to an asset. Although this depends on more variables such as blast radius or the effective range of a projectile, it helps to categorize the risks for the algorithm shown in Appendix F.

2.4.2. Situation of friendly forces

In the friendly forces section, the commander gives more information about the situation of friendly forces which are localized in or nearby the unit's area of operation. This information is very important for deconfliction and coordination between friendly units during the operation.

It is also important for the design of eSupport to know the location of friendly forces, especially when a UV observes persons, vehicles or other systems and has to identify them. In current operations, BMS helps soldiers to localize and identify friendly troops. The design of eSupport will not cover the deconfliction and coordination with friendly troops, but will focus more on the interaction with the operator (DC0003). Also, other coordination measures are not included, e.g. phase-lines (locations that cannot be crossed before or after a certain date-time group) and restricted areas.

2.4.3. Civil and terrain considerations

Important information about the terrain and civil aspects is given during the briefing in the section *civil and terrain considerations*. Terrain aspects include for example fields of fire, (non-military) obstacles, key terrain locations and avenues of approach. Also the weather is covered in this section. The terrain and weather together have a strong influence on military operations, or often said: they dictate military operations.

Civil aspects are the aspects in the environment that do not belong to the enemy or the terrain itself. It is important to take the presence of civilians and their infrastructure into account. For example, a road may be blocked due to a market or there is a gas station with gas tanks along the main road.

Consequences of the terrain and civil aspects are mostly bound to a specific location. A BMS is very suitable to make this information available for all units. Analysts or the intelligence and operations sections of the units often make this information available to all units through BMS. The civil and terrain considerations are not taken into account in the analysis for eSupport. Therefore, no restrictive information about the environment and terrain is considered (DC0004).

2.4.4. Commander's intent

The commander's intent is perhaps the most important part of the briefing. ”While detailed battle plans do not remain viable much past the onset of hostilities, the commander's intent remains viable” (Shattuck and Woods, 2000). Therefore the intent is essential to achieve a successful outcome of an operation. According to the U.S. military doctrine (FM 100-5, 1993), commanders gives their intent to ”focus subordinates on what has to be accomplished in order to achieve success, even when the plan and concept of the operation no longer apply” (Shattuck and Woods, 2000).

The commander's intent consists of three components, as such described in the Dutch military doctrine for land operations (LDP-1) (Dutch Royal Army Doctrine Commitee, 2013). The intent should provide clearness about the effects to be achieved (the endstate), the way to achieve that (the method) and the purpose of the operation (the purpose). Within these definitions, the purpose is described on a more general (abstraction) level than the effects in the endstate. The definition according to the US Field Manual FM-5 contains three similar components called End State, Key Tasks and Expanded Purpose. In this report the terms of the Dutch military doctrine are used. See Appendix C for an example of a commander's intent with its components.

The commander's intent is supposed to provide clarity about the operation, in distributed systems however, ambiguity of language forms an obstacle for proper interpretation of the intent. When actors are separated in time and space, the language is an even bigger barrier since the meaning of a sentence is dependent on the context (Shattuck and Woods, 2000). Since the intent is communicated prior to an event but has to be interpreted when the subordinates are confronted with a situation, the intent might be misinterpreted. Therefore commanders should make every attempt to make their intent clear at the right level of detail. Too much detail would give insufficient freedom for subordinates while too few details would not give sufficient guidelines to be able to operate autonomously within the commander's intent. So it is important to consider the commander's intent at the appropriate level for eSupport. This is the level of the unit to which the STUV-team is embedded: the CAT (RQ0001).

An additional note is required here. Human to human communication includes emotional cues besides the explicitly stated message (intent). ”A formalization of the commander's intent in an electronic system does therefore not replace personal conveyance of the intent by the commander” (Hieb and Schade, 2007). This is not the aim of the current study.

2.4.5. Conclusions

The context of military operations include the enemy and friendly forces, weather and terrain and the commander's intent. The analysis led to several constraints for the design of eSupport.

- **DC0003**: The analysis and design of eSupport do not include the presence of friendly forces, deconfliction with UVs nor the required coordinating measures

- **DC0004**: The analysis and design of eSupport do not include terrain and weather and the consequences for the operation and UVs

2.5. Technology

The analysis about technology describes relevant technology for the design of eSupport. Section 2.5.1 describes the use of BMS, and Section 2.5.2 describes the current developments of UVs. Various planning algorithms are discussed in Section 2.5.4 and the analysis concludes with Section 2.5.3 about the communication of objectives to automation.

2.5.1. Battlefield Management System

Due to improved mobility the area of operations of military units got larger, often making it impossible for commanders to keep direct visual contact with subordinate units. A Battlefield Management System (BMS) supports the commanders to share information among military commanders and provides them with access to all current information. This increases the commanders' awareness of the situation, the so called ”situational awareness”. A BMS is an essential element of the digitized battlefield for military operations at the low tactical level, ranging from a single vehicle to a battalion and increases the operational effectivity (Gmelich Meijling, 1997).

The Dutch Army provided their vehicles with a BMS, a project that already started in 1997 with the first pilot study (Gmelich Meijling, 1997). The BMS consists of a working station and a screen, is coupled with a global positioning system and it communicates with other BMS entities by radio (Gmelich Meijling, 1997). BMS uses a Command & Control (C2) application called OSIRIS that enables the user to interact with the system. OSIRIS shows a map on the screen with all relevant information, such as the actual position of friendly units and indicated enemy units and/or obstacles. A system called ELIAS is going to replace OSIRIS starting from 2015 (Ridder de, Cap, 2015). Both applications are however not open source and thus not available for the current study.

TNO developed the Common Sense Framework to improve Situational Awareness among military commanders and to enable effective communication with other commanders and automation. It is specifically designed for military commanders and has already been evaluated in different test settings with military personnel and thus has been developed and improved according to their requirements. The framework offers the functionalities of a BMS, but also offers the additional functionality to autonomously schedule tasks to multiple heterogeneous vehicles, both manned and unmanned, based on the sensors of the vehicles. The system enables the user to insert assets, enemies and tasks with further information and characteristics on a map. Finally, the user can request an allocation of tasks to the available assets. However, the user cannot exercise further control.

The framework has a screen that is divided in three parts, which are the main screen and a menu bar at the left side respectively at the bottom of the screen, see Figure 2.2. The main screen represents a map on which all location bound information is shown together with some tools. The user can zoom and scroll through the map and interact by manipulating tasks, enemies and assets on the map with specific information about the item and its location. The left menu bar offers several capabilities such as changing the map style and enabling and disabling layers for assets, tasks and enemies. The bottom bar enables the interaction with the different components on the map. There is a tab to add assets, a tab to indicate tasks and a tab to indicate enemies. Another tab enables the user to request a schedule. The framework can also be connected with other systems to exchange information and communicate with other commanders. Other capabilities of the Common Sense Framework are not relevant for the design of eSupport and therefore will not be discussed.

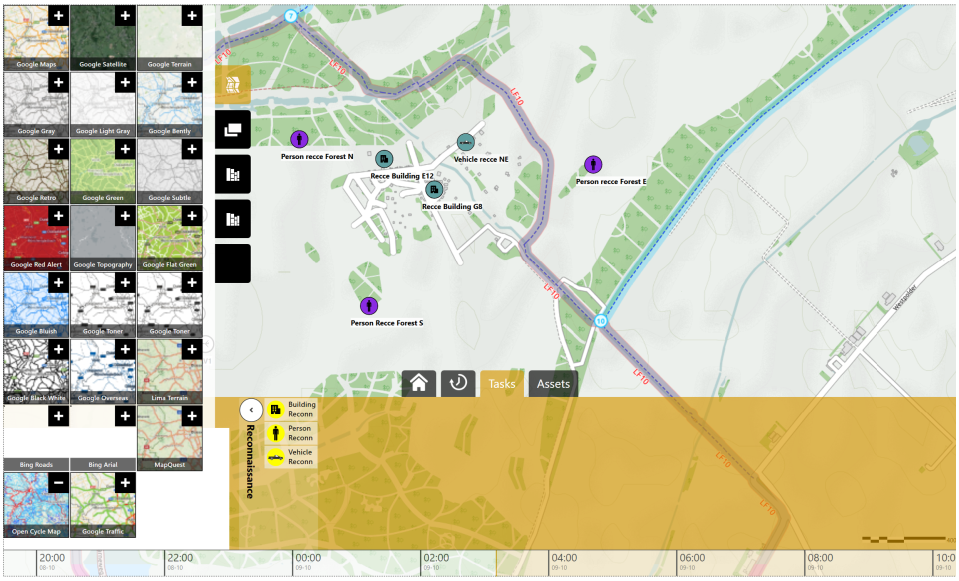


Figure 2.2: Overview of BMS: The CommonSense Framework of TNO

The framework offers useful functionalities for eSupport such as the BMS and a basic scheduling functionality. The current study will therefore make use of the framework to serve as the BMS. The BMS functionalities to show and manipulate the map, assets, tasks and enemy were already included in the framework and evaluated. These will not be evaluated separately (DA0007).

2.5.2. Unmanned Vehicles

The current technological developments resulted in a rapid development of UVs and their capabilities. The Dutch army currently has (operational) experience with various UAVs, e.g. the Raven, the ScanEagle and the Sperwer and UGVs, e.g. Telemax, tEODor and Dragon Runner. These vehicles are all used for explosive ordnance disposal. However, these vehicles do not meet assumption DA0004, since these systems do not have the capability to execute high-level tasks without the control of multiple UV-operators. Therefore, vehicles able to execute high-level tasks autonomously are required.

Several documents describe the intention of the Netherlands, Europe and the US for the development and/or use of UVs in the future (Stam, 2009; Office Secretary of Defense, 2005). However, there are no UVs that can execute high-level tasks autonomously yet. While the current study does not focus on the design of UVs, the experiment requires a set of heterogeneous vehicles to give operators a choice between vehicles based on their characteristics during the experiments. Therefore, a set of five fictive UVs with realistic capabilities were designed, based on the current expectations of these vehicles in the future (Stam, 2009; Office Secretary of Defense, 2005). Appendix E provides an overview of five fully autonomous aerial and ground vehicles with their capabilities, based on existing UVs.

2.5.3. Communicating Commander's intent

A supportive system aims to support the user to achieve certain objectives. Therefore, the objectives of the user have to be clear to the system. In the military domain, the objectives of a commander are encapsulated in the commander's intent, as described in Section 2.4.4. The intent is therefore an essential component to communicate to a system. However, it is very difficult to communicate the intent to a Command and Control (C2) system, since the intent is written in free format text, so incomprehensible for systems. Therefore it is important that users are able to express their intent to the system, but also in a way so the objectives are understandable for the system and all other stakeholders (Miller et al., 2005). When the communication of the commander's intent succeeds, it can reduce the amount of cognitive steps and time required to gain an acceptable schedule (Clare et al., 2012c).

Two approached to communicate the commander's intent to a system are described in the two following sections.

Formal Command and Control language

Human languages contain a lot of semantic ambiguity. To be comprehensible for command and control (C2) systems and future robotic forces, a language needs to be formalized and unambiguous (Hazen and Randall, 2008). To overcome the problem of ambiguity, an initiative was launched in 2001 to create a formal unambiguous grammar for the military domain. The grammar was intended to command and communicate with live forces, simulations and robotics. The language is called the Battle Management Language (BML), which Carey et al. (2001) defined as:

> The unambiguous language used to command and control forces and equipment conducting military operations and to provide for situational awareness and a shared, common operational picture.

The article defined key concepts and principles for BML. Subsequently, Hieb and Schade (2007) presented an extension for BML, which until then was only suitable for tasking and orders, with a grammar for the communication of the commander's intent. Their proposed method to communicate the commander's intent via BML is described in Appendix D.

The BML grammar seems suitable to express certain components of the commander's intent to eSupport, however according to Hieb and Schade (2007) more research is required prior to implementation.

A comparison of an example of a written commander's intent with the same intent described in the

BML lexicon in (Hieb and Schade, 2007, p.14-15), shows that the formal grammar does not cover all relevant information. See Appendix C for the commander's intent and the components. Certain terms used in the original commander's intent described values, such as *rapidly, courageous, hasty* and *aggressive*. These values cannot be expressed in the current formal grammar, however they are very important for subordinates to act upon the commander's intent, so also an automated system should have as many as possible knowledge of the commander's intent to be able to act upon it. This requires that users are able to express all their objectives, both functional objectives and objectives about the execution of the tasks described with values (RQ0001).

The next section describes a method to provide more information about the user's objectives to the automation.

Communication of intent via simulations

Simulations are a good way to share the intent of the commander with subordinates via technology (Hazen and Randall, 2008). Hazen and Randall (2008) based their conclusion on previous work about the creation of scenarios for training and analysis purposes. These scenarios require information about the expected behavior for every entity in a simulation. According to the authors, the required information is fairly similar to the information provided in a commander's intent.

To create a realistic simulation, the authors encoded the commander's intent with several parameters in a simulation algorithm. After proper validation of the parameters, the simulated entities or agents show behavior that corresponds to the commander's intent. The challenges of this method are first of all to get the right settings for the parameters such that the resulting behavior corresponds to the intent. The second challenge for the operator is to trust that the system will behave accordingly also in unforeseen situations.

Based on the experiences of Hazen and Randall (2008) it seems promising to communicate the commander's intent to eSupport with parameters. The number of parameters can get very large though. Besides, the commander's preferences for objectives depend on the specific situation. Therefore, decisions about these parameters cannot be made during the design phase, the operator should be able to set these parameters in real time. As a result, there is no possibility to test and refine the parameters, so the effect of the parameters has to be straightforward. For this reason, only the parameters that can be reflected in the algorithm and the resulting schedule are taken into account for the current design.

Clare et al. (2012b) used a similar approach to allow the operator to change settings for the planning algorithm. The authors chose to use binary parameters, so the users can set a parameter to on (important) or off (not important). They based their choice on a study that explained that users often make insufficient adjustments to the parameter settings, based on the initial value of the parameter, the so called 'anchoring and adjustment' heuristic. The current study however follows the military command & control process and requires the operator to express its intent through several parameters prior to creating a schedule, similar to informing subordinates about the intent. The operator is thus forced to think about its intent including its implicit objectives, to prevent that the implicit objectives start to play a role later in the scheduling process.

Besides, the binary choice used by Clare et al. (2012b) does not allow operators to express their actual intention, since the operator is forced to consider something important or unimportant. In the current study, all parameters are considered important and the operator has to make decisions about the setting of the parameter, e.g. risk should be avoided vs. losses are acceptable. We assume that a five-level Likert scale provides the operator with enough freedom to express its intent for a parameter, while not asking for too much detail. eSupport therefore allows the operator to choose from a five-level Likert scale for the parameters of its intent whenever these are not nominal (e.g. binary) parameters by nature or when a five-level scale would not make sense (DA0005).

The next section describes the scheduling algorithm and the parameters that will be communicated to eSupport.

2.5.4. Planning algorithms

Ghallab et al. (2004) made a distinction between the planning and the scheduling phase. In a planning phase one determines what has to be done from a set of goals described as a set of actions. Subsequently, in the scheduling phase one determines when and how the actions have to be executed, described in a scheduled plan. Planning and scheduling sometimes affect each other, in that case they need to be integrated. This study focuses on the scheduling phase, the high-level tasks are already defined by the military commanders.

For the scheduling task, eSupport requires an algorithm to allocate the tasks to UVs in an (near) optimal way. Clare et al. (2012a) concluded that near-optimal algorithms in some situations have great benefits for performance. Because scheduling tasks typically can become very complex when taking multiple optimization-parameters into account, an optimal solution might require too much time to compute.

The CommonSense Framework included a scheduling algorithm. This algorithm takes all tasks, the characteristics of UVs and the distances into account using random optimization. However, because the algorithm used random optimization, it returned completely different schedules for the same problem. The fitness of these schedules were also significantly different. eSupport requires a stable algorithm to compare the effects of the transparency conditions during the experiment, so this algorithm was not usable.

We had to implement a new algorithm that provides stable results. First, we considered a brute-force algorithm, this is an algorithm that checks all possibilities. However, because the number of possibilities grows exponentially with the number of tasks or resources, the algorithm requires much computational time. A schedule has to be calculated several times during a mission, so the use of this algorithm would have a negative effect on usability.

There are several other scheduling algorithms that produce relatively stable and (near-)optimal results. Most of these algorithms require much analysis and adaptations for the specific problem before they can be implemented. Because the design and implementation of an algorithm is not in the scope of the study, a more simple algorithm using a greedy approach was implemented. Greedy algorithms have the advantage to produce stable results, but they do not consider global optimization, rather these algorithms try to achieve a high utility as soon as possible. We could argue that this is wishful in a highly unpredictable environment.

The greedy algorithm determines for each individual task what UV is best to execute it using a utility function. Then, the algorithm allocates the task with the highest possible utility to the best UV. This step is repeated for the resulting tasks, taking the already assigned tasks into account, until all tasks are allocated. See Appendix F for the pseudocode. This shows that the algorithm only calculates the local utilities (so per UV-task combination) and does not consider the global optimum. However, the results are often optimal given the relatively small amount of vehicles and tasks in the experiment.

Utility function

The algorithm uses a utility function to determine the fitness of a solution. To determine this fitness, the function requires a point of reference. This section analyzes how to calculate the fitness of a solution.

A utility function is bounded by the available information of the environment. The CommonSense Framework contains information about the resources (UVs), the tasks, locations and time. The data available in the CommonSense Framework is listed below.

- Resources
  - Location
  - Maximum and cruise velocity
  - Available sensors
  - Terrain mobility / flight altitude
  - Minimal distance to target for covert operations
  - Action radius (based on fuel consumption)
- Tasks

- Location
- Required time
- Priority
- Required level of detail of the information (resolution)

- Location
  - Enemy presence
  - Terrain & weather conditions
- Time

The user and eSupport are also able to manipulate the environment. The variables that can be manipulated in the CommonSense framework are:

- Resources
  - Velocity
  - Sensors used
  - Altitude / path
  - Minimum / maximum distance to target
- Allocation of tasks to resources
  - Allocation of tasks to resources
  - Order of task execution
  - Time constraints

These variables should be manipulated in order to achieve maximal satisfaction of the user's objectives (and intent). Therefore, eSupport should find the best combination of settings for these variables. Further analysis of these variables resulted in a set of six variables that should be manipulated, each value is described below together with the user's considerations to manipulate these variables.

Information quality and quantity
The ability of a UV to execute a task is expressed in the amount of useful information (information quality and coverage) a UV can gather for a task. This is based on two things: the resolution and the coverage. The resolution of the sensors of the UV determines if a vehicle can achieve the required quality, e.g. a person identification requires high resolution to recognize a person. The coverage is determined by the characteristics of the vehicle and describes how much area the vehicle can reconnoiter, e.g. a large vehicle cannot enter a building for interior reconnaissance. Since the required information is covered in the intelligence requirement, eSupport cannot change the settings. Therefore, the variable will not be taken into account for the intent.

Fuel consumption
During military operations, fuel might be scarce or UVs need to save fuel to be able to execute tasks in a later stadium. A UV can use fuel more efficiently by moving at cruise altitude or at roads at cruise velocity. In the current study, fuel consumption is not considered, so a user doesn't have to worry about fuel, fuel reservoirs and refueling during the experiment (DC0005). Therefore this parameter will not be included in the intent.

Time
Time describes the time required for a UV to move to a location and execute the reconnaissance. The users should consider the time pressure of a mission to indicate their preferences. Time pressure indicates the necessity to execute the tasks as fast as possible. Although tasks can have deadlines, this does not equal time pressure. For example, while multiple tasks may have no restrictive deadlines, it might still be wishful to gain a lot of information on short notice. In that case one might want to be able to set a high time pressure. This also should prevent an operator to set artificial deadlines for tasks.

Secrecy
Operations can be classified into three types of secrecy: overt, covert and clandestine (US Department of Defense, 2010). "Operations (...) that are executed in such a way as to assure secrecy or concealment" are called clandestine operations. "An operation that is so planned and executed as to

conceal the identity of or permit plausible denial by the sponsor” is called a covert operation. ”Openly conducted operations without concealment” are called overt operations. The scope of the current study is however on tactical units, these units will not be involved in clandestine operations. Therefore, the users can choose between overt or covert execution of tasks.

Risk
The environment exposes various risks to the resources, such as heavy weather and enemy fire. The risk is expressed in the chance to lose a resource. Users can define their considerations for risk by indicating their acceptance for risk exposure. Risk exposure defines the acceptance of the chance to lose a resource.

Other variables
Users might have additional preferences for the schedule to achieve their implicit objectives or the objectives of the CAT such as *deter the enemy* or *win hearts and minds*. Mostly, these preferences will affect the physical behavior of the UVs and will not have an impact on the schedule. One such variable that does impact the schedule is the preference to use many or few resources, this is the *massiveness*. This parameter pushes the use of more or fewer resources. Massiveness is defined as *the pursuance to use many resources for the operation*.

The utility function uses five variables to determine the fitness of a solution. These variables with the respective user preferences are listed in **Table 2.5**.

Table 2.5: Algorithm variables and user preferences

Variable	User preference
Information quality and quantity	Always maximized
Time	Time pressure
Overt or Covert	Secrecy
Risk	Risk exposure
Amount of resources used	Massiveness

Almost all variables conflict each other, so the best settings for the values are not straightforward. For example, the distance to an object does improve the level of detail of the information, however this results in a higher risk. Since there is no optimal solution in general for these conflicts, these are dilemmas for eSupport. There is no general solution for these dilemmas and the preferences for these dilemmas depend on the situation. Therefore, the users should indicate their preference for each mission.

Preferences for algorithm dilemmas
The preferences of the users for the dilemmas are reflected in their (purpose) values, defined in Section 2.1. Below you see which preferences are reflected by each user value:

- Effective use of resources
    1. Information quality and quantity
- Efficient use of resources
    2. Time pressure
    3. Fuel consumption
- Compliance with imposed constraints
    4. Secrecy
- Safety
    5. Risk acceptance

- Conservation of combat effectiveness
    6. Massiveness

The user value *deconfliction* is not represented by a variable, since friendly troops are not part of the design analysis (see DC0003). The resulting parameters defined for the utility function with their respective quantities are:

- Time Pressure: Very low pressure (1) - Very high pressure (5);
- Risk exposure: Take no risk (1) - Take risk (5);
- Massiveness: Use few resources (1) - Use many resources (5);
- Secrecy: Overt - Covert;

We assume that these parameters are sufficient for eSupport to produce a schedule that satisfies the user during the experiment (DA0006). The utility function uses the parameters to calculate the fitness of a solution and is described in Appendix F.

Using qualitative terms
Section 2.5.3 suggested to incorporate qualitative terms in eSupport. When these terms, like aggressive or surprising, can be communicated to eSupport, eSupport should be able to use these terms. eSupport could translate these terms to system' understandable parameters. Especially when the physical behavior of UVs will be incorporated in eSupport, it might become interesting to create a fixed set of parameters for these qualitative terms. For example, using the current available variables, the term aggressive might be translated to an acceptance of much risk, high time pressure, high massiveness and Overt. Since the BML is not available for the current study, this is not covered in the current study.

2.5.5. Conclusions

A BMS provides an overview of information about the battlefield. The CommonSense Framework of TNO will be used as BMS as a basis for eSupport. BMS shows the enemy, high-level tasks and resources. The resources for the experiment will be five autonomous UVs defined for the current study, these are fictive vehicles and able to execute three high-level tasks autonomously.

The Battlefield Management Language is designed to support the command and control of forces and supports communication of the commander's intent to the system. The language seems suitable for eSupport, however the BML lexicon requires more development before it can be used. The commander's intent will be communicated to eSupport with parameters for 1. Time pressure, 2. Risk exposure, 3. Massiveness and 4. Secrecy. The user decides for each of the parameters what value fits their intent best. Qualitative terms of the commander's intent can be translated to system variables once the BML is incorporated in eSupport.

A greedy algorithm is used for eSupport, since these produce stable results and don't require much time for computation. Considering the relatively low complexity of the missions in the experiment, the algorithm produces near-optimal to optimal solutions.

The analysis led to a requirement, assumptions and a constraint for the design of eSupport.

- **RQ0001**: eSupport allows users to share all their important objectives and constraints with eSupport at the appropriate level, so for the STUV-team, and eSupport shall compute a (near) optimal schedule that satisfies the objectives and constraints of the user

- **DA0005**: A five-level Likert scale provides users enough freedom to indicate their preferences for each intent parameter
- **DA0006**: The parameters *time pressure*, *risk acceptance*, *massiveness* and *secrecy* are sufficient to align the user's objectives with eSupport in order to allow eSupport to suggest a schedule that satisfies the user during the experiment

- **DA0007**: The analysis and design of eSupport does not include a validation of the Common-Sense Framework as a BMS, the current study assumes that the framework functions well as BMS, covers assumption DA0003 and is a proper basis for eSupport

- **DC0005**: The analysis and design of eSupport do not consider the fuel consumption of UVs and the related consequences

2.6. Human Factors

Supervision of multiple UVs is a highly demanding task, this can easily lead to several problems. Since users have to supervise multiple systems, but several vehicles or situations might require more attention, this could lead to a decreased ability to supervise the other systems adequately. Parasuraman and Riley (1997); Chen and Barnes (2012a,b); Chen et al. (2011) have investigated the effects of supportive autonomy for supervisory tasks. Several negative consequences have been identified, such as degradation of user's situational awareness (SA), a higher cognitive task load due to the increased complexity, misuse (over reliance on automation) and disuse (under utilization of automation) due to varying trust in the automation and possible loss of skills to perform certain tasks manually. Studies show that the performance of a user to supervise multiple systems depends on the characteristics of the automation, the tasks and the environment. These characteristics are for example the reliability of the autonomy, the complexity of the tasks, the degree of coherence and accuracy of the available information about the tasks (Mosier et al., 2007, as cited in Chen et al., 2011) and the effectiveness and usability of the interface (Chen et al., 2011). Chen et al. (2011) concluded that interfaces that do not support effective interactions or that have poor usability affect cognitive task load and overall performance (RQ0002). The current study investigates the effects of additional transparency on the performance, it makes more sense to investigate these effects when the performance is not negatively affected by eSupport itself. Therefore, these consequences have to be considered in the design of eSupport. This section describes the literature about the relevant human factors.

2.6.1. Situational Awareness

Adequate SA is one of the most crucial factors to ensure effective supervisory control over multiple vehicles (Chen et al., 2011). SA is defined by Endsley (2011) as "The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". It basically is the awareness of what is happening around you and understanding what that information means to you now and in the future, defined in terms of what is relevant for the goals and objectives of a specific job or function. According to Endsley (2011), SA can be categorized into three levels, this section describes these levels of SA.

The three levels of SA according to Endsley (2011) are:

Level 1 Perception of the elements in the environment;

Level 2 Comprehension of the current situation;

Level 3 Projection of future status.

A description of the levels is provided below.

Level 1: Perception of the elements in the environment

The first level of SA comprises the perception of the status, attributes and dynamics of the relevant elements in the environment. The domain and job determine what the relevant elements are. For example for a UV operator the relevant elements include the information about UVs, tasks, friendly and enemy forces, other present actors, the weather and the terrain. However all required information might not be available, or even false due to misleading activities by the enemy. To support the user, eSupport must ensure that the necessary information is obtained and presented in such a way that the user can easily process it (Endsley et al., 2003). These functional requirements are covered in the assumption for BMS in DA0003.

Level 2: Comprehension of the current situation
The second level of SA covers a person's understanding of the relation between the data of the first level of SA with the relevant goals and objectives. This means that the user is able to link information about loose data and elements in the environment to the goals and objectives and understands the impact of changes. Therefore the user needs a good knowledge base or mental model to be able to interpret and reason about these relations. Especially the relation with goals is important because these form the basis for decision making in complex environments (Endsley et al., 2003). To support the user, a supportive system should show the link between available data with the goals it affects whenever possible (Chen et al., 2011, cited from 45) (RQ0005). The current study does not investigate the comprehension of new information about the environment, therefore no new information about the environment will become available during a scenario (DC0006).

Level 3: Projection of future status
The third level of SA covers the user's ability to predict what the elements defined in level 1 and level 2 will do in the future. A good mental model of the user is essential for this. A high level of SA at level 3 can prevent unanticipated and undesirable situations, however high cognitive task load or insufficient knowledge about the domain could prevent the user to achieve proper SA at this level. Therefore the system should not overload the user with data (RQ0003). An overload of information, resulting in a significant increase of cognitive task load, can be avoided by only presenting information relevant for the current goals and objectives. New data should be integrated with already known information, instead of presenting loose data (Endsley et al., 2003). The system can support the user to increase level 3 SA by showing the future consequences of specific decisions or changes (RQ0004).

The definition of SA of Endsley (2011) focuses on the point of view of a single user. However, SA can also be described from a system or collaborative view, with more focus on the distribution of knowledge throughout a team and the process of acquisition of knowledge (Salmon et al., 2008). Stanton et al. (2006) indicate that not all actors in a system require the same knowledge, each (human or nun-human) actor requires knowledge (SA) depending on their specific task, so not each individual requires the same knowledge. The available knowledge should be distributed among the system and specific knowledge should be activated for or communicated to an actor once it is required for a specific task. Meta-SA is required to know what knowledge other actors have. Therefore they suggest to consider SA at system level instead of individual level. This point of view and the resulting analysis about what knowledge is required for each phase of an operation, could be used to decide what knowledge should be presented (more salient) for a user at a different phase in an operation. This would support efficient communication of information. The current study focuses only on the scheduling task of one individual, this requires specific knowledge for this user. The required knowledge is analyzed in Section 2.6.2. However, when the design of eSupport is extended to support the user with other tasks, or when UVs need to communicate data to the user, eSupport should consider what knowledge is relevant for the user for each individual task. The suggested theory of Stanton et al. (2006) would be a good basis for this analysis.

Concluding, there are a number of factors that influence the user's SA, such as the user's cognitive task load and the level of autonomy of the supportive system. In the current study, the level of autonomy of eSupport is fixed, to be able to focus on the effect of additional transparency, so the appropriate level(s) of autonomy will not be investigated in the analysis (DC0007). The user's SA can be improved by presenting the right information at the right time. To further investigate what information is relevant for the user, the theory about abstraction hierarchy is discussed in the next section.

2.6.2. Abstraction Hierarchy

In order to determine what information is relevant for the user of a system, an abstraction hierarchy is used. An abstraction hierarchy is a categorization of a system or domain into five levels of abstraction (Rasmussen and Lind, 1981). It represents the full domain for which it is made and is independent of available technology. The five levels of abstraction help to identify what a user is focusing on during an activity, and therefore what the relevant information and support is during that activity. The main focus of a user typically is on one level of abstraction. The abstraction level above describes the purpose (why) of the focus level, the abstraction level below describes the methods (how) to realize the focus level (Rasmussen and Lind, 1981). There relations are also called the means-ends relations. For example, the user focuses on a task called ”building reconnaissance”, then the level above describes the purpose of that task, which is to ”Prevent an enemy attack”. The abstraction level below describes the method, so the physical resource that executes the task, which would be ”UAV-1”. The user typically focuses on one level of abstraction and on the level below and above. This limited span of attention of a user is used for the design of eSupport, so eSupport should only provide information from the three appropriate levels of abstraction (RQ0003).

The abstraction hierarchy can be expanded by adding the decomposition space. This provides the hierarchy with an extra dimension to cover the level of detail (resolution), the resulting two dimensional space is called the abstraction decomposition space (ADS) (Baker et al., 2008; Bennett et al., 2008). Going to the right in the decomposition space shows more detail (higher resolution). For example, a military vehicle is represented as one component in the ADS, going down in the decomposition space (so to the right in the ADS) would show the components of the vehicle such as the wheels. Going up in the decomposition space (so to the left in the ADS) would show the unit of the vehicle.

Defining the abstraction decomposition space

An ADS is useful to determine the span of attention of a user and the relevance of information. An ADS can be defined for a certain domain with a predetermined scope, another domain or scope comprise other activities and often results in a different hierarchy. So to determine the span of attention of the user while working with eSupport, the design requires an ADS specifically for the STUV-team as domain.

Different studies often label the abstraction levels differently (Jenkins, 2012). The five levels of abstraction described by Jenkins (2012) are, from high (more abstract) to low: 1. domain purpose, 2. domain values, 3. domain functions, 4. physical functions and 5. physical objects. Jenkins (2012) specifically chose for these labels to distinguish the top 3 levels, which describe the domain independent of technology, from the other 2 levels, which focus on technology and its functions, independent of the domain. This distinction makes sense because (domain) functions in the military domain typically can be executed by different physical entities and/or functions. This study therefore uses the labels as such defined by Jenkins (2012).

Jenkins (2012) presented an abstraction hierarchy for the domain military air operations and included UAVs besides other manned air vehicles as a technological resource. The authors considered the domain in general and did not focus on a certain echelon. As a result, they defined the purpose of this domain as *supporting military operations effectively*. The study gave more insight in the abstraction levels of military air operations in support to military operations, however, the current study focuses on the integration with land operations.

Martinez et al. (2001) presented an abstraction hierarchy focused on land operations of battle groups (battalion echelon) and brigades. This domain is more similar to the current domain considering the top three abstraction levels. However, Martinez et al. (2001) and Jenkins (2012) interpreted the abstraction levels differently. Martinez et al. (2001) described similar domain values but at one abstraction level higher. Subsequently, the authors defined the term *combat power* at the second level of abstraction, which they defined as the *abstract functions and priority measures* level. They defined *combat power* as the complete set of resources including intangible resources (such as experience and leadership). This information however does not conform to the means-ends (how-why) characteristic, since military activities cannot be a method (how) of the resources.

Another difference is the abstraction level with which the domain functions (or general functions) are described. Martinez et al. (2001) described this level at a more functional level of activities (e.g. countermobility, engagement) while Jenkins (2012) used this level to describe military effects (e.g. prevent enemy attack, deter enemy). Again, the method of Martinez et al. (2001) does not conform with the means-ends characteristic. Considering these differences and the characteristics of the current domain, the structure for the abstraction levels about the domain by (Jenkins, 2012) will be used for the current study. The ADS they presented did not include the full set of functions, so domain functions might have to be added.

The two lowest levels of abstraction comprise the physical objects and their functions, independent of the domain. In (Baker et al., 2008) an abstraction hierarchy is specified for the Mars exploration mission domain and includes unmanned ground vehicles as technological resource. The hierarchy describes the activities of the UGV in high detail. Their description of the physical levels, in combination with the physical levels described by Jenkins (2012), will be used as the basis for the ADH for the current study. The resulting identified levels of abstraction are described as follows:

Domain purpose
The reason why the STUV-team is embedded to the CAT is to support the CAT to achieve its objectives and possibly to gather intelligence for high command. The domain purpose of the STUV-team is therefore split in two, these purposes are defined as *support tactical operation* and *gather intelligence*. This is the general purpose of the team and is independent of the resources, the situation and the time.

Domain values
The domain values are those values that can be used to assess the system's performance. The user values about purpose (see section 2.2.3) indicate what the user's find important about the system's performance. Since the purpose of the system is to support the CAT in achieving it's objectives, the domain values are the combination of the purpose values of the CAT-commander and the purpose values of the user, described in section 2.2.3.

Domain functions
The domain functions describe the abstract functions in the domain, unrelated to the physical resources. These functions are similar to military effects and describe the purpose of the physical functions. Domain functions are related to the domain values that they can affect. Several domain functions are defined by Jenkins (2012), some of the relevant[2] functions for the current domain are:

- Develop situational understanding
- Transport forces and supplies
- Target designation
- Communications extension (relay / gateway)
- Deter enemy attack
- Prevent enemy attack

Other functions relevant for the current domain presented in (Martinez et al., 2001) are:

- Control of terrain
- Information collection
- Air space denial

Physical functions
The physical functions describe the functions of the physical objects independent of the domain. These functions can be described at different levels of the decomposition space, e.g. component level and vehicle level. For example, the component *wheel* has the function *spin*, the object *UGV* has the function *move*, and the unit *platoon* has the function *move in convoy*. Functions at a higher level in the decomposition space typically require several functions at a lower level, such as *radio communication* to be able to *move in convoy*. Physical functions are typically standardized and described with technical and tactical activities (see section 2.3.2). These activities describe the functions that units can

[2]Some of these domain functions (or military effects) suggest that the use of weapons is required, this is outside the scope of the current study

execute, so there is no need to check if all required functionalities for an activity are covered at a lower decomposition level. The current study also relies on these ”standard” functions of units, so component functions will not be described. The functions of the UVs are described in section 2.5.

Physical objects
The physical functions are executed by the *physical objects* such as soldiers, robots, weapon systems and ammunition. These resources can also be defined with different resolutions. The current study does not describe the resources at a more detailed level than vehicle level, since more detail is not required for the current study. The physical resources provide the method (the how) of the military activities.

The resulting abstraction decomposition hierarchy is shown in Figure 2.3. Note that the hierarchy might not cover all activities in the domain, an additional analysis about the full domain is required to find the complete ADS.

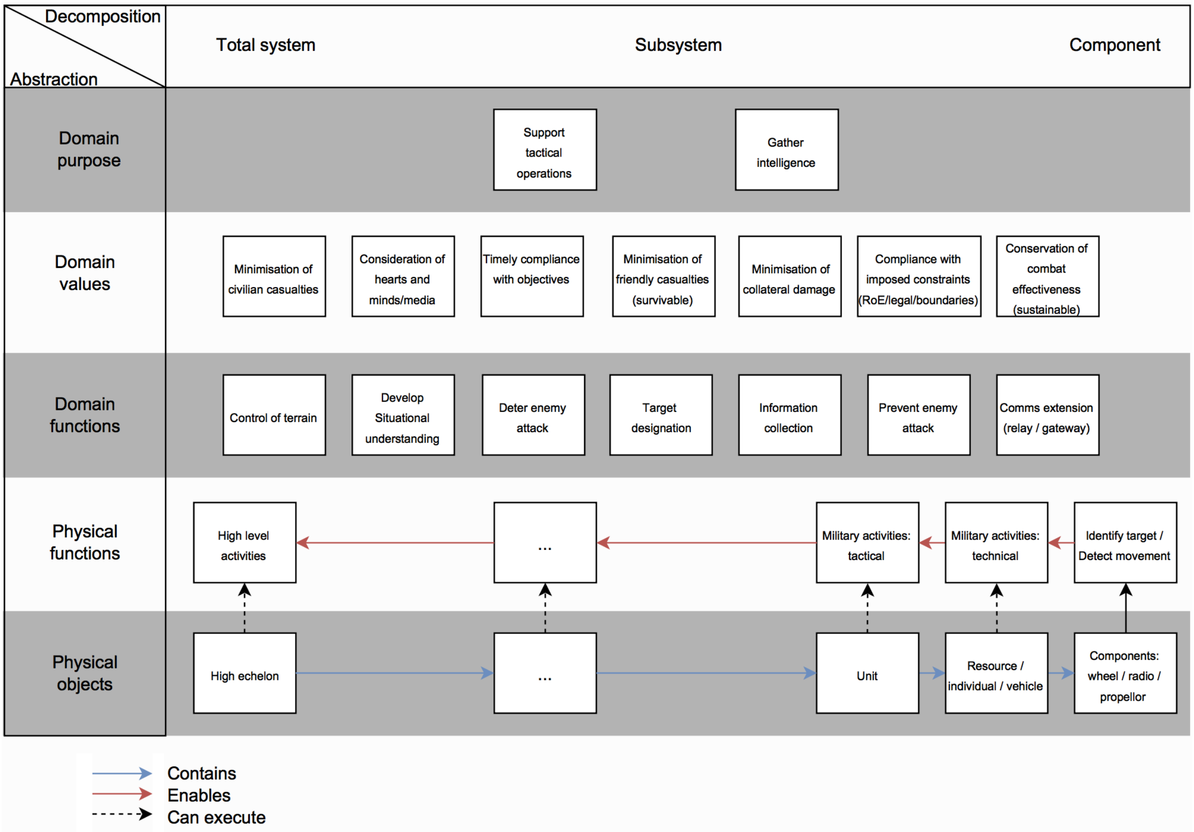


Figure 2.3: The abstraction decomposition space (drawn with www.draw.io)

Commander’s intent
During an operation the user focuses on the tasks, the resources and the intent of the (CAT) commander. To repeat short, the commander’s intent (see Section 2.4.4) contains three elements:

- *The purpose*: describes the purpose of the mission, these are also known as the military effects;
- *The endstate*: describes the physical functions to be achieved within the scope of the mission;
- *The method*: describes the way to achieve the endstate.

The CAT-commander needs to achieve a certain physical endstate. The method describes how the endstate will be achieved physically using the subordinate units. The projection of the commander’s intent to the abstraction decomposition space is shown in Figure 2.4

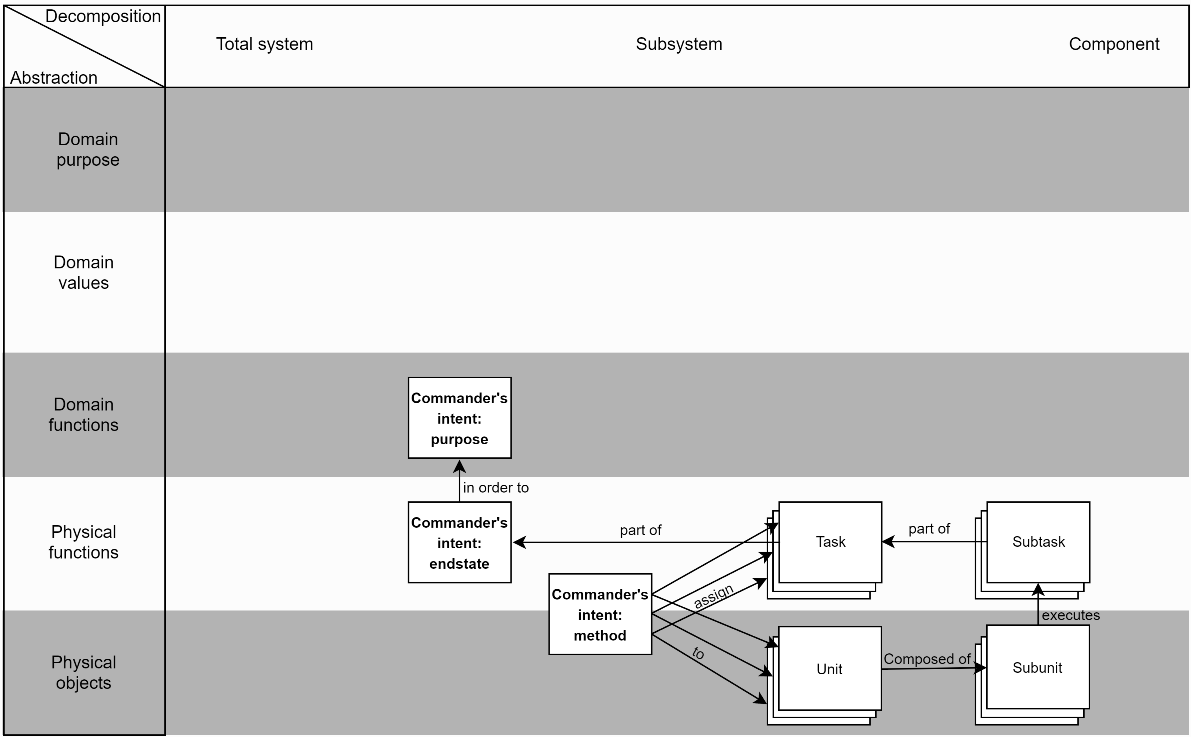


Figure 2.4: The commander's intent related to the abstraction decomposition space (drawn with www.draw.io)

The main focus of users is on achieving the tasks assigned to their unit in order to achieve the task of the CAT, so at the level of physical functions. The abstraction level below describes the physical resources (e.g. UVs), the level above describes the domain functions (military effects). However, on the level of physical functions the user needs to focus on two levels of the decomposition space, the own dedicated tasks and the task of the CAT they contribute to. eSupport should show the relations between information at these three levels of abstraction (RQ0005). The interface should organize the information so the information about the task, the method respectively the purpose is clustered (RQ0006) (Linegang et al., 2006).

2.6.3. Task delegation

The user has to allocate tasks to assets, this can be compared to the delegation of tasks between humans. To delegate tasks to the automation, users must be able to express their objectives together with constraints. (Clare et al., 2012c) found that performance increases when users are able to express their objectives to the automation, but it is important to establish trust in the system.

Miller et al. (2005) propose a human-automation integration architecture called Playbook to delegate tasks to automation. The concept uses predetermined plays that are understandable for both the humans and automation. A play is a high level plan that requires the completion of lower level tasks. For each play the user can provide parameters or constraints for the automation, this will help the user to exercise control on high level. This method ensures that there is a common understanding between the user and the system about the lower level tasks to be executed. The automation creates a plan based on the high level objectives (e.g. tasks and intent) of the user to complete the play while satisfying the constraints.

An evaluation of the system by Parasuraman et al. (2003) with users controlling up to eight agents, proved that the Playbook interface supports effective user supervision of agents, the interface led to better results than with only direct control or fully automated plays. The evaluation provided strong support that ”the Playbook allows for effective tasking of multiple agents while keeping the user in the decision-making loop, without increasing user mental load, and allowing the user to adapt successfully to unpredictable changes in the environment” Parasuraman et al. (2003).

Due to the positive results with Playbook, the principles will be used in the design of eSupport. The plays used in Miller et al. (2005) are similar to the abstraction levels about physical functions, defined in the abstraction hierarchy. Plays also enable the use of constraints. For eSupport, there will be two kinds of constraints available to the user to enable interaction between the user and eSupport for the evaluation: 1. Denying an asset for a specific play (*asset n cannot execute task m*) and 2. The order of plays (*task i [starts/ends] [before/after] task j [starts/ends]*).

2.6.4. Cognitive task load

For optimal performance, the cognitive task load should be in ’the sweet spot’. This means that the user’s tasks should be challenging but still manageable while the user keeps a sufficient level of SA (Chen et al., 2011). Mosier et al. and Cummings and Guerlain (as cited in Chen et al., 2011) conclude that participants made quicker and less-accurate decisions, not using all available information, when the task load was higher. To get the task load in the sweet spot, a balance has to be found between putting the user in the decision loop and more autonomy (Barnes et al., 2014). Optimally, the tasks should be divided over humans and autonomy to utilize the most effective characteristics of both humans and autonomy (RQ0007), resulting in improved system performance (Steinhauser et al., 2009, as cited in Chen et al., 2011).

The sweet spot is affected by various factors such as task difficulty, the type of unreliability of the automation and individual differences between users (Chen et al., 2011). The effects of these factors are not always consistent. For example, even imperfect automation may be preferred to human involvement when the user’s task load is too high (Barnes et al., 2014). This also holds for eSupport, since automation typically performs better than humans at planning and task allocation in complex and high-pressure domains such as military operations (Clare et al., 2012b). Therefore, automation sometimes should take over tasks of the user.

The appropriate balance between execution of tasks by humans or by automation is however not always straightforward and cannot be predicted. According to (Barnes et al., 2014), this balance should not be fixed but rather should adapt to the workload of the user, this is called Adaptive Automation (AA). AA involves the user in the decision process for automated tasks whenever the workload allows it and keeps the user aware of the status of the automated tasks. AA can ”leverage the strengths of humans and computers to improve overall system performance while mitigating the negative aspects of both” (Barnes et al., 2014). However, the users are still responsible for their tasks, so at least initially, the decision authority of the users should not be taken away (Barnes et al., 2014) (RQ0008) and the users should be involved in the system’s process to not feel unnecessary (Steinhauser et al., 2009, as cited in Chen et al., 2011)(RQ0007).

Another important conclusion of Chen et al. (2011) is that the attention of the user is attracted to the visually salient representations, even when the information is not the most critical for the task and especially under time pressure. Therefore, a system should aim to display and accentuate the essential information for a decision using visually salient representations (Nam et al., 2009, as cited in Chen et al., 2011)(RQ0009).

Components of CTL

The cognitive task load (CTL) of a user consists of three load dimensions (Neerincx et al., 2009). The first is the time the user is occupied to complete a task, the second is the level of information processing resulting from the cognitive complexity and the third the number of task switches requiring different sources of human task knowledge. See Figure 2.5 for an overview of the effects. When the cognitive task load of a user is too high (cognitive task overload), the abilities of the users can be negatively affected, possibly resulting in counterproductive interruptions of the automation by the user (Jentsch and Fincannon, 2012; Barnes et al., 2014). Too low CTL (cognitive task underload) might lead to loss of attention and a decrease of SA.

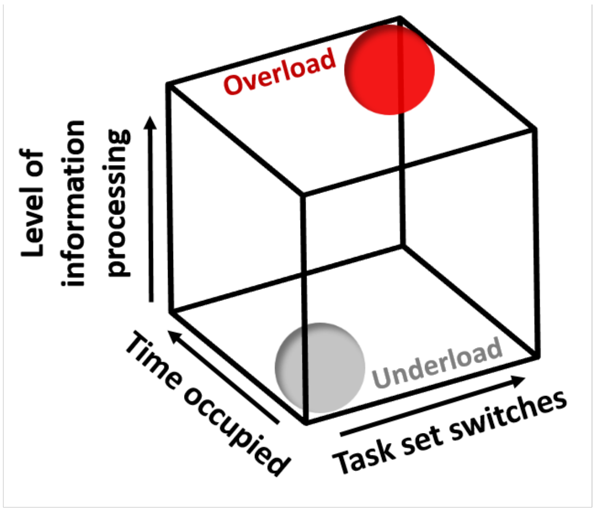


Figure 2.5: The three dimensional model of cognitive task load (Neerincx, 2003). The authors defined four general problem regions, the most relevant are displayed here

In the experiment, the users only have to focus on their task to create a schedule and check for mistakes. No other information, besides one message at most, is shown to the users during a mission. Therefore, participants practically do not have to switch task-set. Besides, all users have a target time to complete the mission, this is necessary to restrict the required time for an experiment. Therefore, not much variance in occupied time is expected. As a result, the CTL of participants during the experiment is expected to be influenced mostly by the level of information processing resulting from the cognitive complexity of the task.

The cognitive task load will be measured during the evaluations. This measurement does not result in an indication if the CTL was balanced, however allows for a comparison of CTL between conditions.

2.6.5. Trust

Trust is ”the attitude that an agent will help achieve an individual’s goal in a situation characterized by uncertainty and vulnerability” (Lee and See, 2004, as cited in Chen et al., 2011). The user’s trust in a supportive tool is a crucial factor for performance, due to the interactive nature of human-automation collaborative systems (Clare et al., 2015). User trust in automation depends on the initial trust and the behavior of the system throughout an operation and possibly fluctuates throughout an operation. Both overtrust and undertrust have a negative impact on the performance and the user’s SA and CTL (Chen et al., 2011; Clare et al., 2015). Overtrust of the user can lead to automation bias, also called automation-induced complacency, causing the user to follow the decisions of the automation, without further checking these decisions, the information they are based on or contradicting information. This can lead to dangerous situations resulting in negative affects and even catastrophic accidents (Parasuraman and Riley, 1997). Undertrust might cause the user to continuously overrule the autonomy, even while this results in worse performance caused by e.g. biased decision making. It might also cause the user to check all automated decisions resulting in a significant increase of workload, distracting focus from essential tasks (Chen et al., 2011; Cummings et al., 2012).

For appropriate human-automation collaboration the trust of the user should be calibrated. Maximizing user trust is not the objective, instead the user should know when to trust and when to override the autonomous system (Lee and See, 2004, as cited in Barnes et al., 2014; Chen et al., 2011). ”Especially in combat, humans will always be the final arbitrator of safety issues, and, as such, they must know when autonomy is safer than human intervention and vice-versa” (Barnes and Evans III, 2010). In previous research various ways to calibrate trust were suggested. An important conclusion of prior studies is that trust can be biased through earlier experiences with the system (Mercado et al., 2016)

Dzindolet et al. (2001) concluded that the order in which automated solutions are presented to the user has an impact on the user's trust in the system. When the automated solution is shown at the same time the problem is shown, the user tends to over rely on the automated solution, however when showing it after the user made their decisions, the user tends to rely on their own decisions even when it was suboptimal. A possible explanation is that the user tries to reduce their cognitive task load, since reconsidering a solution would increase their task load. A system should therefore present a solution or suggestion at appropriate times during the decision making process of the user (RQ0010).

Lee and See suggest that the operations and algorithms of the automation should be simplified to be understandable for the user (Lee and See, 2004, as cited in Chen et al., 2011). This includes dividing (complex) tasks into smaller tasks to simplify the tasks (Steinhauser et al., 2009, as cited in Chen et al., 2011). So, the automation process and algorithms can be made more understandable by dividing complex tasks, simplifying operations and algorithms and revealing the intermediate results to the user (RQ0011). Training the users besides gives them insight in the reliability of the automation and the intended use, training is not included within the scope of the current study.

2.6.6. Measurements

This section describes the factors that will be measured during the evaluation. The first section describes the extraneous factors, these are factors that cannot be manipulated during the experiment but might influence the results of the experiment. The second section describes the dependent variables, identified in Section 1.2.

Independent variables

This section discusses the relevant independent variables as such identified as relevant in prior related research. (Chen and Barnes, 2012a; Mercado et al., 2016) suggested that several individual differences might impact the collaboration between the user and eSupport. These factors are user's spatial ability, attentional control, video gaming experience and working memory capacity. Since the regular tests for working memory capacity took over 10 minutes and time is critical for the experiment, this measure was not included in the current study. The other measures are discussed below. Also decreased color vision might influence the outcome due to the use of colors in eSupport, this measure is also included.

Spatial Ability

Spatial ability (SpA) is the ability to perceive spatial patterns or to maintain orientation with respect to objects in space (Ekstrom et al., 1976). Several studies have identified SpA to be an important factor for the performance of controlling multiple vehicles (Chen and Barnes, 2012a). According to these studies SpA has a significant influence on a user's ability for effective visual scanning and target detection and according to (Mercado et al., 2016) spatial ability may have effects on operator response time. However, the current study assumes a fixed situation without movement of vehicles and without rotation of the map. In the comparable study of (Mercado et al., 2016) the authors measured the effect of spatial ability, they concluded that there were significant effects on fixation duration (a difference of around 20ms (10%)) and pupil diameter. Special equipment is however required to measure these effects. Considering these expected effects, the time it takes to measure these effects during the experiment and to test the participant's SpA prior to the experiment, we choose not to include spatial ability as a measure.

Attentional control

Attentional control (AC) is defined as one's ability to focus and shift attention in a flexible manner (Chen and Barnes, 2012a). Multiple studies have shown that people with higher AC are better in switching their attention flexibly and effectively in a multitasking environment. Studies also repeatedly demonstrated that people with lower AC would rely more on automated systems to save their cognitive resources in a complex information processing environment. This effect should be taken into account in the evaluation of the test results. Therefore, a questionnaire of (Derryberry and Reed, 2002) will be used to evaluate participants' attentional control (PAC). The participants are categorized as either low PAC or high PAC based on their score compared to the median (Derryberry and Reed, 2002).

Video game experience
Former experience with video games influences the performance of users for tasks that require for example visuospatial selective attention, rapid process of visual information and imagery and flexibility in attention allocation (Chen and Barnes, 2012a). In their evaluation the video gaming experience had a significant impact on the overall multitasking performance of the participants. This factor will therefore be taken into account for the evaluation with the users.

Color vision test
Participants who cannot indicate they are not color blind, have to complete the Ishihara Color test. If they fail the test, a test in the training mission is done to see if they are hampered in the task. Participant's will not be excluded from the experiment.

Dependent measures
The current study investigates the changes in user performance during the scheduling phase as an effect of the additional transparency in terms of efficiency, appropriate (non-)compliance and situational awareness. Since the measures for performance, defined in terms of scheduling duration and appropriate (non-)compliance, measure a simple construct (time respectively amount of appropriate acceptances or rejections) no measurement techniques from literature are discussed for these measures.

Calibrated *trust* and balanced *CTL* are important conditions for appropriate (non-)compliance, therefore also *trust* and *CTL* will be measured during the experiment and discussed in this section. Additionally the *usability* is also measured to evaluate the design and to test what elements of the interface contributed most to the effects. This section gives an overview of current measurement techniques for the dependent variables, Section 4.2.3 describes the applied measurement techniques for the current study in more detail.

Situational awareness
Various measurement techniques are available for situational awareness assessment. In (Endsley et al., 1998) the authors compared two well known techniques, the *Situation Awareness Rating Technique* (SART) and the *Situation Awareness Global Assessment Technique* (SAGAT) (Endsley, 1995). The authors concluded that the measurements for SA surprisingly were not correlated and thus were assessing something different. The SART technique is a subjective measurement and strongly correlated with the participant's confidence level in SA, so the authors hypothesized that this was not an objective measure for SA. According to the authors, the SAGAT technique seemed a more objective measure.

To use SAGAT, a number of questions about the current situation are posed to the user whereby only one answer is correct. The assessment technique requires to pose question about a situation directly at that moment. Sometimes that requires to freeze the simulation, however (Endsley, 1995, 2000) concluded that this has no effect on the results. This method fits the purpose of the current study.

Trust
Human trust influences human-machine collaboration in several ways. The questionnaire for trust between people and automation, defined in (Jian et al., 2000), is a common used measure for trust and the questionnaire for the current study is based on this. Several questions are left out, since they seem irrelevant for the current system and could lead to confusion.

Despite trust can be biased through earlier experiences with the system (Mercado et al., 2016), the questionnaire is used at the end of both conditions per participant to identify a possible effect of the other condition within subjects. Besides the questionnaire, other indicators can be used often to measure trust. The current study uses *correct usage* and *correct rejections* of solutions (schedules) as additional measures, Section 4.2.3 describes these measures in more detail.

Cognitive task load
The cognitive task load of the user is likely to influence the performance (Chen and Barnes, 2012a). The cognitive task load can be measured with the NASA Task Load Index (TLX) (Hart & Staveland,

1988, as cited in Chen and Barnes, 2012a). The TLX contains a questionnaire with six measures of CTL. Besides, the test contains a questionnaire about the relevance factors of each measure, so participants can indicate how important a factor was for their experience of workload for the task. There is an ongoing discussion about the usefulness of the additional questionnaire, since the results of both tests are highly correlated (De Winter, 2014). As a result, the relevance factors are often not used. The raw test without the relevance factors is called the Raw Task Load Index (RTLX). Our experience in the preparation of the experiment was that the questionnaire about relevance factors was not intuitive for the current experiment, it was hard to indicate what factor was more important for CTL. Because of this and because the usefulness is heavily doubted, we decided to use the RTLX.

Another used measure for the cognitive task load is the reaction time of the user to answer an electronic question during task execution. This measure can be used to validate the values of the questionnaire for CTL.

Usability
The current study includes the perceived usability of the user as a measure to investigate the possibility for further improvements of the system. (Brooke et al., 1996) developed a short survey to measure usability. Although the questionnaire, besides effectiveness and satisfaction, also covers efficiency, efficiency will still be measured apart. Besides the questions of the short survey, the participants receive additional questions about the interface elements for the mission block with eSupport to evaluate the values of these elements.

2.6.7. Conclusions

Several requirements for the design of eSupport are identified from the literature about human factors. The literature also led to design constraints for the current study. An overview is provided below.

Requirements

- **RQ0002**: eSupport supports good usability to prevent a negative impact on cognitive task load and performance
- Situational Awareness
  - **RQ0003**: eSupport shall not overload the user with data by presenting only relevant data from the appropriate levels of abstraction (the purpose of the mission, the endstate of the mission, the tasks to achieve the endstate and the method that provides the link between the resources and the tasks), with an appropriate level of detail for the current task
  - **RQ0004**: eSupport shall show the consequences of specific decisions or changes of the user (or the environment) at the right level of abstraction. The consequences for the total performance are communicated to the user
- Abstraction Hierarchy
  - **RQ0005**: eSupport shall show the relation between information of physical objects, physical functions (own tasks and the related parent tasks) and domain functions. Other data shall be integrated with this information to avoid loose data
  - **RQ0006**: The interface shall organize the information so the information about the tasks, the method of executing the tasks respectively the purpose of the tasks is clustered
- Cognitive Task Load
  - **RQ0007**: eSupport shall compute a schedule and enable the user to set or change the objectives and constraints for the schedule whenever desired by the user
  - **RQ0008**: eSupport shall (at least initially) not take decision authority away from the user by asking permission to the user to execute the proposed schedule or by allowing the user to override the schedule
  - **RQ0009**: eSupport shall use visually salient representations to attract the attention of the user to the most critical information for the task
- Trust

- **RQ0010**: eSupport shall present a schedule and the resulting performance during each step of the decision making process of the user, so during each step of the scheduling process, not only at the start nor the end of this process
- **RQ0011**: eSupport shall support the user's understanding of its behavior by separating the scheduling process in several steps. The steps are: 1. set intent, 2. set allocation constraints, 3. set task-order constraints and 4. acknowledge schedule

Design constraints

- **DC0006**: No new data about the environment becomes available during a scenario
- **DC0007**: The level of autonomy in eSupport is fixed and not included in the analysis

2.7. Discussion

The previous sections described the analysis of the domain. This analysis led to several requirements for eSupport to support good usability and performance, calibrated trust and balanced CTL. These requirements are the basis for the design of eSupport. Furthermore, the analysis provided sufficient knowledge to find solutions for research questions 1a, 1b and 1c, these solutions are discussed below.

2.7.1. Research question 1a

In Section 1.2, research question 1a was defined as: *What are the user's main objectives considering the allocation of high-level tasks to heterogeneous unmanned vehicles within a military operation?* The analysis resulted in a solution for this question. The main objectives of the operator for the allocation of high-level tasks to UVs are defined in Section 2.5.4 and represent some of the user values. Only those variables the algorithm can reason about are included, to make sure that a setting for an objective has an effect on the solution (the schedule). We defined five objectives for eSupport: 1. Time pressure, 2. Risk exposure, 3. Massiveness, 4. Secrecy and 5. Information quality and quantity. The latter variable is however fixed, since this is an objective specified in the intelligence requirement.

2.7.2. Research question 1b

The second research question, defined in Section 1.2, is: *How can the user's main objectives defined in research question 1a be communicated to eSupport in order to allow eSupport to reason about a schedule's compliance with these objectives?*. In Section 2.4 of the analysis we found that all objectives of the operator are captured in the Commander's intent. The Battlefield Management Language allows to communicate the commander's intent to an electronic system, however this language requires more development. Therefore, we propose a proven method to communicate intent via loose parameters, see Section 2.5.3. So the solution to research question 1b is, that the users can indicate their objective for each parameter defined in research question 1a, on a 5-Likert scale:

- Time Pressure: Very low pressure (1) - Very high pressure (5);
- Risk exposure: Take no risk (1) - Take risk (5);
- Massiveness: Use few resources (1) - Use many resources (5);
- Secrecy: Overt - Covert;

When more parameters (objectives) are introduced, this method requires more effort from the user and possibly more time to get used to the parameters. Therefore, we recommend to create fixed sets of parameters for certain terms defined in the commander's intent in the future. These terms could be communicate using the Battlefield Management Language.

2.7.3. Research question 1c

Research question 1c is defined in Section 1.2 as: *How can eSupport provide more transparency to the user about a schedule's compliance with the user's objectives?*. Section 2.5.4 showed that the objectives of the user are in conflict, therefore an optimal solution is never straightforward, rather depends on the preferences of the user for the specific situation. We refer to these conflicting objectives as dilemmas for the algorithm, since they require knowledge about the user's preferences and objectives to decide what the best solution is. Therefore, we want to show these dilemmas to the user to support understanding of the reasoning. Figure 2.6 shows the conflicts between the defined objectives.

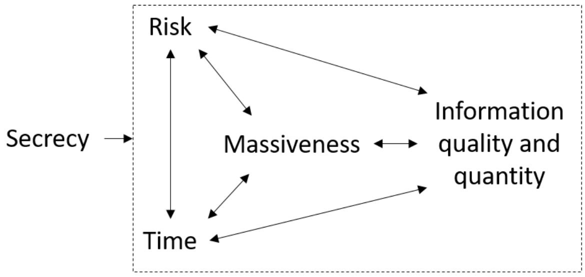


Figure 2.6: Overview of the conflicts between the objectives

The figure shows that there are six dilemmas between four objectives. A method is required to show the conflict between these dilemmas to the user, so to clarify that optimization for one variable affects the other variables. For the design, we also have to take into account that the amount of objectives will increase in the future, thus should not be based specifically on the defined parameters.

To prevent information overload and to support user's understanding, it is also important to consider at what abstraction level the information will be presented. In Section 2.6.2 we defined the focus levels of the operator. To repeat short, the main focus of the operator is on the tasks that have to be executed by the resources. The level above was the endstate of these tasks, which is at the same level of abstraction but at a higher level of decomposition. The endstate has to be achieved in order to achieve a military effect, the purpose, at one abstraction level higher. The transparency should provide information about the consequences of decisions, so the consequences of the allocation of a task to a vehicle. These consequences can be found at the higher levels in the ADS, so the endstate of the tasks and the purpose of the mission. This information is available in the missions hierarchy, therefore the transparency will also be provided in the missions hierarchy.

Considering these requirements, we propose to use a pie chart icon in answer to research question 1c, so to provide more transparency. These icons can show a solution's compliance with each user objective and the mutual ratio. Each component of the pie chart represents a parameter equal to the objectives, indicated with a letter (e.g. R for risk and T for time). The size of each component represents the value of the compliance with an objective relative to the other components. The color of each component indicates the degree to which the solution complies with the user's objective, red indicates very low compliance, yellow indicates average compliance and green indicates full compliance. See Figure 2.7 for the design. The pie chart is similar to the sprocket graphic presented by Mercado et al. (2016). The authors used the graphic to increase transparency for a similar purpose and showed that the graphic improved the transparency of the system. The authors indicated that the pie chart can also be used to show the reliability of the information by using a more transparent color for less reliable values.

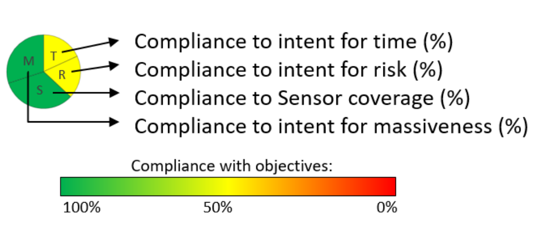


Figure 2.7: The pie chart indicates the compliance of parameters to the user's objectives

The user should be able to request more details about the solution's compliance with the objectives to support understanding. Therefore, the actual performance value related to each objective should be shown. The units of these values depend on the variable, e.g. percent, seconds or a number. The unit for the compliance with the objectives are percentages. A pilot study with three participants showed that it was confusing when the performances values and the degree of compliance with the objectives were shown both when the latter was shown in percentages. The confusion is understandable, for example, when a participant indicates low acceptance of risk exposure, and the performance measure shows there is much exposure to risk for a certain solution (70%), the compliance with intent is low (20%). So, the values have contradicting meanings in the example, the performance value should be as low as possible and the compliance with the intent should be as high as possible. This makes the meaning of values less intuitive for users. Therefore, we introduce rating stars to indicate the solution's compliance with the objectives of the user. Rating stars follow a universal rule that more stars equal a better rating. A pilot study indicated that this led to better understanding, so less confusion of participants.

2.8. Conclusions

The analysis about the domain led to several requirements for the design of eSupport and to solutions for research questions 1a-1c. The requirements are based on prior research and should be met in order to support good usability, user performance, calibrated trust and balanced CTL. Furthermore, the analysis resulted in a method to communicate the user's objectives to eSupport and to provide more transparency about the compliance of a solution with these objectives.

In this chapter, several design assumptions are made together with several design constraints. An overview is given below with the pages where it is provided.

Design assumptions

DA0001: A team of UVs can be embedded to a combined arms team (CAT) during future military operations

DA0002: The UVs are combined in a team, an operator is responsible for the unmanned vehicles and likely several additional human enablers

DA0003: The relevant information about the operation is communicated via a Battlefield Management System (BMS). Authorized stakeholders can insert, prioritize and manage intelligence requirements and can see the status through BMS at all times in a way that the user can easily process it. eSupport has access to all the information available in BMS.

DA0004: The UVs are able to autonomously execute the high-level activities person reconnaissance, vehicle reconnaissance and building reconnaissance

DA0005: A five-level Likert scale provides users enough freedom to indicate their preferences for each intent parameter

DA0006: The parameters *time pressure*, *risk acceptance*, *massiveness* and *secrecy* are sufficient to align the user's objectives with eSupport in order to allow eSupport to suggest a schedule that satisfies the user during the experiment

DA0007: The analysis and design of eSupport does not include a validation of the CommonSense Framework as a BMS, the current study assumes that the framework functions well as BMS, covers assumption DA0003 and is a proper basis for eSupport

Design constraints

DC0001: The study focuses on the design of eSupport to support the commander of a UV-team with their related activities

DC0002: The analysis and design of eSupport do not include the development and the interaction of the user with the Battlefield Management System, nor the interaction of eSupport with the agent(s) in control of the autonomous UVs

DC0003: The analysis and design of eSupport do not include the presence of friendly forces, deconfliction with UVs nor the required coordinating measures

DC0004: The analysis and design of eSupport do not include terrain and weather and the consequences for the operation and UVs

DC0005: The analysis and design of eSupport do not consider the fuel consumption of UVs and the related consequences
DC0006: No new data about the environment becomes available during a scenario
DC0007: The level of autonomy in eSupport is fixed and not included in the analysis

2.8.1. Validation of assumptions

Based on the current initiatives in the Dutch military forces and the US Marine Corps and the opinion of the SMEs (see the interviews in Appendix A), assumption **DA0001** about embedding a STUV-team to a CAT and assumption **DA0002** about dedicating a UV-operator to a team of UVs and possibly additional enablers seem valid. Assumption **DA0003** is also considered valid, moreover it is unthinkable the system would not be integrated with BMS. The recently introduced command and control-support system in the Dutch Army allows all commanders to indicate their intel requirements. However the intel requirements are not visualized in BMS in the current setting, this should be enabled for eSupport and is technologically possible. Once the intel requirements of all units at all levels are available for eSupport, this also allows more proactive behavior of eSupport to advise the operator about additional achievable tasks of other units. Assumption **DA0003** about sharing intel requirements through BMS is considered valid.

3
Design and specifications

This chapter describes the design of eSupport in more detail. The first section describes the scenarios to provide more context for eSupport and claims to justify the requirements for eSupport. The design section goes into more depth about the suggested solutions for these requirements. These suggestions are formalized as proto-patterns as a generic solution to a recurring design problem that has to be tested. The prototype of eSupport for the experiment is discussed in the last section.

3.1. Scenarios and Claims

Scenarios describe the activities of the user and support better understanding of the context of the problem. A problem scenario is created for the current study and described in Section 3.1.1. The activities are structured in use cases to provide the (formal) contextualization (who, when, where) of the activities. Use cases require certain functions or satisfaction of certain user values for specific activities, therefore requirements are attributed to them. The justification of these requirements in terms of user values is formalized with claims. Claims describe which requirements result in user values and their trade-offs.

3.1.1. Problem scenario

The problem scenario describes a realistic scenario for the use of a STUV-team, based on a scenario for the deployment of multiple heterogeneous unmanned vehicles, derived from an inhouse project of TNO. The official document cannot be released. The scenario has two components, a general situation description and more specific descriptions for a mission.

The general situation:

The STUV-team is part of a European Force assigned to maintain a zone of separation between two countries. Both countries house parties that want to attack the other country. EU force has to monitor both sides of the border, to prevent border crossings or invasions from both countries to the other.

The general context of the scenario is described in more detail in Appendix G. Based on this scenario, eight specific scenarios (missions) are created. These missions are similar to a military fragmentary order and describe the specific situation at location, the mission, the commander's intent, specific information from the military analyst and the combat engineer advisor and further restrictions. See Section G.2 for the missions.

The activities of the operator within these missions are described in design scenarios in the next section.

3.1.2. Design scenario

This paragraph describes the operator's activities within the operations in more detail. To provide more structure within the design scenarios the operator's activities are split up in six phases. These phases are identified in the in-house project of TNO and are described below.

Phase 1. Pre-Mission Activities: Mission data is loaded such as terrain, terrain features and the Mission Task Order and the communication with the UVs and the commander is established. The UVs are physically prepared for deployment.

Phase 2. Initiate Mission: The operator analyzes the mission defined by the commander and defines the tasks to be performed by the UVs: the task allocation. The tasks are communicated to the UVs on the end of this phase.

Phase 3. Control Mission Execution: The operator monitors the UVs and the changes in the environment. The operator updates the list of tasks to be performed and adjusts the schedule accordingly.

Phase 4. Monitor Mission Progress: The operator receives updates about the mission progress based on the status of the UVs and the data the UVs collected. The operator updates the list of tasks to be performed and adjusts the schedule accordingly.

Phase 5. Event Management: When an event occurs the operator replans the mission and updates the list of allocated tasks to UVs.

Phase 6. Manage Post Mission Activities: The post mission activities are managed such as the extraction of technical and tactical data from the UVs. The operator decides about the mission termination.

The study focuses on Phase 2: *Initiate Mission*. The activities within this phase are described in more detail in the design scenario. The design scenario holds for all missions defined in Section 3.1.1, since the activities in this phase require the same work flow despite the environment and mission.

The design scenario is provided for two conditions, one for a situation without additional transparency (a) and one for a situation with additional transparency (b). Because the conditions only differ for two activities, the design scenarios are provided in one overview.

Activities from phase 2: Task allocation to assets (without (a) and with (b) additional transparency)
The operator allocates the tasks to the assets

1. The operator sees the mission, the belonging tasks and an overview of the available assets on the screen.
2. The operator specifies their intent for the operation.
3. The operator starts the planning tool and does not want the person identification to be done by a certain UV, so indicates this constraint to eSupport.
4. The operator requests a schedule.

5a eSupport computes the most optimal task allocation that satisfies the constraint and shows the schedule to the operator in a time line. In this time line all assets are shown with their dedicated tasks and the travel time. eSupport also shows the task allocation on the map.

5b eSupport computes the most optimal task allocation that satisfies the constraint and shows the schedule to the operator in a time line. In this time line all assets are shown with their dedicated tasks and the travel time. eSupport also shows the task allocation on the map and the schedule's compliance with the operator's intent.

6 The operator sets a constraint on the building search to take place not earlier than the area reconnaissance has finished.

7a eSupport shows an adjusted schedule that satisfies the constraints.

7b eSupport shows an adjusted schedule that satisfies the constraints and shows the schedule's compliance with the operator's intent.

8 The operator accepts the new schedule and eSupport sends the orders to execute tasks to the agent in control of the UVs.

The activities of the actors are structured with use cases. Use cases define the structure and provide the (formal) contextualization (who, when, where) of the activities. The activities defined in the design scenario can be described in four use cases: *initial information*, *constraints on task allocation*, *first schedule proposal* and *second schedule proposal*. The use cases are described in Appendix H.

3.1.3. Requirements for eSupport

The requirements that resulted from the analysis are used in the design for the prototype of eSupport. The requirements are listed below.

RQ0001: eSupport allows users to share all their important objectives and constraints with eSupport at the appropriate level, so for the STUV-team, and eSupport shall compute a (near) optimal schedule that satisfies the objectives and constraints of the user

RQ0002: eSupport supports good usability to prevent a negative impact on cognitive task load and performance

RQ0003: eSupport shall not overload the user with data by presenting only relevant data from the appropriate levels of abstraction (the purpose of the mission, the endstate of the mission, the tasks to achieve the endstate and the method that provides the link between the resources and the tasks), with an appropriate level of detail for the current task

RQ0004: eSupport shall show the consequences of specific decisions or changes of the user (or the environment) at the right level of abstraction. The consequences for the total performance are communicated to the user

RQ0005: eSupport shall show the relation between information of physical objects, physical functions (own tasks and the related parent tasks) and domain functions. Other data shall be integrated with this information to avoid loose data

RQ0006: The interface shall organize the information so the information about the tasks, the method of executing the tasks respectively the purpose of the tasks is clustered

RQ0007: eSupport shall compute a schedule and enable the user to set or change the objectives and constraints for the schedule whenever desired by the user

RQ0008: eSupport shall (at least initially) not take decision authority away from the user by asking permission to the user to execute the proposed schedule or by allowing the user to override the schedule

RQ0009: eSupport shall use visually salient representations to attract the attention of the user to the most critical information for the task

RQ0010: eSupport shall present a schedule and the resulting performance during each step of the decision making process of the user, so during each step of the scheduling process, not only at the start nor the end of this process

RQ0011: eSupport shall support the user's understanding of its behavior by separating the scheduling process in several steps. The steps are: 1. set intent, 2. set allocation constraints, 3. set task-order constraints and 4. acknowledge schedule

3.1.4. Claims

Claims justify the requirements by describing their effect on the user values. The user values are separated in values about the interaction with eSupport (*system interaction values*), and values about the schedule itself (*purpose (schedule) values*). The values are identified in Section 2.2.3. The study focuses on the use of eSupport, therefore, only the *system interaction values* are discussed in this section.

Effectiveness

eSupport produces a schedule that complies with the user's objectives, therefore it enables the user to work more effectively.

The trade-offs are:

+ The user gets the expected result and does not have to correct the result.

- The user might want to force the automation too much towards its expected result, without considering if that is the most optimal result.

The requirement that results in effectiveness is RQ0001.

Efficiency
eSupport provides appropriate information at the right level of abstraction to minimize the required scheduling time and effort while maintaining the performance of the user.

The trade-offs are:

+ The user requires less time and effort.
- The user might not see important information.

The requirements that result in effectiveness are RQ0002, RQ0003, RQ0004, RQ0005, RQ0006, RQ0007, RQ0009 and RQ0010.

Situational awareness
The user has sufficient knowledge and understanding about the environmental elements with respect to time or space.

The trade-offs are:

+ The user is able to make fast decisions based on the most up-to-date information of the environment.
- The user might experience cognitive overload when too much information is provided, resulting in worse performance.

The requirements that result in situational awareness are RQ0003, RQ0004, RQ0005, RQ0006, RQ0009 and RQ0010.

Cognitive task load
eSupport provides the user with the right amount of information at the appropriate abstraction level. This leads to a balanced amount of mental processes that does not lead to under- or overload of the user.

The trade-offs are:

+ The user experiences an optimal cognitive task load, this supports concentration and performance.
- eSupport might not perform optimally due to the limitation of cognitive capabilities of humans.

The requirements that result in cognitive task load are RQ0002, RQ0003, RQ0004, RQ0007, RQ0009, RQ0010 and RQ0011.

Decision authority
The user has the final right to decide and can always overrule eSupport. eSupport does not start execution of the schedule before the user gave permission.

The trade-offs are:

+ This supports safety and the responsibility is clear.
- eSupport might not perform optimally when it has to wait for approval.

The requirement that results in decision authority is RQ0008.

Trust
eSupport gives the right information about the reliance of a suggested schedule to the user to calibrate the user's trust. This should prevent overtrust, which results in too much compliance of the user to the suggestions of eSupport, and undertrust, which results in overly controlling eSupport by the user.

The trade-offs are:

+ An appropriate level of trust supports the user to actually use the automation with appropriate reliance in eSupport's reasoning. This leads to more efficient human-automation cooperation since both human's and automation's skills are utilized.
- Calibrated trust requires provision of information to the user about the automation's reasoning and reliability, this might have negative effects such as to cognitive task load and efficiency.

The requirements that result in trust are RQ0001, RQ0004, RQ0010 and RQ0011.

Appropriate (non-)compliance
eSupport gives the right information about the performance of the schedule to the user and about the weak spots of eSupport. This leads to an increase of appropriate (non-)compliance.

The trade-offs are:

+ The user does know the weak spots and performance of eSupport and can estimate when it makes a mistake.
- The user only checks for weak spots of eSupport and misses other important information.

The requirements that result in appropriate (non-)compliance are RQ0001, RQ0002, RQ0004, RQ0009, RQ0010 and RQ0011.

3.2. Design

The identified requirements should be embedded in the design of eSupport, so eSupport can provide functionalities to the users. Based on the requirements and use cases, prototypes of interaction design patterns (proto-patterns) are designed. Interaction design patterns support specific functionalities and support generation of solutions for generic recurring human-computer interaction problems. Once proven effective, these patterns enforce re-use of knowledge of former research about human-automation interaction. Because the patterns we propose have not been proven to be effective yet, we refer to these prototypes of interaction design patterns as proto-patterns. This section describes the proto-patterns with examples for eSupport. For each proto-pattern, a rationale is provided that describes how the pattern is supposed to address the problem.

The proto-patterns are introduced per requirement. For convenience, the order of the sections deviate from the order of the requirements.

3.2.1. Aligning objectives and constraints

Requirement RQ0001 is defined as follows: *eSupport allows users to share their objectives and constraints with eSupport at the appropriate level, so for the STUV-team, and eSupport shall compute a (near) optimal schedule that satisfies these objectives and constraints of the user*. So the user has to be able to share their objectives and constraints with eSupport. This should be done at the start of the scheduling process, similar to the actual military command & control cycle. When the objectives are changed in a later phase of the scheduling process, the process will be started over, because these changes affect the whole schedule. Constraints can be added or removed during all phases of the scheduling process.

Proto-pattern design: objectives

eSupport has to enforce the user to think about several of their (implicit) objectives in an early stage of the scheduling process. This support the user to become more aware of their objectives. Besides, it allows eSupport to make a better schedule in terms of the user's objectives, since their objectives are aligned better. Therefore, we propose a pattern to allow the user to set the objectives, defined in Section 2.5.4, when the scheduling process is started. The settings are shown to the user at all times during the process in the interface. eSupport uses the settings to create a (nearly) optimal schedule in terms of the user's objectives. Figure 3.1 shows the design of the proto-pattern.

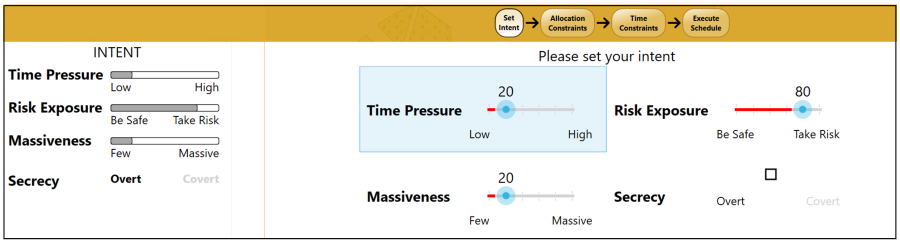


Figure 3.1: Proto-pattern to align the user's objectives with the objectives of eSupport

Proto-pattern design: allocation constraints

eSupport allows the user to set constraints for assets that shall not execute a task. Therefore, when a user clicks on a task, eSupport provides a list of the UVs. Behind each UV, the (local) compliance with the user objectives is shown if the task would be allocated to that UV. There is also a check box behind each UV. When this checkbox is deselected, the asset is not allowed to execute the task. See Figure 3.2 for the design of the proto-pattern. The bottom panel of eSupport provides a list of all allocation constraints.

Proto-pattern design: order constraints

eSupport allows the user to set constraints for the order of tasks. The user can indicate if a task should start or finish before another task starts or finishes. The user can drag an arrow from the start or end of a task to the start or end of another task to indicate that the first point has to take place before the

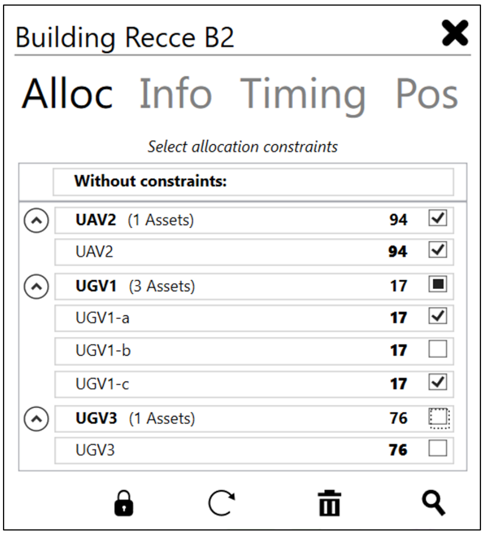


Figure 3.2: Proto-pattern for the menu to set allocation constraints using check boxes.

second point. Each task has a diamond at its start and end so an arrow can be drawn from and to these diamonds. See Figure 3.3 for the design of the proto-pattern.

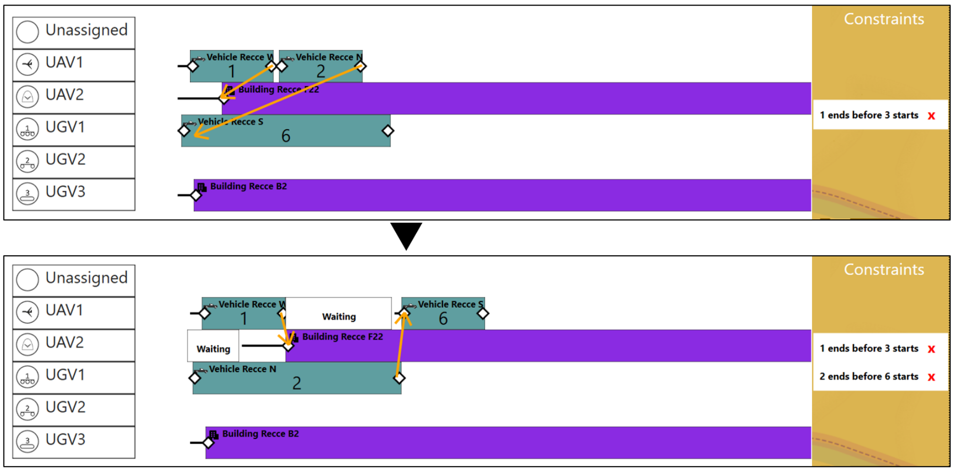


Figure 3.3: Proto-pattern for the menu to set order constraints using arrows. The figure on top shows the initial state, the figure at the bottom shows the result

Claims

The patterns to indicate the objectives and constraints result in better user trust in eSupport and the schedule and better effectiveness. The alignment of user's and eSupport's objectives reduces the chance that the user wants to incorrectly adjust the proposed schedule or accept an incorrect schedule (inappropriate compliance).

3.2.2. Good usability

Requirement RQ0002 is defined as follows: *eSupport supports good usability to prevent a negative impact on cognitive task load and performance*. This means that the design of eSupport should support ease of use and users should not experience trouble using eSupport. This is not something we can capture in a proto-pattern, rather all proto-patterns should be designed with this requirement in mind.

Claims

Good usability results in better performance in terms of efficiency and more appropriate (non-)compliance and in a decrease of CTL.

3.2.3. Appropriate information

Requirement RQ0003 is defined as follows: *eSupport shall not overload the user with data by presenting only relevant data from the appropriate levels of abstraction (the purpose of the mission, the endstate of the mission, the tasks to achieve the endstate and the method that provides the link between the resources and the tasks), with an appropriate level of detail for the current task*. So eSupport shall only provide information relevant for the scheduling task. Since this is the only task for the prototype, the requirement will not be integrated in eSupport but will be met manually by creating a fixed set of information relevant for the task for each abstraction level.

Claims

Showing only appropriate information at the right level of abstraction results in more efficiency, better situational awareness and a decrease of cognitive task load.

3.2.4. Transparency about performance

Requirement RQ0010 is defined as follows: *eSupport shall present a schedule and the resulting performance during each step of the decision making process of the user, so during each step of the scheduling process, not only at the start nor the end of this process*. eSupport should be transparent to the users about the performance of a schedule related to their objectives and the dilemmas. This supports the user's understanding of the reasoning behind a schedule. The additional transparency should be provided every time a schedule is proposed or updated. The additional transparency can be provided at a high level, so the total performance for each endstate and the total performance for the purpose of the mission. It can also be provided at local level about the performance of a resource for a task. Both levels seem meaningful, the design for both methods is described in two proto-patterns. Salient representations are used for both proto-patterns to attract the attention of the user to the most valuable information (RQ0009).

Proto-pattern design: Global compliance

eSupport shows the schedule's compliance with the conflicting objectives of the user for each endstate and the purpose of the endstate(s). The performance of these levels are shown with an icon, as defined in Section 2.7.3, see Figure 3.4.

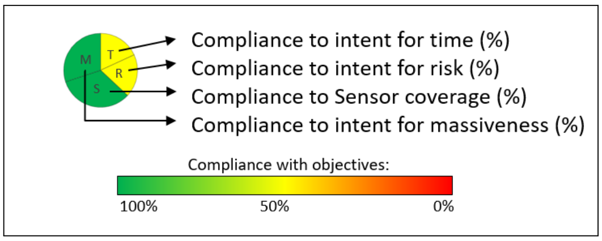


Figure 3.4: Proto-pattern to indicate the compliance of parameters to the user's objectives

Proto-pattern design: Local compliance

The user can click on a task, an endstate or their purpose to see more details about the actual performance and the extent to which it complies with the user's objectives. Section 2.7.3 suggested to use rating stars to indicate the compliance with the objectives. See Figure 3.5.

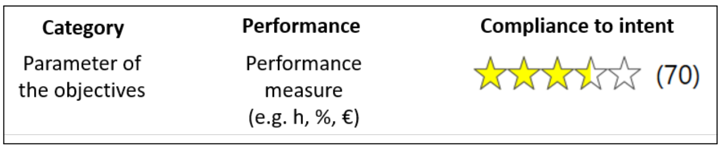


Figure 3.5: Proto-pattern to indicate a solution's compliance with the user's objectives for one parameter

Claims
Providing appropriate information about the compliance of their objectives at the right level of abstraction results in more efficiency, better situational awareness and trust, a decrease of cognitive task load and an increase of appropriate (non-)compliance. The suggested proto-patterns also support salient representations to attract the attention of the user to the most important information (RQ0009).

3.2.5. Visually salient representations

Requirement RQ0009 is defined as follows: *eSupport shall use visually salient representations to attract the attention of the user to the most critical information for the task*. eSupport uses visually salient representations to attract the attention to the most important information. This is the information about the compliance with the user's objectives, used in the proto-pattern defined in Section 3.2.4.

Claims
The use of visually salient representations to attract the attention to the most critical information results in more efficiency, better situational awareness, a decrease of cognitive task load and more appropriate (non-)compliance.

3.2.6. Consequences of decisions

Requirement RQ0004 is defined as follows: *eSupport shall show the consequences of specific decisions or changes of the user (or the environment) at the right level of abstraction. The consequences for the total performance are communicated to the user*. So eSupport should support a quick and intuitive understanding of the consequences of a decision. The user objectives are an intuitive way to project the consequences on. Every time when the user requests a new solution (schedule), the consequences of adjustments (e.g. additional constraints) or changes in the environment are shown.

Proto-pattern design
The rating stars designed for requirement RQ0010 are the basis for the design. This proto-pattern design will be expanded with red stars, to represent a decrease of compliance with the intent. Resulting, more decrease leads to more red stars and thus to a more salient representation. See Figure 3.6 for the design.

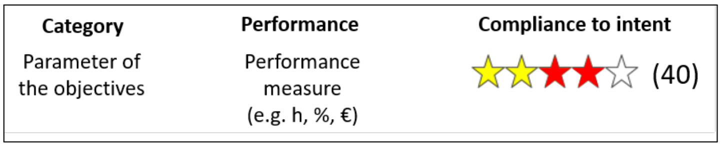


Figure 3.6: Proto-pattern to indicate a decrease of the solution's compliance with the user's objectives for that parameter

Claims
Showing the consequences of decisions results in more efficiency, better situational awareness, a decrease of cognitive task load, an increase of trust and more appropriate (non-)compliance.

3.2.7. Clusters of information

Requirement RQ0006 is defined as follows: *The interface shall organize the information so the information about the tasks, the method of executing the tasks respectively the purpose of the tasks is clustered*. A structure allows the user to find the required information quickly and to focus on the desired level of abstraction. It also allows to link new information with already available information and to show consequences and performance at the right level of the ADS. The structure should be maintained during all phases of the mission.

Proto-pattern design
The information on the screen will be structured according to the abstraction hierachy as defined in Section 2.6.2. The interface is therefore separated in three clusters. The main screen shows the BMS, this shows the locations of the resources, enemy and tasks at the map. The information about the purpose of the tasks is shown on the left side of the screen. The method, that describes how the tasks

are allocated to the resources, is shown at the bottom of the screen in a schedule. See Figure 3.7 for the clusters.

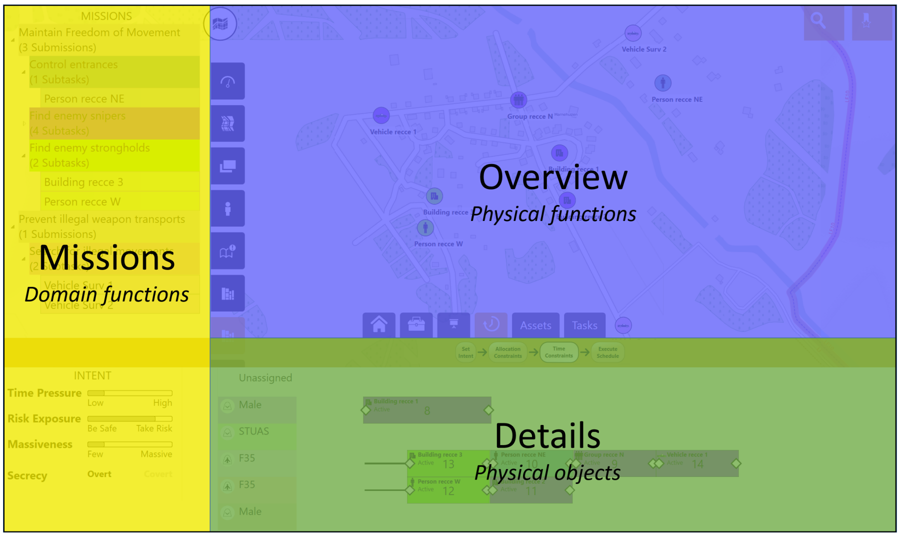


Figure 3.7: Proto-pattern to indicate information clusters for different levels from the abstraction decomposition space

Claims
Showing the information separated in clusters results in more efficiency and better situational awareness.

3.2.8. Relation of information

Requirement RQ0005 is defined as follows: *eSupport shall show the relation between information of physical objects, physical functions (own tasks and the related parent tasks) and domain functions. Other data shall be integrated with this information to avoid loose data*. eSupport should support the user's understanding of the relation between the information of the resources, the high-level tasks, the endstate(s) and the purpose. This also supports RQ0003 to be able to provide information at the appropriate level. The relation between the levels should be clear during all phases of the scheduling process. The relevant information is defined in Section 2.6.2 and includes the purpose of the mission, one or several endstates to achieve the mission's purpose and several high-level tasks to achieve the endstate(s).

Proto-pattern design
eSupport shows the relations between the purpose of the mission and the endstate in the left screen, according to the user interface structure defined in Section 3.2.7. All endstates have a distinctive color, the icons of the tasks on the map, so the overview screen, have the same color as the endstate they contribute to. See Figure 3.8 for an overview of the relations.

Claims
Providing the relation between information leads to better situational awareness and more effectiveness.

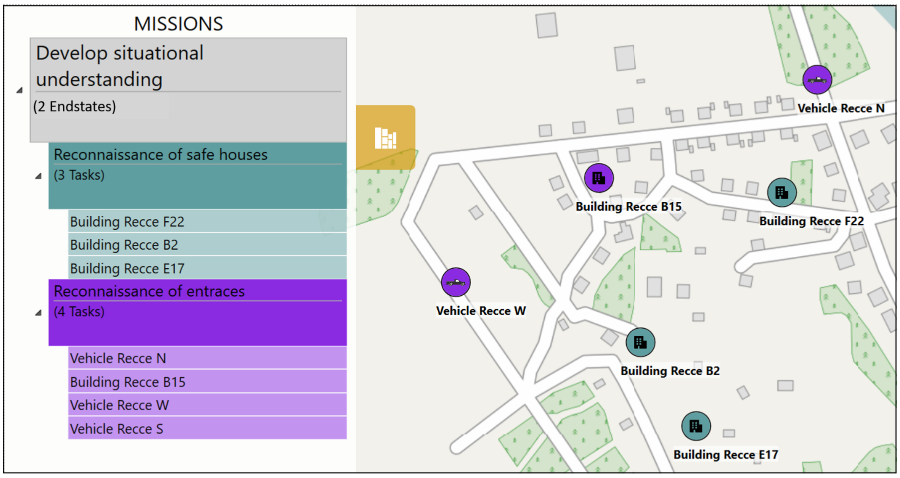


Figure 3.8: Proto-pattern to show the relations between the information about different levels of the abstraction decomposition space

3.2.9. Separated scheduling steps

Requirement RQ0011 is defined as follows: *eSupport shall support the user's understanding of its behavior by separating the scheduling process in several steps. The steps are: 1. set intent, 2. set allocation constraints, 3. set task-order constraints and 4. acknowledge schedule*. Dividing the scheduling steps supports the utilization of the most effective characteristics of the user and makes the process easier to understand. The user is in control and can set any desired constraint that eSupport has to deal with whenever desired.

Proto-pattern design

The scheduling process is divided in three steps: setting the intent, setting the allocation constraints and setting the task-order constraints. The user can go to each phase of the scheduling process whenever desired using the buttons. Going back to reset the intent however will also remove the constraints, since the basis of the schedule is then changed. See Figure 3.9 for the design of the proto-pattern.

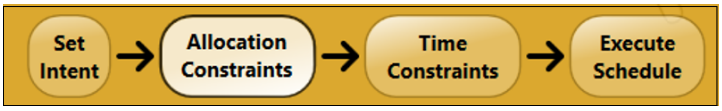


Figure 3.9: Proto-pattern to navigate through the steps of the scheduling process

Claims

Separating the scheduling process in several steps should make the scheduling tasks more simple for the user. This results in a decrease of cognitive task load, more trust and better appropriate (non-)compliance. Furthermore, the design is usable to support decision authority of the user (RQ0008) and to support users to set and change their objectives and constraints whenever desired (RQ0007).

3.2.10. Flexibility for user

Requirement RQ0007 is defined as follows: *eSupport shall compute a schedule and enable the user to set or change the objectives and constraints for the schedule whenever desired by the user*. eSupport shall utilize the user's strengths by enabling them to go to another phase of the scheduling process whenever they desire. The proto-pattern defined in Section 3.2.9 supports this requirement.

Claims

A flexible system that allows users to set their intent and constraints at all times results in more effectiveness, a decrease of cognitive task load, more trust and better appropriate (non-)compliance.

3.2.11. Decision authority

Requirement RQ0008 is defined as follows: *eSupport shall (at least initially) not take decision authority away from the user by asking permission to the user to execute the proposed schedule or by allowing the user to override the schedule*. eSupport shall allow the user to decide if a schedule will be executed. The proto-pattern defined in Section 3.2.9 supports this requirements.

Claims

Providing the user with decision authority over eSupport results in (the user value) decision authority.

3.3. Prototype

The proto-patterns are combined for the design of the prototype of eSupport. This section describes the claims that will be validated in the experiment. Furthermore, the prototype is described in more detail.

3.3.1. Claims for prototype

Claims can be verified through research experiments, in the current study however only the effects of additional transparency are investigated. Requirement RQ0010 covers the demand for additional transparency and two proto-patterns are designed to meet the requirement. Therefore, we shall investigate the effect of a prototype that includes both proto-patterns with a prototype without the proto-patterns. However, since the participants of the experiment have no domain knowledge, the participants would have no reference and cannot reason about the fitness of a solution in relation to their objectives. Therefore, we decided to include the proto-pattern for the local compliance (so the compliance with the user's objectives of the allocation of one task to one resource) in both prototypes.

In the claims, we defined that requirement RQ0010, so additional transparency, results in more efficiency, appropriate (non-)compliance and trust, better situational awareness and a decrease in cognitive task load. Therefore, these five main effects are tested. This corresponds with the focus of the current study. To repeat short, the focus is to investigate how eSupport can provide additional transparency about a proposed schedule's compliance with the user's objectives and thereby improve the performance. The performance is defined in terms of *efficiency*, *appropriate (non-)compliance* and *situational awareness*. However, since balanced *cognitive task load* and calibrated *trust* are important for appropriate (non-)compliance and because the requirement effects *trust* and *cognitive task load*, also these variables are measured. Besides, the *usability* of eSupport is tested to evaluate the design and to check if the minimal requirement for good usability of eSupport is met, since poor usability causes high cognitive task load and a decrease of performance.

Concluding, the claims that will be evaluated with the prototype are: **efficiency**, **situational awareness**, **trust**, **cognitive task load** and **usability**. Other effects are not expected nor intended to change. Section 4.2.3 describes the measurement methods in more detail.

3.3.2. General software description

eSupport is build upon the CommonSense Framework. The framework therefore determined for a great deal the software, the software architecture and interface style of eSupport.

The CommonSense Framework and the prototype of eSupport are built using the Model-View-ViewModel (MVVM) pattern, an extension of the Windows Presentation Foundation. The Views take care of the presentation of the data, so the interface components, and are built using XAML. The Model contains the data and business logic and is implemented using C#. This is for example the scheduling algorithm, the look-up tables with the fixed values for the algorithm and the enemy types of threat. The ViewModels convert the data so the views can easily manage and present it. So the ViewModels support the communication between the views and the models, this is implemented using C#.

The CommonSense Framework contains the plugin COPPR. This plugin is the BMS we build eSupport for and includes the management for tasks, resources and enemy threats. COPPR also contains a first scheduling method, however this functionality needed to be extended for eSupport. eSupport is build as a plugin for COPPR and as such can be turned on or off without affecting other plugins.

eSupport contains several component for the interface, these are the main Views. Each View has a ViewModel to manage the data it shows and to manipulate the interface. The main Views used for the prototype of eSupport are:

- Left View to show the missions hierarchy and user's objectives
- Left View to manipulate the missions (only in experimenter mode)
- Bottom View to show the scheduling process

Other Views were used for smaller components of the interface, we will not describe these Views. An overview of the main components of the software is provided in Figure 3.10.

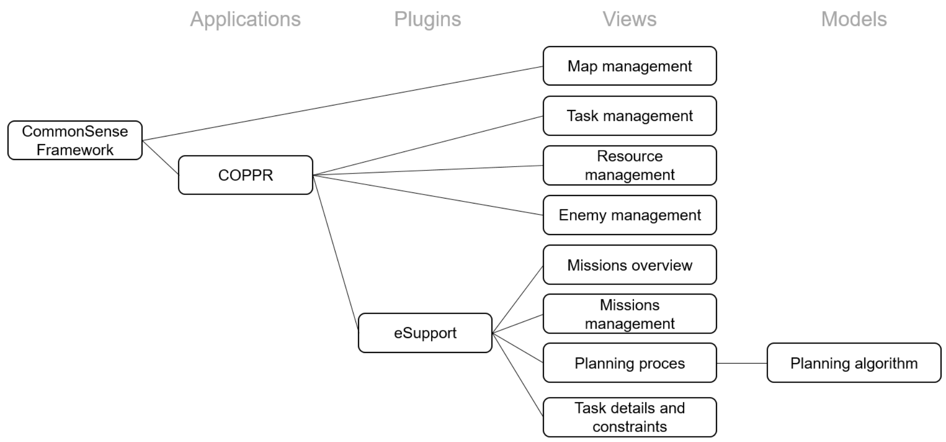


Figure 3.10: Most relevant software components for the prototype of eSupport

3.3.3. Storyboard

The design scenario and use cases describe the activities of the operator. The prototype of eSupport supports these activities. The storyboard in Figure 3.11 shows the interface of the prototype for the different steps of the scheduling process.

3.3.4. Pilot study results

A pilot study was carried out with four participants in order to test for usability issues, to test the understanding of the participant of the instructions and their task and to get more insight in the duration of the experiment. Since we had no funding to execute the experiment, we aimed for an experiment duration of 75 minutes, since that seemed like an acceptable duration for volunteers.

The pilot study showed no usability issues with the prototype of eSupport, however several small improvements have been made to support the ease of use. Three examples of improvements are: 1. The buttons to close a window are now easier to click on to support ease of use, 2. The paths of resources to the tasks, that were provided in the CommonSense Framework, have been removed to reduce the confusion of participants and 3. The colors of the icons for the resources have been removed to make the resources more distinct from the tasks.

The study also showed that the instruction to explain the relevant information, how to use eSupport and the method to check the schedule for mistakes required much improvement. It was important that participants had a clear understanding of these aspects, however, the instruction should also be completed in approximately half an hour. After various improvements, the participant understood their task for the experiment well, however the whole experiment still took two hours.

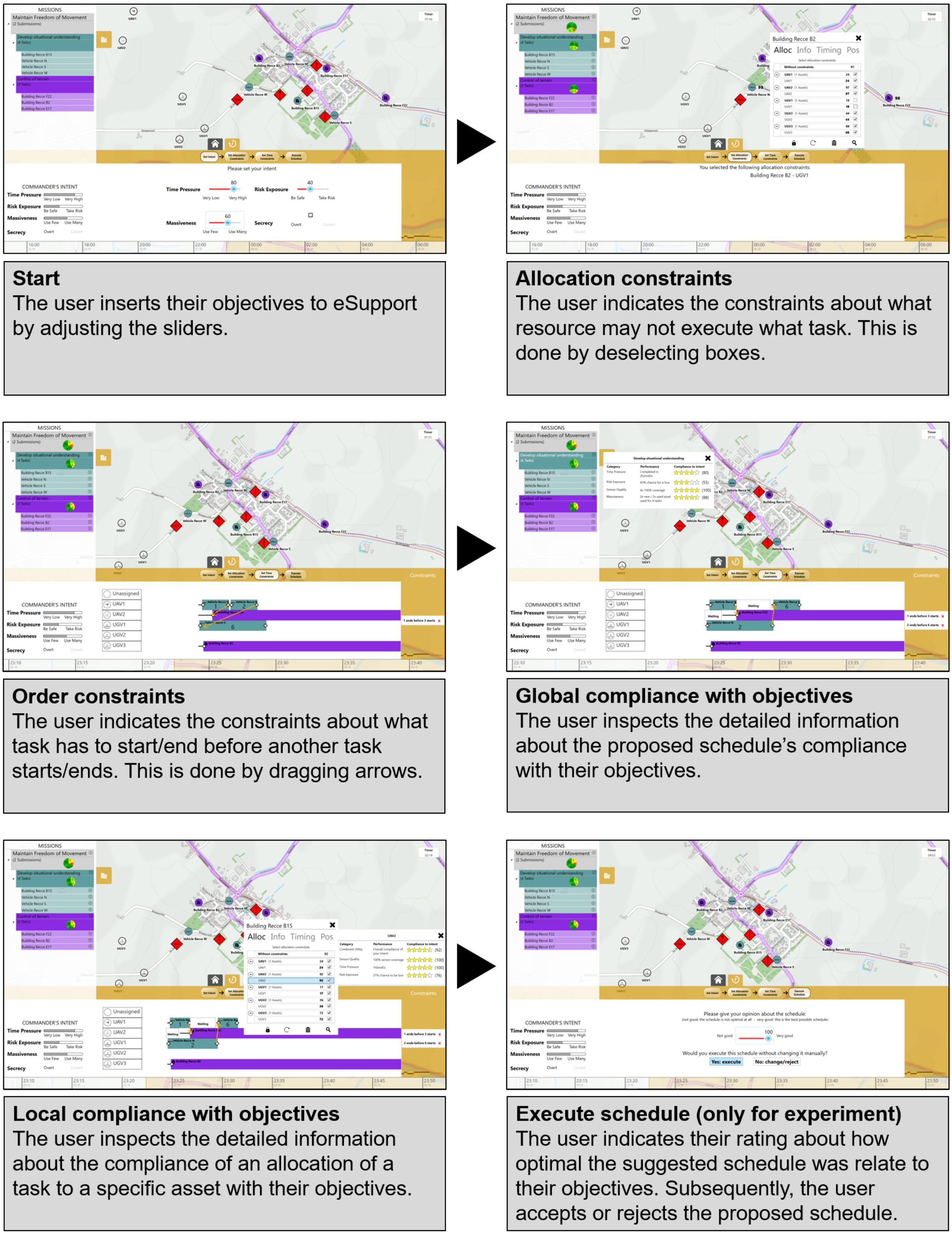


Figure 3.11: The interface of eSupport to support the user's activities

To reduce the required time for an experiment, we took several measures. The first measure was to decrease the number of missions from five to four missions per condition. This required an adjustment of the experimental design, since the questionnaires had to be distributed over four missions per condition and at least one mission still had to present a wrong schedule. The second measure was to restrict the time required for each mission. However, because we also wanted to use the scheduling duration as a measure of performance, we did not set a hard threshold. Instead, we only indicated that each mission should be completed within four minutes, this was told to each participant in the instruction. To support the participants with keeping an eye on the time, we included a small timer in the right top corner in eSupport.

The first measures resulted in a decrease of duration, however not sufficient. So, we had to find more measures. During the pilot study, we saw that participants required a lot of time to understand the constraints during the instruction and to set the constraints during the missions. Therefore, we decided to exclude the order-constraints from the experiment. As a consequence, we will not see when participants try to manipulate a schedule by setting unnecessary order constraints to investigate the effect of a change to the schedule. This effect is called sensitivity analysis ((Cummings et al., 2012)).

With these measures, the minimum duration of a full experiment was 1.5 hours. We refrained from further measures and decided to continue with the experiment.

4
Evaluation

This chapter describes the experiment in more detail to evaluate the effects of additional transparency on users. The introduction in Section 4.1 describes how the effects are investigated using the claims. This is followed by a description of the method of the experiment in Section 4.2. The results of the experiment are presented in Section 4.3.

4.1. Introduction

The purpose of the experiment is to answer the research question described in Section 1.2 by testing several claims, described in Section 3.1.4. The claims are indicated in the research question:

How can a human-automation collaborative system (eSupport), used for scheduling high-level tasks to multiple heterogeneous unmanned vehicles, provide additional transparency about a proposed schedule's compliance with the user's objectives, and thereby improve the user's performance? Hereby we defined performance in terms of **efficiency**, **appropriate (non-)compliance** and **situational awareness**. **Cognitive task load** and **trust** are also measured since these are crucial factors for appropriate (non-)compliance. At last, **usability** is measured to draw conclusions about the design of eSupport and to validate if requirement RQ0002 is met.

It is probable that experienced users (like actual UV operators) pay attention to other things than the participants and would use different indicators to check a schedule for mistakes. Therefore, we will not be able to generalize the results about the performance of participants. We specifically will look at the effect of additional transparency on the participants compared to the condition without additional transparency. The order of providing additional transparency is counterbalanced and included in the statistical analysis.

We hypothesize that additional transparency will improve user performance and trust and will lead to a decrease of CTL compared to the situation without additional transparency. Section 4.2.3 describes how the dependent measures will be measured and the hypotheses about the effect of additional transparency on them. Furthermore, we checked how the order of providing transparency affects the possibly observed effects of additional transparency.

4.2. Method

4.2.1. Participants

The participants were not required to have a military background. Cummings et al. (2010) investigated a system to support operators for scheduling various assets, they used a similar experimental design. The authors concluded that military experience or the amount of years of service did not make a difference for the performance or workload (Cummings et al., 2010; Clare et al., 2012b).

Based on the current requirements for military functions that show similarity with the UV operators (e.g. fighter pilots, air controllers and officers of technical disciplines such as the engineers), we assumed that future UV operators at least require a bachelor degree including mathematics. Therefore, we assume that a sample group only containing students, former students and academic staff is representative for the future population of UV operators.

Participants are a convenience sample of 24 adults and included 16 men and 8 women. The sample included 22 European, 1 Asian and 1 Middle eastern participants. All participants were students, former students or employees of the TU Delft, the sample included 8 bachelor, 10 graduate, 2 graduated and 4 PhD students. 2 participants had a military background, each had 2 years of military experience.

4.2.2. Apparatus

The experiment required two monitors, a Dell 24-in. LED color monitor was used to show the missions and required a mouse for interaction. A Philips 17-in. LED color monitor was used to show the instructions and required keyboard for interaction. See Figure 4.1 for an overview. The experiment required the CommonSense Framework of TNO with the additional eSupport application for the scheduling task. The instructions were shown using Microsoft PowerPoint. An Hewlett Packard Spectre 13-3000 notebook was used to run the CommonSense Framework with eSupport applications and a Dell Latitude E6530 notebook to run PowerPoint. An additional A4-paper (cheat-)sheet (see I) provided the participants with basic information, which was also shown during the instruction. The participants were allowed to use this sheet during the missions.

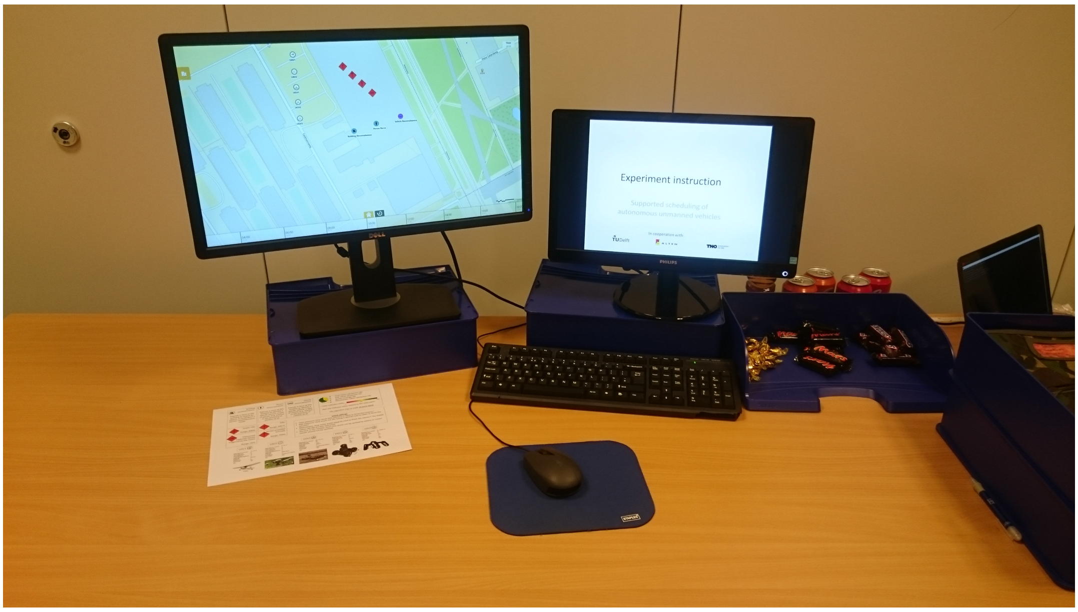


Figure 4.1: The experimental setup

4.2.3. Dependent measures

There are five effects described in the claims that we want to measure. The methods to assess these measures are described below.

Trust

When a user has better calibrated trust, it typically results in increased performance of the participant in several observable ways. The level of trust of the participants is measured during the experiments and compared to the other performance measures. The level of trust is however no indicator for calibrated trust, to verify if trust was calibrated better we have to look to the relation between trust and performance.

Trust is measured using a questionnaire that assesses the overall trust level of the participant in the system. The questionnaire for trust between people and automation of (Jian et al., 2000) is used to measure the user's self assessed trust. Since a number of questions are not applicable to the current experiment, we use a modified version of the questionnaire. The questionnaire is provided in Section I.1.3. Although trust can be biased through earlier experiences with the system, we measured trust for each participant in both mission blocks to check for within-subject effect due to transparency.
Hypothesis: the additional transparency of eSupport results in higher trust in eSupport. We expect that higher trust also results in better user performance, so we expect more appropriate (non-)compliance, fewer unnecessary adjustments and less scheduling time.

Situational awareness
The situational awareness of participants is measured with two questionnaires per mission block. The questionnaire is constructed based on the SAGAT (see Section 2.6.6). The assessment method normally requires to freeze the simulation, however in this case the evaluation can be done directly after the acceptance or rejection of the schedule, because the missions are ended abruptly once the participants accepted or rejected a suggested schedule. Each questionnaire contains 8 questions picked from Section I.1.4, the questions are specified for the mission at hand. The number of correct answers is compared between both conditions per participant. The SA scores are compared within-subjects to evaluate the effect of the use of additional transparency on SA. *Hypothesis:* Additional transparency does not effect SA. Because the additional information should increase the participants understanding of the reasoning behind the schedule, the additional transparency information also should require fewer manual checks. When a user requires fewer manual checks, we expect that they have fewer knowledge about the elements in the environment (SA level 1). However, the schedule's compliance with the user's objectives is presented to the user in the transparency condition, so we expect then that users will have more insight in the effect of the schedule on the user's objectives (SA level 2 and possibly 3). However, the questionnaire does not measure SA at level 1 and at level 2 and 3 separately.

Cognitive task load
The cognitive task load of the operator is measured using the NASA-TLX (see Section 2.6.6). To save valuable time, the additional questionnaire to allow the participants to indicate the relevance of factors for the CTL is not included, since we are particularly interested in the difference of CTL between the condition with and without additional transparency. To allow us to compare the CTL within subjects, the questionnaire is provided for the second mission for each mission block. This method does not indicate if the CTL was balanced, however indicates an increase or decrease due to the transparency information.

A second measure is introduced to give an indication about the actual cognitive task load during the execution of the missions. The measure uses the message response time (MRT) to a simple question (which does not require SA) that pops up on the screen during the second and third mission of each mission block. All participants are instructed to answer these questions as soon as possible. Very short respectively long response times could indicate that task load is too low respectively too high and thus is not balanced. The questions that will be asked are:

- *Commander: Did you already start with scheduling?*
- *Commander: Communications check, do you receive this message?*

Hypothesis: additional transparency initially results in more cognitive task load due to the extra information, but will overall result in a decrease of cognitive task load because the participants have to reason less about the proposed solution. However, since there is no actual deployment of assets during this phase, we expect relatively low scores for CTL and thus short response times.

Appropriate (non-)compliance
Appropriate (non-)compliance is one of the measures for the user's performance. The following indicators are used:

1. *Appropriate compliance*: The ratio the participant accepts a correct schedule suggested by eSupport [Right or Wrong];

2. *Appropriate non-compliance*: The ratio the participant rejects an incorrect schedule suggested by eSupport [Right or Wrong];
3. *Appropriate rating*: The rating of the participant for a correct or incorrect schedule. Ratings are between 0 and 100 and are provided at the end of a mission. Correct and incorrect schedules require high respectively low ratings, the measure is inverted for incorrect schedules [0-1].

Efficiency
Efficiency, composed of *required time* and *effort* is the other measure for the user's performance. The required time for scheduling is measured during the second, third and fourth mission of each mission block. The time starts after the participant read the mission description and ends once the participant accepted (compliance) or rejected (non-compliance) the suggested schedule. The indicator for *effort* is the amount of times a new schedule and detailed information is requested from eSupport.
Hypothesis: the additional transparency of eSupports results in better efficiency. In the condition with transparency information we expect less required time for scheduling. Besides we expect a decrease in effort. Since the extra information at first might lead to extra effort, we expect that participant's would employ sensitivity analysis (Cummings et al., 2012), so manipulate the constraints to see the effects on the schedule, in the conditions without transparency to gain more understanding of the system's reasoning.

4.2.4. Other measures

Other measures were used for further analysis, these are a demographic survey (see Section I.1.1), attentional control (see Section 2.6.6) and usability, see below.

Usability
The perceived usability of the operators is measured with the short usability questionnaire (see Section 2.6.6). We added questions to the questionnaire to evaluate the usefulness of specific interface elements, see Section I.1.6 for the full questionnaire. The participants fill out the questionnaire after all mission blocks are completed. The usability of the transparency information is also specifically asked for. The degree in which a user used the transparency information (both the piechart and detailed information) compared to the other interface elements is referred to as the *usage of transparency information*. The rating of the usefulness of the transparency information is referred to as the *usefulness of transparency information*.

4.2.5. Procedure

The participants were tested individually in the New Delft Experience Lab. At the start of the experiment they received a description of the experiment and were asked for their informed consent. Besides, they were informed about the chance to win €50,- if they were among the 5 participants who performed best during the missions. After they signed their informed consent they filled out a demographic survey and the questionnaire about attentional control.

After they filled out these questionnaires, the participants received an instruction how to work with the CommonSense Framework and the eSupport application and were instructed about their task to check the schedules for mistakes. All participants were besides instructed to try to finish each mission within 4 minutes, this was necessary to limit the total experiment duration. The participants were also informed that eSupport can make mistakes in scheduling. The instruction was finished with a proficiency test to ensure the participants had sufficient understanding of the system and their task.

At the completion of the instruction and proficiency test, the participants were asked to execute 2 mission blocks of 4 missions each. One of two mission blocks provided additional transparency information. For each mission, the participants received a written partial operations order about the situation, the commander's intent and the restrictions, which they had to read before starting each mission. Before the first mission they also received a more general description of the military situation and environment. All separate missions take place within this general context. During each mission, the participants had to indicate the commander's intent to eSupport and insert the given restrictions. These restrictions only consisted of the constraints that a certain UV shall not execute a certain task. The constraints about the order of tasks was excluded from the experiment, since it would have taken too much time for the instruction and during the missions. Thereafter, the participants had to check if the schedule suggested by eSupport complied to their intent most optimally. They finished each mission by indicating if they would execute the schedule as such, or if they would have liked to adjust it manually. Besides, they gave a score to each schedule about how optimal it was.

During the first mission of each mission block, the participants were allowed to ask questions to the supervisor. After the second mission of each mission block, the participants were asked to fill out the questionnaire about SA and CTL in that specific order. eSupport provided an erroneous schedule during the third mission of each mission block in which two tasks were assigned badly. After the fourth mission of each mission block the participants filled out the questionnaires about SA and trust. At last, after the completion of both mission blocks, the participants were asked to fill out the usability questionnaire.

4.2.6. Experimental Design

The experiment employs a mixed-factorial design with the transparency condition as independent within-subjects variable and the order of providing the additional transparency as independent between-subjects variable. Each participant executes two mission blocks of four missions each. In one of two mission blocks, additional transparency was provided, the block that provided transparency was counterbalanced.

Besides, to prevent undesired effects of unforeseen differences in the two sets of missions, these were also counterbalanced. Each mission within a mission block was paired with the equivalent mission in the other mission block and contained similar and the same amount of tasks and assets. Since there were 2x2 conditions to counterbalance, the participants were assigned to one of four conditions based on their participant number, see Figure 4.2. To minimize the side effects of participants who had to get used to a new condition and switching between conditions, the performance of the first mission of each mission block was not included in the results.

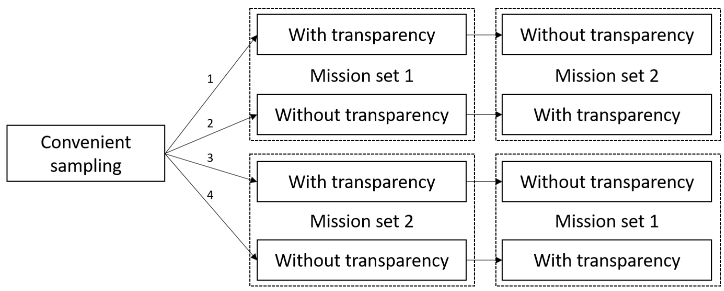


Figure 4.2: Assignment of participants to experimental conditions

Scenario’s

For each mission, the participants received a written scenario (partial mission order) that describes the current situation of the mission. Each mission order describes the mission and intent of the CAT-commander and several restrictions for the current mission. All scenario’s take place in a more broad scenario (the General Crisis State and Special Crisis State), these were provided to the participants prior to the experiment. The scenario’s are included in Section G.1.

4.3. Results

The participants were very positive about the experiment and rated eSupport as useful, professional and promising. This section describes the observed results of the experiment. The next paragraphs describe the reliability of the questionnaires and the compliance with the assumptions for normality. Subsequently, the outliers are described and the results per measure are analyzed. Using G*Power, with a sample size of 24 participants we should be able to find medium effects between dependent measures and large effects between independent measures. All effects are reported as significant at $p < .05$.

4.3.1. Reliability of measures

Five different questionnaires were used to assess several dependent measures. Three of these questionnaires (Trust, SA and Usability) were created or adjusted specifically for the current study, these will be checked for reliability. Besides, during the experiments participants asked questions about particular questions of different questionnaires and indicated that it was not clear what they should answer. This also included the CTL questionnaire. Since these questions might not have been clear for other participants as well, a reliability analysis of these four questionnaires is done. Besides, because several participants had comments about the schedule for some missions, the measures for correct (non-)compliance are checked.

CTL questionnaire
Several participants raised questions about question 2 about *physical demand* and question 4 about *performance*. Since the participant's job was not physical at all, question 2 was generally rated low and considered irrelevant. The additional questionnaire to indicate the importance of a factor for CTL, that was not included (see Section 2.6.6), typically would have eliminated question 2 in this case. Question 4 about the participant's performance sometimes led to confusion, since it was not always clear for the participants if the question had to be answered related to their performance for the actual task, or related to filling out the SA questionnaire. We conclude that the performance question was too ambiguous and might have not been a good indicator of CTL.

Because of the confusion and irrelevance, we performed a reliability test. The test showed that, when all questions were included for CTL, Cronbach's α = .66. The confusion about question 4 had relatively low reliability, Cronbach's α = .187, as expected. Question 2 also had low reliability, Cronbach's α = -.115.

After exclusion of question 2 and 4 from the results the reliability considerably increased, Cronbach's α = .778. For the further analysis of results, question 2 and 4 are excluded from the measure for CTL. This has no further consequences, since the original questionnaire can also exclude variables based on the additional questionnaire.

Message response time
There were two measures for *CTL*, the questionnaire and the response time to messages. Results showed that the two indicators were not correlated, τ = 0.120, BCa 95% CI [-0.109, 0.325], p = .296. Transformation of the message response times using the Log-function, so normality can be assumed, did not lead to better results, r = .134, BCa 95% CI [-0.207, 0.422], p = .417. Furthermore, *MRT* was not correlated with other measures except with *SA*. Participants with higher *SA* typically had lower *MRT*. During the experiments, we noticed that participants often did not pay attention to a message, even though the message was clearly visible and the participants were instructed to respond as soon as possible. From these observations, we conclude that the messages should be more salient to drag the attention of the participant. Because there is no correlation with the questionnaire for *CTL*, we cannot be sure what construct we measured with *MRT*. For example, we might have measured the focus or concentration of the participants, or the memory capacity considering the correlation with *SA*, and in Section 2.6.6 we concluded that spatial ability might influence the user's response time (Mercado et al., 2016). Therefore, we exclude the measure of *MRT* from the results.

Trust questionnaire
Various participants indicated that question 2, *The execution of the schedule will have a harmful or injurious outcome for the environment*, was more related to their attitude towards the use of UVs for military purposes than about the system. Since this question is perceived differently by multiple participants, this question will be excluded from the results. To check the questionnaire for comparable differences in perception, a reliability test for the questionnaire without question 2 was performed, Cronbach's α = .686. The test statistics showed that question 6, which measures familiarity of the user with eSupport, was unrelated to the other questions, all inter-item correlations are under α = .270.

In this specific situation, we hypothesize that familiarity is perceived by the participants as how well one knows how to interact with the interface of eSupport and to find the required information. This is not directly related to the participant's trust in the performance of eSupport. The answers to question 6 show a significant increase over time to familiarity, $t(23) = -3.50$, $p = .002$, between mission block 1 ($M = 4.46$, $SE=1.59$) and mission block 2 ($M=5.08$, $SE=1.35$). This effect is not seen for the answers to the other questions. Based on the questionable relation of familiarity with trust in this situation and the lack of correlation between question 6 with the other questions, this question is excluded from the measure for trust. The reliability of the questionnaire for trust without question 2 and 6 increased to Cronbach's $\alpha = .794$.

Usability questionnaire
There were no comments of participants about this questionnaire. The reliability test of the questionnaire showed that Cronbach's $\alpha = .792$. The results of the questionnaire are used unchanged.

Appropriate (non-)compliance
The comments of participants about mistakes in schedules were reason to check the schedules that eSupport proposed for each mission. From this, we had to conclude that the schedules for mission 4 and 8 were slightly suboptimal (a very low increase of performance for one task was achievable) while these had to be optimal. This is a consequence of using a greedy algorithm. Although this was only noticed by a few participants, we chose to exclude the measures for (non-)compliance for these missions since we cannot be sure what participants have seen these minor mistakes. However, we do not exclude the rating for trust, since the participants knew there could be mistakes in the schedules and the additional unexpected mistakes occurred in both conditions.

The measures for appropriate (non-)compliance in mission 2, 3, 6 and 7 will be used. However, this measure has only three outcomes per transparency-condition and as expected is not normally distributed. Therefore this data will be treated ordinal and analyzed with the Wilcoxon matched-pairs signed-rank test.

4.3.2. Assumptions for linearity of data

We checked the data for the assumptions of homogeneity of variance, sphericity and normality to allow the use of parametric tests. Unless otherwise stated, all tests met the assumptions of homogeneity of variance, using Levene's test, and Sphericity, using Mauchly's test. The distribution of data of the measures variables is investigated in the following sections, the Kolmogorov-Smirnov test, Q-Q plots and histograms were used to check for the assumption of normality.

Situational Awareness
The SA of participants is measured using the amount of correct answers out of 16 questions per transparency-condition. Normality was checked for the mission block with transparency, $D(24) = 0.130$, $p > .200$, and without transparency, $D(23) = 0.122$, $p > .200$. The values were also checked with histograms and Q-Q plots, from the analysis we can assume the values do not deviate significantly from normal.

Message response time
The message response time has a large positive skew for the condition with and without transparency information, this can be expected for response time data (Whelan, 2008). Use of a logarithmic transformation to make data normal is normal for single-sides distributions, like times starting from 0. Also in this case, a logarithmic transformation returned normally distributed scores for both conditions. The further analysis uses the logarithmic transformation of the message response times.

The results for appropriate (non-)compliance are not normally distributed nor can be described with a Poisson distribution due to it's large under dispersion, this is expected due to the very low number of possible outcome values (0, 1 or 2). Since a comparison of means would not make sense, the scores are analyzed using non-parametric tests. Since the scores have a large number of tied ranks, more specifically most scores have rank 2, Kendall's τ is used.

Other variables
The values for SA, CTL, trust, scheduling time meet the assumptions for normality, this also holds between transparency conditions. Also the values for AC did not deviate significantly from normal. Therefore, these variables can be analyzed using linear regression.

Usability has a positive skew, therefore we used a Log-transformation of the reversed data and rescaled the data to the interval [0,1]. After transformation, the Q-Q plot, histogram and the Kolmogorov-Smirnov test ($D(24) = 0.113$, $p > .200$) showed that the values did not deviate significantly from normal. For further analysis of usability, the log-transformation of the values is used.

The values of usefulness and (reversed) usage of transparency information show a positive skew and contain high rankings of the highest value. Therefore, no normal distribution for these variables can be assumed. Kendall's τ is used for the analysis of correlations because of the high rankings of certain values. The values for appropriate (non-)compliance each have 2 possible values (correct or wrong) and will not be treated as normally distributed. The ratings for compliance in the transparency and non-transparency condition have positive skew, we found however no transformation that results in a distribution to assume normality for both conditions, so the ratings will be treated non-linear. The scores for VGE, attitude towards UVs and attitude towards military UVs (and obviously gender) did not pass the assumptions for normality. Closer inspection of the values for attitude towards UVs using a histogram and a Q-Q plot show that there are no high ranks for the values 1 and 5, and the values look normally distributed. Therefore, we assume normality for attitude towards UVs and thus will use linear regression for further analysis. The values for VGE show two modi and high ranks for the value 1, VGE will therefore not be treated as linear model. Since the scores contain high ranks of the value 5, Kendall's τ is used for further analysis about the correlations.

4.3.3. Exclusion of data

One of the participants was confronted after the first mission block that the experiment would take longer than average since the instruction took a lot more time than average. The participant indicated to have an appointment planned after the experiment, and thus was in a rush during the second mission block. The results showed that the second mission block was executed significantly faster than the first and the rating for cognitive task load was significantly higher. Because of the external time pressure and the relation with the outcome, the results of the participant for the second mission block have been excluded from the data set.

After the first mission block, a participant indicated that it was not clear what asset would execute what task. Although this was explained and asked for during the instruction, this might have been forgotten due to an information overload. After an additional explanation directly after the first mission block, the participant indicated that everything made sense now. Since this understanding was crucial for a proper execution of the task, the participant's data for the first mission block is excluded from the data set.

In Section 4.3.1 we already indicated that the results considering (non-)compliance for mission 4 and 8 will be excluded from the analysis. Participants were allowed to ask questions during the first mission of each mission block. Some participants didn't require help, however others asked more suggestions about their task. Since the first mission of each block was considered a trainings mission, all measures for mission 1 and 5 will be excluded from the analysis.

The attitude towards military UVs was not on the demographic survey for the first four participants. This makes it less convenient to investigate interaction effects with this variable, since that would eliminate 4 or 8 data values per test. Therefore, we often did not include this variable in the analysis.

4.3.4. Outliers

All dependent variables were checked for outliers using boxplots, however most of the identified outliers were not excluded since these values were not extraordinary. However, a comparison between the increase of each dependent variable due to the additional transparency showed a clear outlier for the increase of trust. The analysis was performed using scatterplots between the increased values (see

Figure 4.3). The scatter plots show that one value for trust was always isolated from the clusters of values. Closer inspection of the value showed that the participant's trust decreased significantly in the second mission block. This participant indicated with a comment on the usability questionnaire that their trust decreased, since the system did not reschedule an asset on a task during mission 6. The inappropriate noncompliance (so rejecting a correct suggestion) in the second mission block indicates that the participant made a mistake. The participant's reaction on this event was very strong, and the reaction has a strong impact on the linear relation between variables. Therefore this value is excluded from the data set.

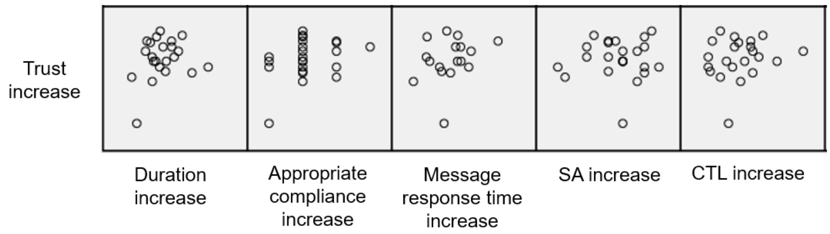


Figure 4.3: Comparison of the influence of additional transparency on various dependent variables per participant

4.3.5. Overview of results

An overview of the mean and standard deviation of *SA*, *CTL*, *trust*, *(non-)compliance*, *duration* and *MRT* for each mission block and transparency condition is provided in **Table 4.1**. The mean, median and standard deviation of the between-subjects variables are provided below in **Table 4.2**.

Table 4.1: Overview of means and standard deviations of variables per mission block and transparency condition

	First mission block				Second mission block				Total		Total		Total		Total	
	With transparency		Without transparency		With transparency		Without transparency		With transparency		Without transparency		First mission block		Second mission block	
Var	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SA	8,58	3,18	7,55	1,75	9,33	3,39	10,08	2,97	8,96	3,28	8,82	2,36	8,07	2,46	9,71	3,18
CTL	0,45	0,19	0,46	0,16	0,46	0,15	0,46	0,19	0,46	0,17	0,46	0,18	0,46	0,18	0,46	0,17
Trust	0,67	0,11	0,71	0,11	0,72	0,12	0,69	0,09	0,70	0,12	0,70	0,10	0,69	0,11	0,71	0,11
Approp compl.	0,83	0,39	0,82	0,4	0,83	0,67	0,75	0,75	0,83	0,53	0,79	0,58	0,83	0,40	0,79	0,71
Approp. non-comp.	0,83	0,39	0,64	0,5	0,39	0,49	0,45	0,45	0,61	0,44	0,55	0,48	0,74	0,45	0,42	0,47
Score	0,7	0,1	0,67	0,21	0,68	0,21	0,61	0,17	0,69	0,16	0,64	0,19	0,69	0,16	0,65	0,19
Duration	3,89	0,9	3,79	0,2	3,67	0,92	4,03	0,78	3,78	0,91	3,91	0,49	3,84	0,55	3,85	0,85
MRT	0,3	0,31	0,2	0,12	0,17	0,23	0,25	0,29	0,24	0,27	0,23	0,21	0,25	0,22	0,21	0,26

Table 4.2: Overview of means, medians and standard deviations of between-subject variables

	Mean	Median	SD
Usability	0.82	0.85	0.10
VGE	3.17	3.50	1.63
Attitude towards UVs	3.79	4.00	0.78
AC	2.82	2.86	0.31

4.3.6. Correlations

The correlations between variables are investigated to get more understanding of the relations between variables. The correlations and a schematic overview of the correlations are provided in Appendix J.

Note that we cannot draw conclusion about causality from the correlations.

The correlations show that, although the participants were randomly assigned to conditions, the first mission set (missions 1-4) contained significantly more participants with a positive attitude towards the use of UVs, $\tau = -.41$, BCa 95% CI [-.714, 0], $p = .038$. There was however no other effect of the mission set.

There were also correlations within the demographic data. The correlations show that males had more video gaming experience (VGE) than females within the sample group and had a more positive attitude towards the use of UVs (in general and in military operations). VGE and attitude towards UVs were also positively correlated.

4.3.7. Effect of additional transparency

This section investigates the mixed-factorial effects of the transparency condition as independent within-subjects variable and the order as independent between-subjects variable on *Performance (scheduling time, appropriate (non-)compliance and correct compliance rating)*, *SA*, *trust* and *CTL*.

Performance: scheduling time and compliance

Figure 4.4 shows the effects of transparency and the order of the condition with and without transparency on *duration*, *correct compliance rating*, *appropriate compliance* and *appropriate non-compliance*.

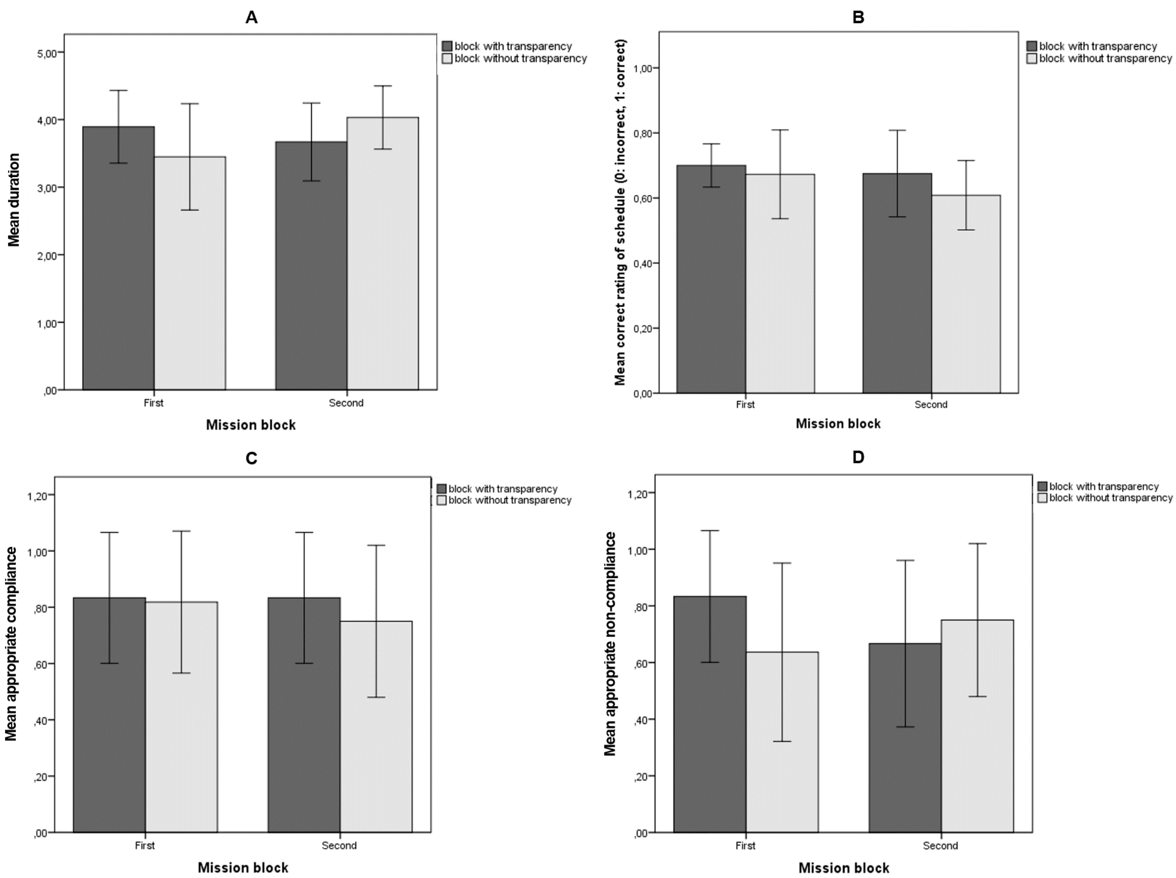


Figure 4.4: The mean values of performance per transparency condition and per mission block with 95% confidence interval for A. the scheduling time, B. the correct rating of the schedule, C. the amount of appropriate compliance and D. the amount of appropriate non-compliance

Chart A shows the effect of the transparency condition and the order of the condition with and without transparency on the average duration of the mission. The figure suggests that both groups, so with transparency in the first versus the second mission block, require on average slightly more time for scheduling in the second mission block. However, further analysis showed that there was no significant main effect of additional transparency on scheduling time, $F(1,20) = 0.441$, $p = .514$, nor a significant main effect of the order of the condition with and without transparency, $F(1,20) = 0.595$, $p = .449$, nor an interaction effect between these factors, $F(1,20) = .003$, $p = .960$.

Chart B shows the effect of the transparency condition and the order of the condition with and without transparency on the correct rating of the schedule. The chart suggests there was a slightly better rating in the transparency condition and that rating decreased over time. However, further analysis showed there was no significant main effect of additional transparency on correct rating of compliance, $F(1,21) = 1.64$, $p = .214$, nor a significant main effect of the order of the condition with and without transparency, $F(1,21) = 0.215$, $p = .648$, nor an interaction effect between these factors, $F(1,21) = 0.735$, $p = .401$.

Chart C and D show that, between mission blocks, the average appropriate (non-) compliance was slightly better in the condition with transparency. Using Wilcoxon matched pairs-signed ranks test, no significant main effect of additional transparency was found on appropriate compliance, $T = 7$, $z = -0.82$, $p = .414$, $r = -0.120$, nor on appropriate non-compliance, $T = 12$, $z = -0.38$, $p = .705$, $r = -.06$. Also no significant main effect of the order of the condition with and without transparency was found on correct compliance to a schedule, $T = 10.5$, $z = 0$, $p = 1$, $r = 0$, nor on correct non-compliance to a schedule, $T = 18$, $z = 0$, $p = 1$, $r = 0$.

Figure 4.5 provides more insight in the effect of the additional transparency on the average duration of the missions and the appropriate (non-)compliance.

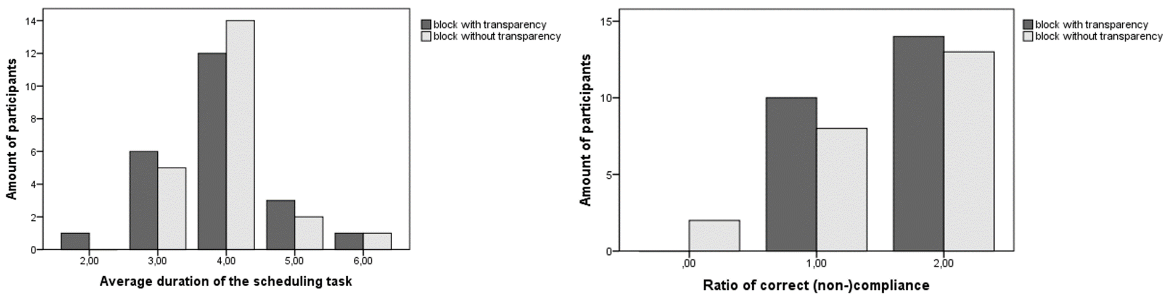


Figure 4.5: Average duration and appropriate (non-)compliance of participants per transparency condition

Situational awareness

The bar chart in Figure 4.6 shows the effects of transparency and the order of the condition with and without transparency on *SA*.

The scatter plot in the left chart suggests that there is no correlation of the transparency condition on SA. The bar chart on the right side of the figure shows how the SA of the participants developed. Participants with the transparency condition in the first mission block had better SA than those in the other condition. In the second mission block however, these participants (so those without transparency in the second block) still had better SA than the other group.

Further analysis showed there was no significant main effect of additional transparency on SA, $F(1,21) = 0.413$, $p = .527$, nor a significant main effect of the order of the condition with and without transparency, $F(1,21) = 0.416$, $p = .526$. However, there was a significant interaction effect between the factors, $F(1,21) = 8.270$, $p = .009$. This is supported by the bar chart, that shows the effect that both groups got better SA in the second mission block, so there was a significant order effect on SA. This is supported by the comments of participants, who indicated that they understood better what they had to focus on after several missions.

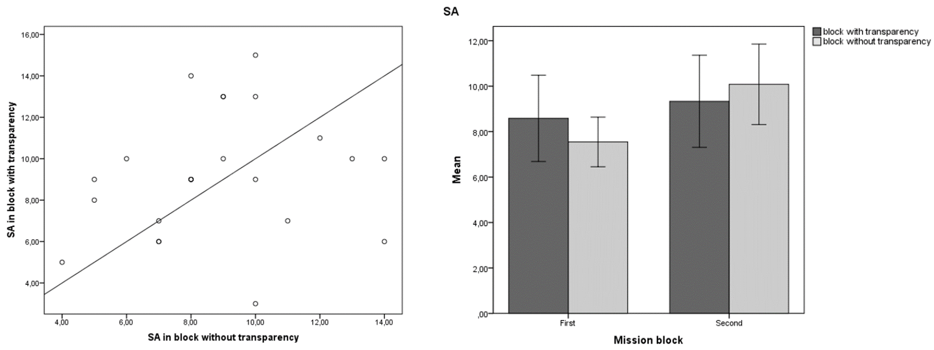


Figure 4.6: Left: values of SA in condition with and without additional transparency, Right: SA in the transparent and non-transparent condition, separated per mission block, with the 95% confidence interval

Trust

The bar chart in Figure 4.7 shows the effects of transparency and the order of the condition with and without transparency on *trust*. The bar chart on the right side of the figure suggests that trust was slightly lower or decreased in the condition with transparency compared to the condition without transparency. Analysis of the results however showed no significant main effect of additional transparency on *trust*, $F(1,20) = 0.25$, $p = .621$, nor a significant main effect of the order of the condition with and without transparency, $F(1,20) = 0.42$, $p = .523$, nor an interaction effect of these factors, $F(1,20) = 1.10$, $p = .306$.

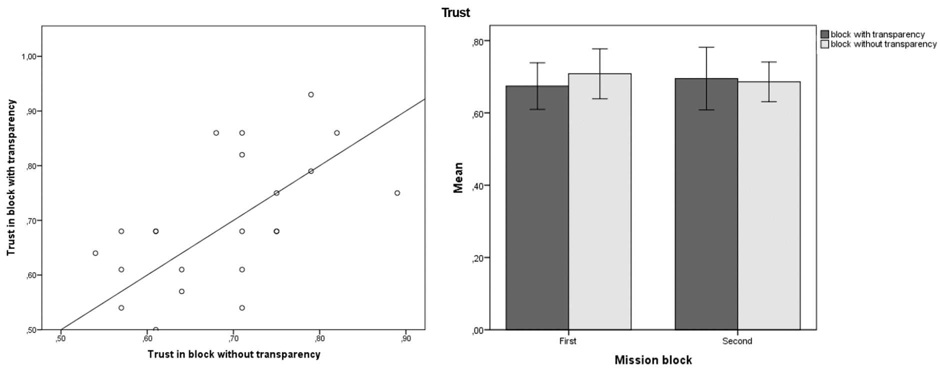


Figure 4.7: Left: Effect of the transparency-condition on trust with the reference line y=x. Right: trust in the transparent and non-transparent condition, separated per mission block, with the 95% confidence interval.

CTL

Figure 4.8 shows the effects of transparency and the order of the condition with and without transparency on *CTL*. The right chart suggests that providing transparency information and the mission block that provided the transparency had no interaction effect on CTL. However, further analysis showed no significant main effect of additional transparency on *CTL*, $F(1,20) = 0.42$, $p = .524$. There was also no significant main effect of the order of the condition with and without transparency, $F(1,20) = 0.89$, $p = .889$, nor an interaction effect of these factors, $F(1,20) = 0.20$, $p = .656$.

Conclusions

No main effects of additional transparency on the dependent variables were found. Table 4.1 shows an increase in performance for all variables in the transparency condition (and equal values for *trust* and *CTL*), however these effects were small. From this, we conclude that additional transparency had no medium or large effects on *scheduling time*, *appropriate (non-)compliance*, *SA*, *trust* and *CTL*. Controlling for *attitude towards UVs*, *gender* and *VGE* also did not show significant effects of additional

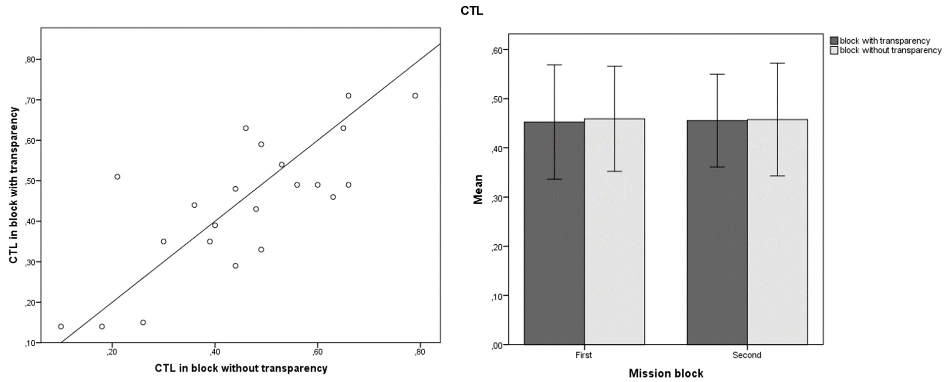


Figure 4.8: Left: Effect of the transparency-condition on CTL with the reference line y=x. Right: CTL in the transparent and non-transparent condition, separated per mission block, with the 95% confidence interval.

transparency on the dependent variables.

For better understanding of the lack of effects of the additional transparency information, the next section investigates the participant's rating of the usefulness and usage of the transparency information. Other observed effects are investigated in the subsequent sections.

4.3.8. Usefulness and usage of transparency information

This section goes into detail about the factors of influence on the perceived *usability and usefulness of the transparency information*. The correlations in Appendix J showed that the *mission block that showed transparency* was significantly correlated with the *usage of the transparency information*. The bar chart in Figure 4.9 gives more insight in this effect.

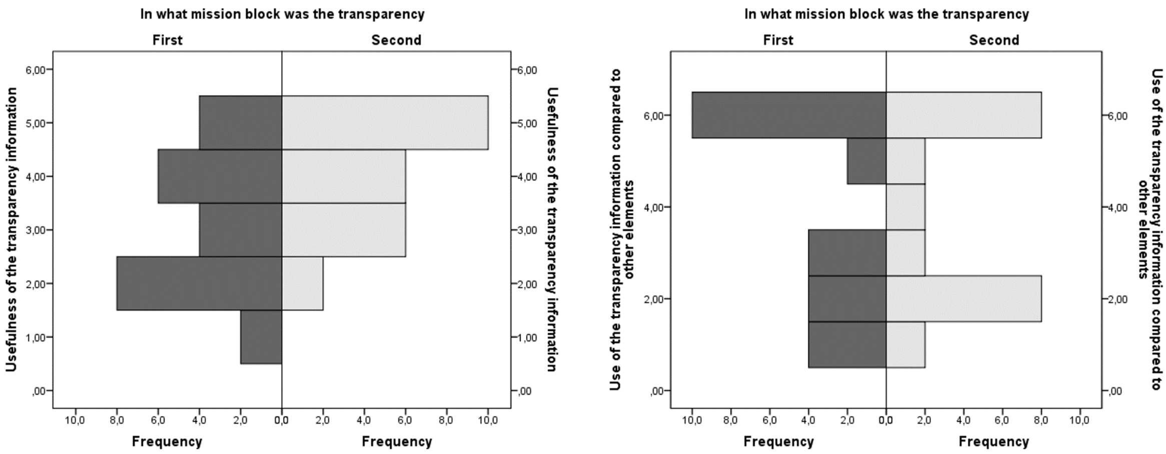


Figure 4.9: The relation between the usefulness and usage of the transparency information with the mission block that provided the transparency information

The figure shows that the *usefulness of the transparency* was rated higher when the transparency was provided in the second mission block (M = 4.0, SD = 1.04) than in the first mission block (M = 3.08, SD = 1.31). An explanation could be that in this case the participants were exposed to the transparency information just before filling out the usability questionnaire.

Another explanation is that participants knew better how to use the transparency information once they were more experienced. Because participants indicated during the experiment that they got more comfortable with their task after several missions, this implies that they were more capable to use relevant available information. The effect of the mission block that showed the transparency information on the *usefulness of the transparency* was however non-significant, U = 101.5, z = 1.754, p = .089, r =

0,36. The figure also shows that participants who received transparency information in the first mission block made more use of it (M = 3.92, SD = 2.11) than in the second mission block (M = 3.75, SD = 1.96). However, the effect was also non-significant, U = 68.5, z = -.210, p = .843, r = -0.43.

Because there was a strong correlation between the *usefulness and usage of the transparency information*, we investigated the effect on the interaction between these variables with MANOVA, however, since these variables differ significantly from a normal distribution we used a bias corrected bootstrapped confidence interval based on 1000 samples. Using Pillai's trace, we found a significant effect of *the mission block that provided transparency* on the *usefulness of transparency information* and the *usage of transparency information*, V = .41, $F(2,21)$ = 7.31, p = .004. The graph in Figure 4.10 shows the effect.

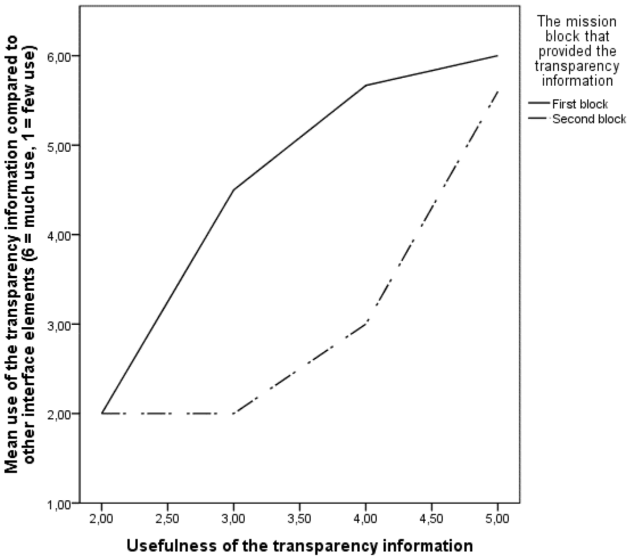


Figure 4.10: Interaction between usefulness and usage of the transparency information, separated by the block that provided the transparency information

The figure shows that the ratings of participants for *usefulness* ranged from very low to very high in both conditions. The *usage of the transparency* however, was significantly higher when it was provided in the first mission block for participants who rated the usefulness as average. The effect was also seen during the experiments. Users who received transparency information in the second mission block, generally used the transparency information less. Users who received transparency in the first mission block, tended to use the transparency information more. No other significant main effects on the *usefulness and usage of transparency information* were found.

No significant main or interaction effects of the rated *usefulness* and *usage* of the transparency information on the dependent variables were found. An interaction between *usefulness* and/or *usage* with the *transparency condition* also had no significant effect on the dependent variables.

4.3.9. Performance: scheduling time and compliance

A higher rating of *usability* was significantly correlated with a decrease of *appropriate non-compliance*, τ = -0.28, BCa 95% CI [-0.474, -0.067], p = .030, so more wrong solutions were accepted when *usability* was higher. Higher *usability* was almost significantly correlated with an increase of *appropriate compliance*, τ = 0.24, BCa 95% CI [-0.013, 0.467], p = .064, so also more correct solutions were accepted when *usability* was higher.

Although participants were instructed to try to finish each mission in 4 minutes and several participants indicated that it was too short, other participants did require only approximately 3 minutes per mission. The findings suggest that there was no learning effect over time for the participants about the execution of their task.

A difference between the mission sets could also have affected the performance. While the mission sets were counterbalanced we have to check specifically for a possible effect. There was however no significant main effect of the *mission set* on the measures of performance, see **Table 4.3**.

Table 4.3: Average values per mission set and statistics for the effect of the mission set on scheduling time, appropriate (non-)compliance and correct compliance rating

	Value missions 2-4	Value missions 6-8	Statistic	*p*
Duration	*M* = 3.97	*M* = 3.73	*F*(1,43) = .960	.330
Appropriate compliance	*M* = 0.87 *Mdn* = 1	*M* = 0.74 *Mdn* = 1	*U* = 238.5, *z* = -1.171, *r* = -0.171	.242
Appropriate non-compliance	*M* = 0.75 *Mdn* = 1	*M* = 0.70 *Mdn* = 1	*U* = 261, *z* = -0.412, *r* = -0.060	.680
Correct compliance rating	*M* = 0.68 *Mdn* = 0.70	*M* = 0.65 *Mdn* = 0.70	*U* = 259, *z* = -0.370, *r* = -0.054	.711

4.3.10. Situational awareness

Appendix J show there is a significant correlation between the *scheduling time* and *SA*. Figure 4.11 shows the interaction between *scheduling time* and *SA* for the condition with and without transparency.

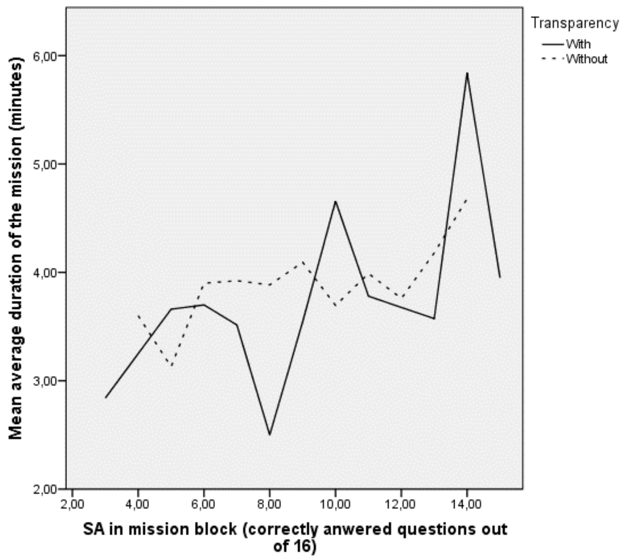


Figure 4.11: Mean of the average scheduling time (minutes) versus the participant's SA in the mission block (correct answers), separated for the conditions with and without transparency

The positive slope of the lines in Figure 4.11 suggest there is a moderate increase of SA as a function of scheduling time. A univariate ANOVA however only showed a significant main effect of the scheduling time on the SA in the transparent condition, $F(1,21) = 4.38$, $p = .049$. In the condition without transparency no significant effect was found, $F(1,20) = 2.32$, $p = .144$. The graph does however suggest there is a small increase for both conditions and this is also what we would expect, since participants are exposed longer to the situation.

4.3.11. Trust and CTL

The effects of *trust* and *CTL* are analyzed in the same section, since many correlations involved both variables.

Appendix J shows significant correlations between *trust*, *CTL* and *appropriate compliance*. An additional inspection of the correlation between the *attitude towards UVs* with *appropriate non-compliance* showed a non-significant correlation but with a predominantly negative confidence interval for the coefficient, τ = -.20, BCa 95% CI [-.453, .072], p = .160.

The correlations also showed that *CTL* was lower for participants with higher *trust* or who had a more positive *attitude towards UVs*. We investigated the effects of more demographic information about the participants on *CTL* and *trust*. We tested for main effects of the independent variables *attitude towards UVs*, *gender*, *usability* (the transformed values) and *AC* on *trust* and *CTL* using an ANOVA test and a Kruskall-Wallis test for the effects of *VGE*. The results in **Table 4.4** show that most effects on *trust* and *CTL* are significant, except for the effect of *VGE* on both variables and the effect of *AC* on *trust*.

Table 4.4: Main effects of variables on trust and CTL

	Trust	CTL
Gender	$F(1,44) = 10.51$, $p = .002$	$F(1,43) = 4.87$, $p = .033$
Attitude towards UVs	$F(1,44) = 5.77$, $p = .021$	$F(1,41) = 6.13$, $p = .002$
Video gaming experience	$H(5) = 9.261$, $p = .099$	$H(5) = 8.788$, $p = .118$
Usability (transformed)	$F(1,44) = 7.52$, $p = .009$	$F(1,43) = 5.88$, $p = .020$
Attentional control	$F(1,44) = 0.03$, $p = .861$	$F(1,43) = 4.06$, $p = .050$
Trust	-	$F(1,42) = 6.42$, $p = .015$

The correlations in Appendix J showed that almost all of these effects were negative, so participants with more *trust*, a better *attitude towards UVs*, males and participants with higher perceived *usability* generally had less *CTL*. Participants with higher *attentional control* however had more *CTL* in general.

We investigated the effect of the independent variables on an increase of *trust* and a decrease of *CTL*. Table J.1 and Table J.2 show that *gender* and *attitude towards UVs* are significantly correlated with *usability* and *VGE*. These correlations suggest there is an underlying relation between variables. Based on the correlations, we tested for mediation effects of independent variables on *gender* as a predictor of *CTL* and *trust*. Therefore we used regression analysis with the analysis tool PROCESS ((Hayes, 2012)).

Analysis showed an almost significant indirect effect of *gender* on *CTL* through *attitude towards UVs*, b = 0.075, 95% CI [-0.007, 0.150]. The probability of the direct effect of *gender* on *CTL* decreased to p = .528, so after controlling for *attitude towards UVs*, *gender* was no longer a significant predictor of *CTL*. The analysis also showed nearly significant indirect effects of *gender* on *CTL* through the combination of *usability*, b = -0.537, p = .048, and *attitude towards UVs*, b = -0.072, p = .076. The resulting direct effect of *gender* on *CTL* was b = .003, p = .969. Although one of the two indirect effects was nearly but not significant, the results show that *gender* is no longer a predictor of *CTL* after controlling for *usability* and *attitude towards UVs*.

We also checked for an effect of the demographic variables on *CTL* through *trust*. Analysis showed no significant indirect effect of *attitude towards UVs*, *usability* or *VGE* on *CTL* through *trust*. An interaction between *attitude towards UVs* and *VGE* on the contrary, which values did not deviate significantly from normal distribution, showed a significant indirect effect. However, the total effect of this interaction on *CTL* was nearly but not significant, p = .055. There was a significant indirect effect of the interaction on *CTL* through *trust*, b = -0.058, 95% CI [-0.185, -0.001], see Figure 4.12. Although the total effect was not significant, the results show that *attitude towards UVs* $*$ *VGE* together were no significant predictor of *CTL* after controlling for *trust*. Also without interaction, *attitude towards UVs* and *VGE* individually were no significant predictor of *CTL* after controlling for *trust* and *VGE* respectively *attitude towards UVs*. From these results we will not conclude that mediation occurs, however the results show that *trust* likely is a mediator of variables that describe the attitude of a participant towards the system as a predictor of *CTL*

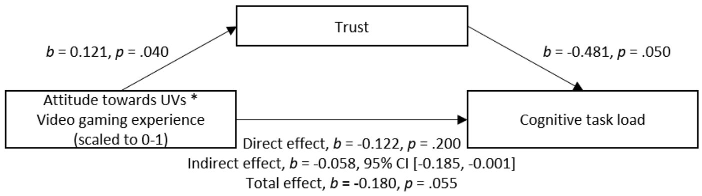


Figure 4.12: Model of attitude towards UVs and VGE as predictor of CTL, mediated by trust. The confidence interval for the indirect effect is a BCa bootstrapped confidence interval based on 5000 samples

Trust
Usability had a significant main effect on *trust*, $F(1,44) = 7.52$, $p = .009$. The Kruskall-Wallis test indicated that *trust* had a significant main effect on *appropriate compliance*, $H(1) = 5.58$, $p = .018$. For completeness, participants with more trust accepted a wrong schedule more than participants with lower trust, however the effect was not significant, $H(1) = 0.46$, $p = .500$, .

Cognitive task load
Section 4.3.6 showed a significant correlation of *CTL* with *usability* and *scheduling duration*. The main effect of *usability* on *CTL* was significant, $F(1,43) = 6.01$, $p = .018$. Section 4.3.6 already showed that this was a negative correlation, so participants with a better perception of *usability* generally had less *CTL*. Section 4.3.6 also indicated that *CTL* was positively correlated to *scheduling duration*, this effect was also found in follow-up analysis using linear regression, $F(1,43) = 6.80$, $p = .012$.

Figure 4.13 shows the CTL, categorized into five levels of scheduling durations. The figure shows the CTL per transparency-condition for each level of scheduling duration. The figure shows that the CTL increases with an increase of scheduling time, except for two participants who required 6 minutes (and didn't notice the time). Section 2.6.4 described that CTL is composed of three load dimensions, these are the time the user is occupied, the level of information processing resulting from the cognitive complexity and the amount of task switches requiring different sources of human knowledge. Since the amount of task switches is zero for all participants, because there is only one type of task the users focus on, we investigate the CTL categorized over 5 levels of scheduling duration. This should show a difference in CTL, due to a possible difference in cognitive complexity between both transparency conditions, for each level of duration. The box plots show a clear difference in CTL for participants with an average mission duration of 5 minutes, however there seems to be no difference for the other participants. Further analysis shows no significant interaction effect of scheduling duration (both categorized in 5 and 10 levels) and additional transparency on CTL, $F(1,36) = 0.60$, $p = .617$.

There was also a significant main effect of *attentional control* on *CTL*, $F(1,43) = 4.06$, $p = .050$. The coefficient of the correlation is positive, so more *attentional control* generally led to more *CTL*, this effect was not expected. Therefore, we investigated the effect of *attentional control* on *CTL* for both transparency conditions. The analysis showed that *transparency* and *attentional control* had a significant interaction effect on *CTL*, $F(1,20) = 4.79$, $p = .041$. Therefore we investigated the difference of the rating of *CTL* per participants between the transparency conditions. The result is shown in Figure 4.14, the color indicates if the transparency condition was provided in the first or second mission block.

The figure shows a positive relation between the increase of *CTL* due to the additional transparency for

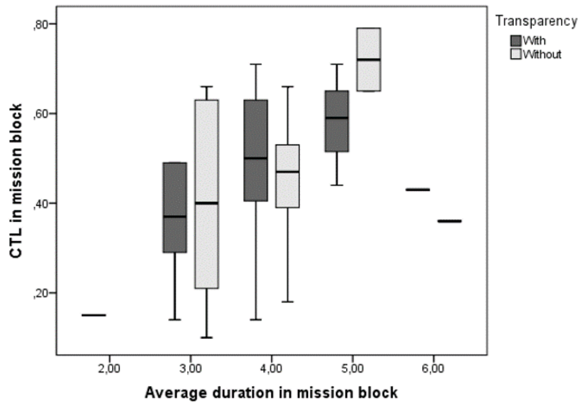


Figure 4.13: CTL per transparency condition, categorized over the required scheduling duration

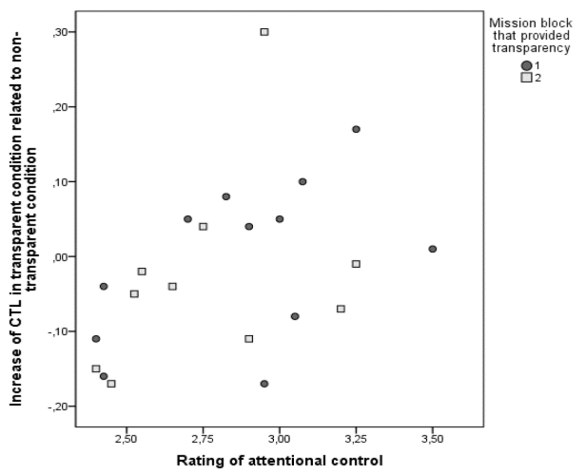


Figure 4.14: Scatter plot of the rating for CTL versus AC, the color indicates if transparency was in the first or second mission block

participants with higher *attentional control*. So a higher level of *attentional control* of a participant resulted in more increase of *CTL* in the condition with additional transparency compared to the condition without transparency. In the condition with additional transparency, *attentional control* had a near significant main effect on *CTL*, $F(1,21) = 4.17$, $p = .054$. In the condition without additional transparency, *attentional control* did not have a significant main effect on *CTL*, $F(1,21) = 0.365$, $p = .552$[1].

For closer inspection of the effect we use the logfile of eSupport about the amount of times the local utilities of the tasks and the overall utilities were inspected by the participant[2]. The analysis showed that participants with more *attentional control* in the condition without transparency inspected all available information almost significantly more, $F(1,22) = 3.76$, $p = .065$, and the local task utilities significantly more, $F(1,22) = 5.70$, $p = .026$. Since the overall utilities were not available in this condition, it makes sense to use the amount of inspections of the local task utilities as a measure instead of the total. In the condition with transparency, participants with higher *attentional control* inspected all information significantly more, $F(1,22) = 4.47$, $p = .046$.

Regression analysis was used to investigate indirect effects of *AC* on *CTL* through the *amount of inspections of information*. The regression analysis indicated a positive indirect effect of *AC* on *CTL* through the *amount of inspections of utilities* in the condition with transparency, $b = 0.11$, 95% CI [0.018,

[1] Although we did not identify the value at (2.95, 0.3) in Figure 4.14 as an outlier, it does seem to have a large impact. Excluding the value from the test did however not result in a significant main effect, $F(1,20) = .511$, $p = .483$

[2] These numbers were highly correlated with the ratings of usefulness and usage of transparency

0.284]. The probability of *AC* as a predictor of *CTL* decreased much after controlling for the mediator in the condition with transparency, *b* = 0.11, *p* = .31.

In the condition without transparency, the results indicated an indirect effect of *AC* on *CTL* through the *amount of inspections of task utilities*, *b* = 0.144, 95% CI [0.018, 0.339]. The probability of *AC* as a predictor of *CTL* decreased after controlling for the mediator in this condition as well, *b* = -0.06, *p* = .648. Figure 4.15 provides an overview of the direct and indirect effects of *attentional control* on *CTL*.

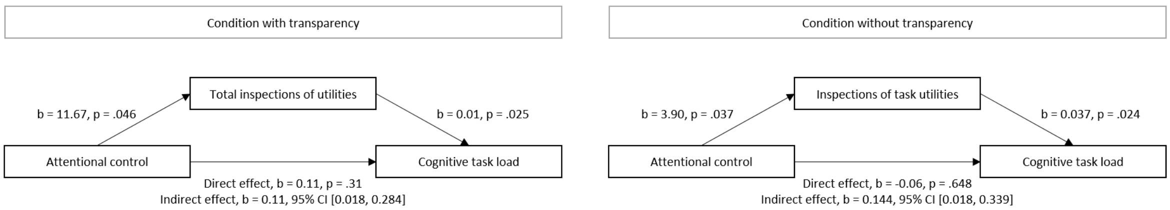


Figure 4.15: Model of AC as a predictor of CTL, mediated by the amount of inspections of utilities, for the conditions with and without transparency. The confidence interval for the indirect effect is a BCa bootstrapped CI based on 5000 samples

Because there was only a near significant direct effect of *AC* on *CTL*, the results explain the observed nearly significant effect.

5

Discussion and conclusions

5.1. Discussion

This section starts with a discussion about the experimental results and the design of eSupport. The section after the discussion describes the validation of the claims, defined in Section 3.1.4, followed by the recommendations, the research limitations and the recommendations for future work.

5.1.1. Experimental results

Additional transparency and the order of the conditions with and without transparency did not have a significant main or interaction effect on the dependent measures. However, interesting effects were found.

There were many correlations between the *gender*, *VGE*, *attitude towards UVs* and *perceived usability* of participants and *trust*, but there was not enough statistical power to draw conclusions about them. However, the correlations suggest there is an underlying structure between these variables. We found that a higher level of *trust* was significantly correlated with more *appropriate compliance* and a decrease of *CTL*. Besides, a better attitude towards UVs was significantly correlated with higher *trust* in eSupport and with lower *CTL*. Figure 4.12 shows a significant mediation effect of the interaction between video gaming experience (*VGE*) and *Attitude towards UVs* on *CTL* through *trust*. These findings match the theory about automation-induced compliance, described in Section 2.6.5. The theory defines that a higher level of *trust* typically results in automation-induced compliance, resulting in a decrease of *CTL* and causing the user to follow the decisions of the automation without further checking these decisions, the information they are based on or contradicting information.

Mediation analysis also suggested that there is no direct effect of *gender* on *trust* and *CTL*, rather has an indirect effect through *usability*, *attitude towards UVs* and *VGE*. This suggests that *gender* was not the cause of higher *trust* and thus automation-induced compliance, *usability*, *attitude towards UVs* and *VGE* were more probable causes.

Usability is not a characteristic of participants, so we investigated the effects on the perception of participants of *usability*. Several participants indicated that they found the *usability* very good, but rated *usability* lower because eSupport was not designed specifically for them but for a specific type of user: military operators. This suggests that participants that could identify themselves less with military operators rated usability lower and subsequently had less *trust* and less automation-induced compliance.

Participants who received additional transparency in the first mission block used transparency information more than the participants of the other condition. Observations during the experiment showed that participants developed a method to check the schedule for mistakes when there was no transparency information available. Participants who received the additional transparency in the second mission block tended to check the schedule the same way as they did in the first mission block, even when additional transparency was provided. This effect was also identified by Dzindolet et al. (2001), see Section 2.6.5.

Participants with high attentional control were inclined to inspect all available information. While there were no significant differences between the conditions with and without transparency information on *CTL* in general, participants with high attentional control perceived a significant increase of *CTL* in the condition with transparency information. This suggests that *AC* is an important moderator for *CTL*.

There was no order effect on compliance and non-compliance, so participants generally did not get better in their task over time. This suggests that there has been sufficient training for the task, although there was a steep learning curve for the instruction. During the experiments, we observed that several participants checked the schedule without using the overall transparency information, but often considered the local performance only. These participants sometimes still had high *appropriate (non-)compliance*. These observations, together with the lack of a learning effect, suggest that the missions could have been made more complex. However, participants already indicated that the task was difficult and complex. They probably thought so, because it was not possible for them to tell if a schedule was optimal with certainty, since the participants did not get feedback during the (training)missions. Therefore, it seemed that they felt uncertain about their performance and thus found the task difficult.

The order (learning) effect on *SA* can be explained with the comments of participants, who indicated that their understanding about what they should pay attention on, increased after the first *SA* and *trust* questionnaires. Since the performance of participants did not improve over time, this suggests that the participants knew what to do but needed more experience to get comfortable with the task and to understand what information was important for them.

5.1.2. Design of eSupport

The participants gave the usability of eSupport a good rating and found the system professional and useful. When talking to the participants after the experiment, they were enthusiastic about eSupport and really liked to do the experiment and to work with eSupport, even though executing the task for eight missions was rather monotonously.

Participants had trouble with deciding what the setting for their intention should be when the briefing indicated that a certain parameter was not very important. There are plenty situations for which there is no need to set a certain parameter or where several parameters are important, but one is more important over the other. For example, a situation in which safety is more important than completing the mission fast. Also, in some situations, users might want to indicate a deadline to complete all missions, instead of only indicating a high pressure. These findings suggest that the identified intention parameters as such do not yet provide sufficient freedom for users to indicate their objectives for actual use of eSupport.

Design assumptions

The design of eSupport was based on assumptions about the future application of UVs in military operations, listed in Section 2.8. Assumptions DA0001, DA0002 and DA0003 were validated through interviews with military subject matter experts (SMEs). The other assumptions have not yet been validated. Assumption DA0004 is about the autonomous capabilities of UVs to execute high-level tasks. It is not possible yet for SMEs to draw conclusions about these aspects. However, based on current practice and international developments, SMEs found this assumption fair and reasonable. The assumptions DA0005 and DA0006 describe that the defined parameters are sufficient to align the user and system objectives. We discussed that the intent parameters should be expanded to represent the user objectives sufficiently. Assumption DA0007 is about the functioning of the chosen representation of a BMS. This assumption require a user study to be validated, the experiment gave no reason to question this assumption however.

5.1.3. Validation of claims

This section discusses the claims for requirement RQ0010 about additional transparency. Since no significant effects were found, the means of the values for each transparency condition provided in Table 4.1 are used for the validation of the claims.

- **Efficiency**: Additional transparency improves the user's efficiency by providing appropriate information about the compliance with their objectives at the right level of abstraction.
  Partially validated (without sufficient statistical power): No significant effect of improved transparency on mission duration was found. The mean scheduling time for the condition with transparency (M = 3.78 min, SD = 0.91) and without transparency (M = 3.91 min, SD = 0.49) in Table 4.1 shows there was a small decrease of duration in the condition with additional transparency.

- **Appropriate (non-)compliance**: Additional transparency improves the understanding of the user about the compliance with their objectives, this should support them to make a better estimation about the correctness of a schedule.
  Partially validated (without sufficient statistical power): no significant effect of additional transparency on appropriate compliance and appropriate non-compliance was found. The mean of appropriate compliance in the transparent condition was however higher (M = 0.83, SD = 0.53) than in the non-transparent condition (M = 0.75, SD = 0.75), see Table 4.1. The mean of appropriate non-compliance in the transparent condition was also higher (M = 0.61, SD = 0.44) than in the non-transparent condition (M = 0.55, SD = 0.48). This shows there was a small increase of appropriate (non-)compliance in the condition with additional transparency.

- **Situational awareness**: Additional transparency improves the user's *SA* by providing information about the consequences of the schedule on the user's objectives.
  Partially validated (without sufficient statistical power): no significant effect of improved transparency on SA was found. The mean of SA in the condition with transparency (M = 8.96, SD = 3.28) was however higher than in the condition without transparency (M = 8.82, SD = 2.36), see Table 4.1. This shows there was a small increase of SA in the condition with additional transparency.

- **Trust**: Additional transparency improves calibrated *trust* by improving the user's understanding of the consequences of the schedule related to the operator's objectives.
  Partially validated: no significant effect of improved transparency on trust was found. Table 4.1 shows no difference in the mean of trust between the condition with improved transparency (M = 0.70, SD = 0.12) and without (M = 0.70, SD = 0.10).

- **Cognitive task load**: Additional transparency decreases *CTL*. By providing additional transparency, the participants need to handle more information, however we expect that this information supports the understanding and decreases the necessity of checking all other available information, and thus we expect a decrease in the level of information processing. Since the time occupied and task set switches are fixed, we expect a decrease of *CTL*. *Not validated: no significant effect of improved transparency on CTL was found. Table 4.1 shows no difference in the mean of CTL between the condition with improved transparency (M = 0.46, SD = 0.17) and without (M = 0.46, SD = 0.18).*

5.1.4. Recommendations

The sample group mostly included participants without military experience, since we assumed that military experience would not influence the results for the current study based on prior studies (see Section 4.2.1). However, this study showed that the attitude towards UVs, VGE and characteristics of participants did have an impact on *trust*, *CTL* and *performance*, through automation-induced compliance. It is very probable that domain experts, let alone UV operators, have specific values and characteristics. For example, UV operators probably will have an attitude towards UVs at least as good as participants. This finding can be useful to calibrate the *trust* of users in a system during a training, this might result in more *appropriate (non-)compliance* and can be used to balance *CTL*. More research is required to investigate the effect of the attitude towards the system and the user experience on *trust* and on resulting effects.

The order of providing transparency information had an impact on how the information was used. For future implementation of supportive systems, this effect can be used for the calibration of *trust* and to balance *CTL* in future systems to ultimately improve performance in terms of *appropriate (non-)compliance* and *duration*. Future developers have to think carefully about how they introduce a supportive system to the user. Exposing the user to supportive information directly might cause a higher level of *trust* than when the user first has to carry out the task without or with less support.

The opinion of participants about the complexity of the missions was contradicted by the results for the performance. It seemed that participants perceived complexity based on their capacity to achieve good performance. Therefore, we recommend to design the complexity of the missions with great care and investigate what the opinions of participants about the complexity are based on. Furthermore, we recommend to use more training missions and to provide transparency about the performance during these missions, this would have allowed participants to improve their strategy and adjust the level of effort for the task (Hart and Staveland, 1988). The consequence is, that this probably requires a between-subjects approach.

While the proto-patterns have not been validated yet, the comments and ratings of the participants about eSupport were positive and promising. The prototype of eSupport offers many options for further development to execute experiments with users. Because good usability has to be guaranteed to investigate human-factors such as *trust* and *CTL*, we recommend to use eSupport as a basis for experiments, instead of designing and building a new prototype.

5.1.5. Research limitations

Two important limitations of the study were identified.

Design of experiment

There was no funding available for the user study, so we had to take into account for the experimental design that it would probably be hard to find participants. Because of this, we decided to test the effect of additional transparency using a within-subjects design, this requires fewer participants than a between-subjects design to achieve the same statistical power. However, since *trust* can be biased through earlier experiences with the system (see Section 2.6.5), the effect of the additional transparency in contrast to the condition without additional transparency might have changed. For the same reason of funding, we decided to reduce the *duration* of the experiment by decreasing the number of (training)missions per participant. Subsequently, participants received less training and did not receive much feedback on their performance, this might have affected the confidence of participants in their performance and subsequently their *trust* and *CTL*.

Near optimal algorithm

The algorithm used in the experiment did not produce optimal solutions. Some participants observed that the schedules of missions 4 and 8 could have been slightly better (while these should have been optimal). Therefore, the participants gave these schedules a worse rating (ranging from 40 to 80) and rejected them, since they noticed that the schedule could be improved. Our method to measure the compliance of the participants, only allowed us to use the data for schedules that were optimal or bad. For near-optimal schedules, participants did not know if they should accept or reject them and how these schedules should be rated. The measures for compliance, so the choice between accepting or rejecting a schedule and the rating of a schedule, should have been explained to the participants in more detail. This would have resulted in more values for the measure of compliance and thus would have given more statistical power.

5.1.6. Future work

The observation of automation-induced compliance for participants without knowledge about the domain, raises the question if subject matter experts (SMEs) would be less sensitive for automation-induced compliance. A similar study with SMEs would create more insight in these effects and would be useful for future development of supportive systems.

We discussed that the missions should have been more complex to increase the necessity and value of the additional transparency information. A follow-up study with more complex missions allows to investigate the influence of complexity on the effects of additional transparency.

Section 5.1.1 suggested to expand the intention parameters with priorities and to allow users to set constraints for parameters. Section 2.5.4 describes the analysis of intent parameters, these are however based on the current scope of the study and the current algorithm. An additional analysis should be done once there are more details about the use of UVs for military operations. Especially the future ideas of the military forces about the use of UVs during tactical operations to create military effects are important, since these describe the objectives of the UVs. The user should be able to insert military effect as objectives into the system like they do to their subordinate commanders.

The high-level tasks should be expanded with sub-objectives for the task with respective priorities. This allows eSupport to make better decisions about the allocation of tasks to UVs and to reason about the consequences, solutions and the progress in case of a change of situation by for example an anomaly or an event. Future research could investigate how eSupport can use the additional information about the tasks to improve its transparency.

An indispensable feature for eSupport is path planning with related measures for (3D) deconfliction, area restrictions and phase lines (phases indicate at what time a geographical line may or should be crossed) for different UVs. The expansion of the algorithm for path-planning requires the introduction of more intent parameters and constraints.

eSupport
eSupport could be expanded with capacity to reason about and anticipate on situations to increase its supportive value. To utilize more components of eSupport, we suggest to expand eSupport with the functionality to create hypothesis about events that could reasonably occur given the purpose of the mission. For example, the application could hypothesize that a convoy will be ambushed in a certain district, based on for example historical data. Indicators are used to validate or invalidate an hypothesis with certain probability. For example, an enemy ambush requires a channeling location in the terrain, surrounded by higher ground or buildings in the direct environment. So the channeling location and higher ground at that location are indicators for the hypothesis about an ambush. By checking for indicators, the probability of the hypothesis being valid increases or decreases. Hypotheses and indicators can be integrated in eSupport in the defined hierarchical structure of tasks, their endstate, and their purpose. So, the indicators would be tasks, the validation of hypothesis would be endstates, and the purpose is equal to the purpose of the whole mission.

5.2. Conclusions

A tool called eSupport to support users with scheduling high-level tasks to multiple UVs in military operations was designed and developed in this study. The focus of the study was to investigate how eSupport can be more transparent about the compliance of a solution with the user's objectives and thereby improve the user performance, so improve *appropriate (non-)compliance*, scheduling time and *SA*. The domain analysis resulted in the design of nine proto-patterns. These are designed to:

1. Align the user and system objectives
2. Set constraints about the allocation of tasks to UVs
3. Set constraints about the order of task execution
4. Provide transparency about a solution's compliance with the user objectives
5. Provide transparency about a solution's compliance with a single user objective
6. Show the consequences of a decision on a single user objective
7. Order the information by the level of abstraction in clusters
8. Show the relation between information about different levels of abstraction
9. Separate the steps of the scheduling process

The proto-patterns were used in the design of the prototype for eSupport. A user experiment was employed with this prototype to investigate the effects of additional transparency.

The domain analysis resulted in a set of requirements for eSupport, listed in Section 3.1.3. The requirements were used for the design and development of the prototype. In the analysis, a set of five main user objectives was identified for the allocation of high-level tasks to UVs. The identified objectives are:

- Time pressure: The necessity to execute the tasks as fast as possible
- Risk exposure: The acceptance of the chance to lose a resource
- Massiveness: The pursuance to use many resources for the operation
- Secrecy: The operation is openly conducted without concealment (overt) or needs to conceal the identity of or permit plausible denial by the sponsor (covert)
- Information quality and quantity: The required resolution of information and the desired coverage of the area

The objective for *information quality and quantity* should be specified in the intelligence requirements, so the user does not have to indicate the desired value for this objective. The domain analysis also resulted in a solution to communicate the user's objectives to eSupport. Users set their desired value for each objective using a slide bar on a five-level Likert scale at the start of a scheduling process. The possible values for each objective are: very low versus very high time pressure, take no risk versus take risk, use few versus use many resources and, at last, the binary choice between overt or covert.

An abstraction decomposition space (ADS) was defined for the current domain in the analysis. The ADS was used to identify the appropriate levels to provide information to the user. This resulted in two levels to show transparency information for. The endstates are the first level, these are functions that can be executed physically, like the tasks, but described at a higher decomposition level (so for a hierarchically higher military unit). The second level is the purpose of an endstate, this is a military effect that has to be achieved.

eSupport uses a pie chart icon to provide additional transparency. This icon shows the compliance of a solution (at a certain level) with each user objective and the balance between the compliance with each objective. Showing this balance is important, since the objectives are conflicting and users should understand this conflict. eSupport shows this icon in the transparency condition at the two defined appropriate levels of the ADS.

The effect of the additional transparency on participants was investigated using a mixed-factorial design. The transparency condition was the independent within-subjects variable and the order of the conditions with and without additional transparency was the independent between-subjects variable. The experiment did not show a significant main or interaction effect of the independent variables on the participants' performance (duration and *appropriate (non-)compliance*), *SA*, *trust* and *CTL*. However, there were non-significant effects of additional transparency on *performance* and *SA*, all effects were positive and in line with the hypotheses. *Trust* and *SA* were not effected.

Several other effects were observed. There was an effect of automation-induced compliance, so an increase of *trust* that results in more compliance to correct and incorrect suggestions of eSupport and a decrease of *CTL*. This effect was mostly observed for participants with a more positive attitude towards the use of UVs and for males. The effect of gender on *trust* and *CTL* however, was caused by indirect effects through *attitude towards UVs*, *VGE* and *usability*.

The order of the conditions with and without additional transparency affected the degree to which participants used the information. When the additional transparency was provided in the first mission block, participants generally made more use of it.

Attentional control (AC) affected the way participants dealt with their task, participants with higher *AC* did more thorough checks of the available information. The additional transparency information subsequently resulted in an increase of *CTL* for participants with high *AC*. Participants that rated usability of eSupport higher, however, had a lower perception of *CTL*.

eSupport proofed to be a platform with good usability and high ease of use. eSupport is flexible for further development and based on seemingly valid assumptions about the future use of UVs in the military domain. The prototype of eSupport provides a good basis for further development and investigation of supportive systems for the application of UVs in military operations and possibly in similar domains such as USAR.

A
Interviews

Two interviews with captains of the Dutch Army were done on Februari 11, 2016 to improve the understanding of the current use of UVs. A third interview was done with an engineer and a captain of the Defense Material Organisation. The most relevant outcomes of the interviews are provided in this appendix. No names are provided in this chapter for security and privacy reasons.

A.1. Interview Chief of Operations JISTARC

This interview was done with a captain of the Joint Intelligence, Surveillance, Target Acquisition & Reconnaissance Commando (JISTARC). The captain served as the chief of operations in Mali as part of an intelligence (ISTAR) module, which goal is to gather intelligence. The module can be classified in roughly three components, the first component is the Collection Coordination and Information Requirements Managements (CCIRM) group, this group is in charge of the other components and manages the plan to collect intelligence information. The second component is the operations cell, this cell is responsible for the actual collection of information through various sensors. These sensors are the units that collect the information in various ways, e.g. by means of electronic warfare, reconnaissance, field human-intelligence and aerial vehicles. The Chief of Operations (CHOps) is in command of this cell and communicates with the deployed sensors through respective liaison officers called Mission Managers. The third component is the All Source Intelligence Cell (ASIC), the cell is responsible for the analysis of the information. The cell contains analysts with different specializations such as GeoSpatial, Human terrain and Military.

An ISTAR module works with a list of required intelligence: the Intelligence Collection Plan (ICP). The CCIRM-group manages the ICP and gives tasks to collect information to specific sensors in a Collection Exploitation Plan (CEP). The respective analysts of the ASIC analyze the resulting information returned by the sensors. The operations of the sensors are typically planned days ahead and performed at great distance.

The captain has confidence in future deployment of autonomous vehicles in military operations, however only if deconfliction of UVs with other forces or areas is guaranteed. According to the captain, this is a hard restriction which has to be addressed before autonomous UVs should be deployed together with other troops.

A.2. Interview Officer UAS flight operations

This interview was done with a captain of the Dutch Army, who was deployed multiple times in the role of Officer UAS (Unmanned Aerial System) flight operations (Mission Manager) in Afghanistan and Mali. According to the captain, the UV they use, the ScanEagle, requires at least two hours preparation for planning tasks and technological checks, however this could go up to days whenever air space permission is not arranged yet. The typical course of events is that the Mission Manager first receives the task to collect certain information after which he produces a formal mission order for the task. After that, he takes care of air space permission, deconfliction and coordination and finally he provides a target brief to the UAS team to actually execute the mission. At that point, the actual execution of the mission can start given that the UAS team finished all preparations. Concluding, deploying the ScanEagle requires a lot of planning and preparation time.

UAVs are also used in different ways in current operations. Reconnaissance units of the Dutch Army possess small UAVs called the Raven. The Raven can be launched with a very short preparation time and can be deployed whenever the operational commander decides to. Because small UVs typically are light weighted, have a small action radius but are able to operate almost instantly, these systems are very suitable for these units. Small autonomous systems therefore seem to be appropriate for the purpose of tactical units, since they give them the advantage of rapid deployment whenever required.

A system is composed of an analogue system with six planes or a digital system with 3 planes each. The crew includes the ground commander, a vehicle in control of the vehicle, the payload (camera) operator, an IVES (INSITU Video Exploitation System; actors that analyze the videos in real time and make the intelligence report/product) and the ground crew.

A.3. Interview Defense Material Organisation

The interview was done with an engineer and an army captain of the Defense Material Organisation and focuses on the expectations of the future employment of UVs in military operations. Because the interviewees indicated that they had no knowledge about a concept of operations for small tactical UVs in the Netherlands, we spoke about the expectations based on the current developments in the Netherlands and the US. A more specific concept how to embed UVs in the military structure will most probably be developed once more small UV systems are available for the military forces.

Current initiatives

There are several initiatives in the Dutch military forces for the use of teleoperated small UVs. Several units at company level are testing various teleoperated UVs during military exercises. In the US Marine Corps there are similar initiatives, more over they tested the usage of a dedicated UV-operator per squad (a team of approximately 12 marines) to operate the UV and to support the squad leader with the analysis of the information. These initiatives indicate that tactical units are open minded towards the use of UVs, moreover are possibly willing to appoint a dedicated UV-operator to a squad. Currently, the units only employ one teleoperated UV at a time, however the captain indicated that once more types of UVs are available to a unit, he expects that multiple (autonomous) UVs will be deployed concurrently, also in cooperation with other units. Currently, in the Dutch army eSupport would fit better at battalion level according to the captain, since not many UVs are available yet. In the future however, a system like eSupport would also be appropriate for lower tactical levels once they have multiple small tactical vehicles embedded.

User values

In addition to the defined values for the system in the current study, the captain indicated that the balance for a task between deployment of soldiers, and thus putting soldiers to risk, or accepting more time to prepare and deploy UVs (presumed that the required staff and technology is available) will be a major value. Besides, the trustworthiness and validation of information collected by UVs is an important concern, since different sensors (e.g. acoustic, visual and magnetic) have different reliability. The captain indicated that decision authority indeed is very important, since a system should always allow an operator to redirect UVs once the dynamics of the environment or operation change. The value deconfliction will surely be addressed according to the captain, since the dynamics of warfare will change in future electronic warfare and separate (aerial, maritime and ground) vehicles and other systems will constantly communicate with each other and thereby assure deconfliction themselves.

It is likely that a team of UVs will be used to create certain military effects in the future. This conclusion is based on the use of teleoperated UVs to create military effects in prior military operations. For example, only the presence of certain vehicles (such as the ScanEagle or Apache) were enough to deter the opponent or to enforce them to stop a fight. It would therefore be meaningful to share these objectives about the desired effects with eSupport.

Integration with systems
The current design of eSupport is based on the assumption that it will be integrated with a BMS and that it can use all available information about the situation. The captain indicates that cooperation between these systems is a definite requirement. More over, a recently introduced system to support command and control and share information between all levels of command in the Dutch army allows commanders to indicate their intelligence requirements. This results in a global intelligence requirements list available to all levels. Once an intel requirement is met the result is displayed in BMS if it is geographical related information. The captain suggested that, on top of the vehicles of the STUV team, eSupport could also consider to assign (or advise to assign) tasks to vehicles that are already standby in the nearby area, such as a ScanEagle or Reaper.

B

Military activities

The military doctrine LDP-1 (Dutch Royal Army Doctrine Commitee, 2013) identifies four categories of activities at the tactical level: offensive, defensive, stability and enabling activities. A fifth but different category is *information operations* that concern the coordination of activities in the information domain, they aim to influence knowledge, attitude and behavior. Information operations are planned and executed in close cooperation and coherence with the tactical activities, but are not a separate form of operation.

Offensive activities The goal of offensive activities is to eliminate the factors that keep a crisis going, in most cases by eliminating the opponent. The doctrine identifies eight offensive activities:

- Attack
  - Deliberate attack;
  - Hasty attack;
  - Counter attack;
  - Spoiling attack.
- Raid;
- Reconnaissance in force;
- Pursuit;
- Demonstration;
- Feint attack;
- Ambush;
- Breakout of encircled forces.

Defensive activities Defensive activities are defined as the activities to resist enemy offensive activities (Dutch Royal Army Doctrine Commitee, 2013). The purpose of these activities is to remain own power and involve the combination of essential information, personnel, materiel, infrastructure and terrain (Noël, 2009).

The doctrine identifies two types of defensive activities:

- Defense;
- Delay.

Stability activities Stability activities are executed in close collaboration with other actors to maintain, recover or create a situation to ensure proper functioning of the responsible governments. The doctrine identifies four types of stability activities:

- Security and control (enforcing and controlling a secure environment);
- Support to security sector reform (support for the reform of local authorities responsible for public order and security);
- Initial restoration of services (for infrastructure and utilities);
- Interim/initial governance tasks.

Enabling activities Enabling activities support and link the transition between offensive, defensive and/or stability activities and are therefore never independently executed. The doctrine identifies thirteen enabling activities:

- Reconnaissance;
- Security;
- Advance to contact;
- Meeting engagement;
- Link-up;
- Relief of encircled forces;
- Relief of troops in combat;
- Retirement;
- Withdrawal;
- March;
- Obstacle breaching/crossing;
- (Tactical) Assembly area;
- Extraction.

C

Components of the commander's intent

Hieb and Schade (2007) provide an example of a commander's intent for a multinational force (MNF). We included this example to give an idea about the content of a commander's intent. Note that a combined arms team is at a much lower level in the military hierarchy and only plays a small role within the intent of the MNF commander.

> 3. EXECUTION.
>
> The purpose of this operation is to enable establishment of regional military stability in operations zone A. This will require our forces to maneuver rapidly from an attack position along river B to seize objectives C and D, destroy enemy forces occupying key terrain in operations zone A and secure the international border. Destroy enemy forces that engage our forces or occupy positions on key terrain along our axes of advance. The key to our success during the attack will be that our forces gain a twofold surprise by courageous and insistent attacking while the enemy is still deploying his forces and by aggressive reconnaissance of suspected and known enemy locations along the axis of advance to identify and destroy hostile reinforcing forces as early as possible. At the conclusion of the operation, our forces will have:
>
> - Destroyed enemy occupying key terrain in zone A;
> - Established hasty defensive positions in objectives C and D;
> - Established our forces on the international border

We can distinguish the three components:

- Purpose: Enable establishment of regional military stability in operations zone A
- Endstates:
  1. Destroyed enemy occupying key terrain in zone A
  2. Established hasty defensive positions in objectives C and D
  3. Established our forces on the international border
- Method
  1. Maneuver rapidly from an attack position along river B to seize objectives C and D
  2. Destroy enemy forces occupying key terrain in operations zone A
  3. Secure the international border.

4. Destroy enemy forces that engage our forces or occupy positions on key terrain along our axes of advance.

Two parts of the commander's intent could not be defined. These were however defined as the key to success by the commander. These components were:

- Gain a twofold surprise by courageous and insistent attacking while the enemy is still deploying his forces
- Aggressive reconnaissance of suspected and known enemy locations along the axis of advance to identify and destroy hostile reinforcing forces as early as possible.

These keys to success describe the execution for the method section, however these sentences cannot be expressed in terms of physical functions. Rather, the commander uses values such as aggressive, surprise and courageous, to describe them.

D

Battle Management Language

The Battle Management Language (BML) is a formal unambiguous grammar for the military domain. The grammar is intended to be suitable for commanding and communicating with live forces, simulations and robotics. (Hieb and Schade, 2007) extended the grammar to communicate commander's intent. The grammar to communicate commander's intent consists of three components: the expanded purpose (Purpose), end state (End state) and key tasks (Method). All three components are expressed in the same format to report a task, event or status in BML. The task report is built up as follows:

```
Task-Report Verb Executor (Affected|Action) Where When (Why) Certainty Label
(Mod)*
```

With:

`Verb`:	The task to be executed
`Executor`:	The unit to execute the task
`Affected`:	Someone that is directly affected by the task (the patient), e.g. the enemy. It depends on the tasking verb whether this term is included
`Action`:	An action that is directly affected by the task, e.g. assist. Usage similar to affected
`Where`:	Location of the task. The form depends on the tasking verb and can be a location (At-Where) or route (Route-Where)
`When`:	Temporal condition for the task, nominators are allowed (e.g. No later than (NLT)), defined in JC3IEDM as action-task-end-qualifier-code
`Why`:	Term to indicate the purpose of the task, preceded by `in-order-to` followed by a Task, to `cause` an End state or to `enable` the Purpose
`Certainty`:	Term to indicate the certainty of the report from the sender's perspective
`Label`:	A unique identifier for the task
`Mod`:	Modifier to add additional information for further description
`*`:	Terms with an asterix mean that arbitrary many respective expressions can be concatenated together
`()`:	Terms between brackets are optional
`|`:	OR, one or both of the terms around the symbol are allowed

The key tasks (Method) can be expressed in a standard order using BML as follows:

```
Verb Tasker Taskee (Affected|Action) Where Start-When (End-When) Why La-
bel (Mod)*
```

With:

`Tasker`: The entity that gives the order
`Taskee`: The name of the unit ordered to execute the task (*OPEN* is used in case this is not known yet)

The `Why` section allows to indicate the end state and the purpose of the method. The end state can be described in a similar way as the method, however on a higher unit level. The purpose can be described by using a status from the BML-vocabulary, e.g. `prevent enemy attack` or `develop situational understanding`.

For more details about the grammar, see (Hieb and Schade, 2007).

E

Unmanned Vehicles

The current study has a particular set of fictive heterogeneous unmanned vehicles available with different capabilities. The capabilities are based on the capabilities of currently available vehicles but are assumed to be fully autonomous and unarmed. The section below provides a list of five heterogeneous UVs, of which two aerial and three ground vehicles.

E.1. Vehicles

E.1.1. UAV-1

The first UAV is based on the currently used Raven, although does not fully represent the specifications and capabilities of the Raven. The Raven is a fixed wing teleoperated aerial vehicle, categorized as mini-UAS and equipped with a color video camera or a infrared night vision camera. The specifications of the UAV-1, based on the Raven, are:

Table E.1: Specifications of UAV-1

Dimensions (w x l x h) (cm^3)	137 x 92 x 20
Action radius (km)	10
Cruise speed (km/h)	30
Cruise altitude (m)	150
Lowest operating altitude (m)	10
Weight (kg)	1,9
Propulsion	Fixed wing
Endurance (h)	4
Covert ability	No
Icon	

Figure E.1: UAV-1: Raven, developed by AeroVironment. Reprinted from AeroVironment, retrieved June 14, 2016, from http://www.avinc.com/av_uas_flash/UAS1/images

E.1.2. UAV-2

The second UAV is based on current available non-military quadcopters, these are defined as micro-UAV. Quadcopters have the ability to hover at one place, although they are typically not silent and thus have no ability to be covert nearby humans or technology with acoustic sensors. They can operate covert at appropriate distance however. The specifications of the UAV-2 are:

Table E.2: Specifications of UAV-2

Dimensions (w x l x h) (cm^3)	50 x 50 x 20
Action radius (km)	6
Cruise speed (km/h)	30
Cruise altitude (m)	50
Lowest operating altitude (m)	0
Weight (kg)	0,9
Propulsion	Rotary wing (4)
Endurance (h)	3
Covert ability	Yes (given proper distance)
Icon	

Figure E.2: UAV-2: AirRobot, developed by AirRobot. Reprinted from Swiss Securitas Asia Pte Ltd, retrieved June 14, 2016, from Swiss Securitas Asia Pte Ltd-Google+ account

E.1.3. UGV-1

The first UGV is based on the reconnaissance vehicle Fennek in combination with the UGV Crusher, developed by DARPA. The UGV-1 is wheel based and the size of a vehicle to allow higher speed movement. This requires the vehicle to keep appropriate distance from a target to stay covert. The specifications of the UGV-1 are:

Table E.3: Specifications of UGV-1

Dimensions (w x l x h) (cm^3)	250 x 570 x 210
Action radius (km)	900
Cruise speed (km/h)	115
Weight (kg)	10.400
Propulsion	Wheel (6)
Endurance (h)	9
Covert ability	Yes (given proper distance)
Icon	

Figure E.3: UGV-1: Crusher, developed by DARPA. Reprinted from: National Robotics Engineering Center, retrieved June 14, 2016, from www.nrec.ri.cmu.edu/projects/crusher

E.1.4. UGV-2

The second UGV is based on the Cobra MK2, developed by ECA robotics (www.army-technology.com/downloads/whitepapers/mines/cobra-mk2/). The UGV-1 is a small wheel based vehicle capable of operating covert. The specifications of the UGV-2 are:

Table E.4: Specifications of UGV-2

Dimensions (w x l x h) (cm^3)	37 x 39 x 17
Action radius (km)	40
Cruise speed (km/h)	25
Weight (kg)	6.1
Propulsion	Wheel (4)
Endurance (h)	8
Covert ability	Yes
Icon	

Figure E.4: UGV-2: Cobra, developed by ECA group. Reprinted from ECA Group, retrieved June 14, 2016, from www.ecagroup.com/en/solutions/cobra-mk2-i

E.1.5. UGV-3

The third UGV is based on various tracked small robots. The UGV-3 is capable of operating covert and to execute a full building reconnaissance. The specifications of the UGV-3 are:

Table E.5: Specifications of UGV-3

Dimensions (w x l x h) (cm^3)	66 x 100 (60) x 23
Action radius (km)	35
Cruise speed (km/h)	20
Weight (kg)	7.5
Propulsion	Track
Endurance (h)	8
Covert ability	Yes
Icon	

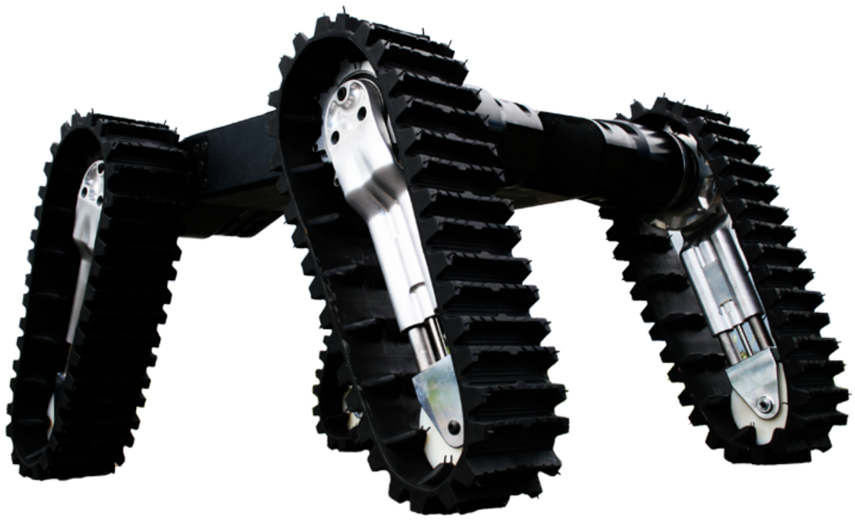


Figure E.5: UGV-3: Chaos, developed by ASI robots. Reprinted from ASI robots, retrieved June 14, 2016, from www.asirobots.com

E.2. Capabilities

The tables below shows the capabilities of the UVs to execute tasks overt/covert with the respective expected duration respectively coverage.

Table E.6: UGV capabilities: Time in minutes for overt / covert execution,
*: For the performance calculation of the algorithm, these times are corrected due to low sensor quality

	Person reconnaissance	Vehicle reconnaissance	Building reconnaissance
UAV-1	5 / -	2 / -	20* / -
UAV-2	5 / 10	3 / 6	60 / -
UGV-1	5 / 60	5 / 60	20* / 60*
UGV-2	10 / 45	10 / 45	25* / 60*
UGV-3	15 / 45	15 / 45	90 / 60*

Table E.7: UGV capabilities: Sensor coverage (%) for overt / covert execution

	Person reconnaissance	Vehicle reconnaissance	Building reconnaissance
UAV-1	100 / 0	100 / 0	25 / 0
UAV-2	100 / 100	100 / 100	100 / 0
UGV-1	100 / 100	100 / 100	20 / 20
UGV-2	100 / 100	100 / 100	50 / 30
UGV-3	100 / 100	100 / 100	100 / 30

Table E.8: Chance for critical hit (%) for overt / covert execution within 50m of the threat (The chance the asset will be neutralized)

	Rifle	RPG	Sniper	VOIED
UAV-1	9 / 2	1 / 0	5 / 1	0 / 0
UAV-2	14 / 3	2 / 0	15 / 3	0 / 0
UGV-1	5 / 1	33 / 7	16 / 3	48 / 10
UGV-2	24 / 5	20 / 4	35 / 7	15 / 3
UGV-3	24 / 5	20 / 4	35 / 7	15 / 3

F

Scheduling algorithm

The scheduling algorithm uses a greedy approach. The pseudocode is provided below.

```
List UNASSIGNED = collection of all tasks
List NOT_EXECUTABLE = empty

WHILE tasks in UNASSIGNED
  Set highest_utility to 0
  FOR each task in UNASSIGNED
    FOR each resource
      IF utility(task,resource) > highest_utility
        Assign utility(task,resource) to highest_utility
      END IF
    END FOR
  END FOR
  IF highest_utility < 0
    Assign highest_utility.task to NOT_EXECUTABLE
  END IF
  ELSE
    Assign highest_utility.task to highest_utility.resource
  END ELSE
  Delete highest_utility.task from unassigned
END WHILE
```

The pseudocode uses a utility function to calculate the utility of a resource for a task, this function is described below. Note that this function is specifically designed for the experiment.

$$V_{i,j} = Value_{i,j} \cdot Factor_{i,j}^{Time} \cdot Factor_{i,j}^{Risk} \cdot Factor_{i,j}^{Massiveness} \tag{F.1}$$

$$Factor_{i,j}^{Category} = 1 - \left(Priority_{i,j}^{Category} \cdot \left(1 - WeightedScore_{i,j}^{Category}\right)\right) \tag{F.2}$$

$$WeightedScore_{i,j}^{Category} = \frac{1}{2} \cdot \left(Cos\left(\pi \cdot \left(1 - Score_{i,j}^{Category}\right)\right) + 1\right) \tag{F.3}$$

$$Score_{i,j}^{Time} = MAX\left(0, 100 - 20 \cdot \left(\frac{Time_{i,j}}{Minimum\_Time_{i,j}} - 1\right) \cdot Intent_{Time}\right) \tag{F.4}$$

$$Score_{i,j}^{Risk} = \left(1 - (1 - Intent_{Risk}) \cdot Risk_{i,j}\right)^2 \tag{F.5}$$

$$Risk_{i,j} = 1 - \prod_{i \cap k} \left(1 - P(k) \cdot visibility_j \cdot critical_{j,k}\right) \tag{F.6}$$

$$Score_{i,j}^{Massiveness} = 1 - \delta_j^3 \tag{F.7}$$

$$\delta_j = \begin{cases} Intent_{Massiveness} & \text{if resource already scheduled} \\ 1 - Intent_{Massiveness} & \text{if resource not yet scheduled} \end{cases} \tag{F.8}$$

where (on order of appearance)

$V_{i,j}$:	Utility of asset j for the task i
i :	Task
j :	Asset
$Value_{i,j}$:	Value of asset j for task i determined by the information quality and quantity the asset would retrieve. The values are from a table, see Appendix E.2
$Factor_{i,j}^{Category}$:	Utility of asset j for task i for the Category [0-1]
$Priority_{i,j}^{Category}$:	Priority of the intent for the Category (fixed at 0.5 for all categories)
$WeightedScore_{i,j}^{Category}$:	Weighted score to punish for low utilities, rationale: a decrease from 0.5 to 0.4 is worse than a decrease from 1 to 0.9
$Score_{i,j}^{Category}$:	The extent to which the allocation of task i to asset j complies with the intent for the Category [0-1]
Category :	Intent category (*Time*, *Risk* or *Massiveness*)
$Time_{i,j}$:	Time required for asset j to complete task i
$Minimum\_Time_i$:	Time required for the asset that completes task i as fastest[1], the time is adjusted if the information quality or quantity is below 100%
$Intent_{Time}$:	The time pressure set by the operator in the intent *[0 (low pressure), 0.2, ... , 1 (high pressure)]*
$Intent_{Risk}$:	The allowed risk exposure set by the operator in the intent *[0 (be safe), 0.2, ... , 1 (take risk)]*
k :	An enemy threat
$i \cap k$:	All threats k that have task i within their range
$Visibility_j$:	The chance of asset j to be spotted (Fixed at 0.1 for Covert and 1 for Overt)
$Critical_{j,k}$:	The chance of asset j to be critically hit by enemy k, based on fixed values for the chance of a hit, the fatality of a hit and the distance
$Intent_{Massiveness}$:	The preferred massiveness set by the operator in the intent *[0 (few assets), 0.2, ... , 1 (many assets)]*

[1] Note that all times are compared to the fastest time per task, since it is important to retrieve intelligence as soon as possible for each individual task.

G

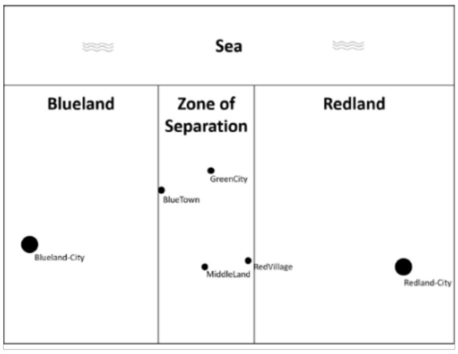
Scenario description

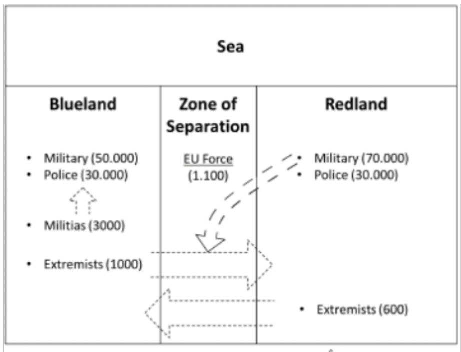
The scenario provides the general context of the operation. The scenarios are inspired by mission scenario's describe in an internal document of TNO, meant for inhouse use only. The chapter first describes the general situation in G.1. Military scenario's typically provide more information including the political situation, infrastructure, electronic measures, command & control organisation, the terrain and the weather, this information is not provided in the current study since the participants do not need this information for this specific experiment. Section G.2.1 - section G.2.6 describe the more specific event, the commander's intent and restrictions for each mission.

G.1. Situation

Current situation

The military unit you are assigned to is part of a European force: EU Force. EU force is assigned to maintain a zone of separation between Redland and Blueland. The region has had years of ethnic and religious tensions, which led to violence regularly. The situation has cooled down since the EU force intervened by establishing a 15-25 km Zone of Separation between the two countries.





Enemy intentions

Enemy parties

There are several tensions between Blueland and Redland. Militias in Blueland undermine the authority of the central government. Some of these militias keep claiming their former territory in Redland and carry out cross-border armed operations, including guerrilla raids and terrorist attacks on civilian targets.

The government of Redland is frustrated by Blueland's inability to control their land and to deal with the militias and extremists. This resulted in small invasions of Blueland by Redland's military forces to neutralize the threats a couple of times.

The Blueland militias attempt to protect and cover their assets and activities by misleading the EU force and by providing misleading disinformation

Both extremist groups are likely to be hostile towards EU forces and their patrols. They attempt to create the conditions for attacks on the other land and they will monitor the activities of EU force and their patrol patterns closely. They might target EU force personnel with VOIED (victim operated improvised explosive device) attacks, especially at roads to/from the Zone of Separation, when EU force is interfering too much in their business.

EU Force role

The EU maintains the zone of separation, prevents militia border crossings from Blueland to Redland and an invasion from Redland to Blueland. The objective of the EU is to monitor both sides of the border to detect and deter potential threats to the stability.

G.2. Missions

G.2.1. Mission 1

PARTIAL OPERATION ORDER Nr 2016-0301 of Commander 72 INFBAT 36INFCOY

Timezone: B

1. Situation: Based on intelligence: Redland military forces possibly hide in the Zone of Separation to gather intelligence about the Blueland Extremists. The negotiations with Blueland and Redland start soon, so we need to find proof for their illegal presence as soon as possible.
2. Mission: Execute a reconnaissance in GreenCity and in the outer region of GreenCity at the indicated locations in order to find hiding Redland military forces.
3. Commander's intent: This is an Overt operation with a moderate-high time pressure. All assets may be used at once, but try to keep them safe from risk.
4. Appendix G (Combat engineer advisor) summary: We expect the soldiers to use improvised explosive devices to guard the location of Person Recce Forest S. Specifically UGV3 is extremely vulnerable to the expected type of VOIED. Besides, the location of Person Recce Forest S is harmful for UAV2 due to electric wires.
5. Restrictions: Do not execute Person Recce Forest S with UGV3 and UAV2.

Commander 72INFBAT 36INFCOY

Captain
R. Jones

G.2.2. Mission 2

PARTIAL OPERATION ORDER Nr 2016-0302 of Commander 72 INFBAT 36INFCOY

Timezone: B

1. Situation: Soldiers of Redland have been spotted in the Zone of Separation by your successful reconnaissance of Building G8, well done! EU Force enforced Redland to retreat it's soldiers, we are going to monitor the proper execution of the retreat and gather intelligence about their troops secretly.
2. Mission: Execute covert reconnaissances to control the retreat of Redland soldier's and to gather intelligence about their troops.
3. Commander's intent: This is a Covert operation with moderate-low time pressure. All assets may be used at once, but do not take risk.
4. Appendix D (Military analysts) summary: The Redland forces might be aggressive towards unmanned vehicles, keep your vehicles safe.
5. Appendix G (Combat engineer advisor) summary: The terrain around Person Recce Forest N is not passable for small wheeled vehicles, the vehicle will get stuck, so do not enter this area with UGV2.
6. Restrictions: Do not execute Person Recce Forest N with UGV-2.

Commander 72INFBAT 36INFCOY

Captain
R. Jones

G.2.3. Mission 3

PARTIAL OPERATION ORDER Nr 2016-0303 of Commander 72 INFBAT 36INFCOY

Timezone: B

1. Situation: Soldiers of Redland retreated according to plan. The Blueland extremists are angry the EU Force let this happen and threaten with hit and run actions. Several safehouses of the extremists are identified from intelligence, we are going to gather more intelligence about these houses.
2. Mission: Execute a massive overt reconnaissance on possible safe houses of the Blueland' extremists in order to gather information about the presence of extremists.
3. Commander's intent: This is an Overt operation with average-low time pressure. Use as many assets as needed, but do not take much risk (so average-low).
4. Appendix D (Military analysts) summary: The extremists will be aggressive towards all EU forces, do not take much risk.
5. Appendix G (Combat engineer advisor) summary: Entering the boobytrapped Building E17 with UGV2 will likely result in the loss of this vehicle.
6. Restrictions: Do not execute Recce Building E17 with UGV2.

Commander 72INFBAT 36INFCOY

Captain
R. Jones

G.2.4. Mission 4

PARTIAL OPERATION ORDER Nr 2016-0304 of Commander 72 INFBAT 36INFCOY

Timezone: B

1. Situation: The safehouse reconnaissances led to valuable information. Now the EU Force is going to take counter measures by entering these houses in BlueTown and arrest extremists. Just before the action, we will do a reconnaissance at several other locations to ensure these locations are not hostile, furthermore it will distract the enemy.
2. Mission: Execute reconnaissance at indicated locations in order to identify possible Blueland' extremists and to distract the enemy.
3. Commander's intent: This is an Overt operation with very high time pressure. Use as many assets as possible, some risk is allowed to support the attack of the EU forces.
4. Appendix D (Military analysts) summary: The vehicles may drag enemy attention and likely will attract fire. You may take some risk. UAV2 might provoke enemy fire that is dangerous for the environment, do not use UAV2 at tasks with civilians living near.
5. Restrictions: Do not use UAV2 for Vehicle Recce S.

Commander 72INFBAT 36INFCOY

Captain
R. Jones

□

G.2.5. Mission 5

PARTIAL OPERATION ORDER Nr 2016-0305 of Commander 72 INFBAT 36INFCOY

Timezone: B

1. Situation: Redland extremists heard about the situation with Blueland extremists and want to take counter measures. Therefore they infiltrated the Zone of Separation. EU Force is threatened by their presence and aims to find these extremists soon.
2. Mission: Execute reconnaissance at indicated locations in Middleland in order to find Redland' extremists and to check enemy presence on the routes of approach.
3. Commander's intent: This is an Overt operation with a moderate-high time pressure. All assets may be used at once, but try to keep them safe from risk.
4. Appendix D (Military analyst) summary: The use of UAV2 for building Recce C3 would cause to much risk for the environment, since it may attract fire.
5. Appendix G (Combat engineer advisor) summary: The extremists use a VOIED around Person Recce S, do not use UGV3 for this reconnaissance.
6. Restrictions: Do not execute Recce Building C3 with UAV2 and Person Recce S with UGV3.

Commander 72INFBAT 36INFCOY

Captain
R. Jones

G.2.6. Mission 6

PARTIAL OPERATION ORDER Nr 2016-0306 of Commander 72 INFBAT 36INFCOY

Timezone: B

1. Situation: Your reconnaissance led to valuable information about the enemy. EU Force is going to search and arrest armed extremists in a large scale operation, the UVs are deployed to create situational understanding in a quick manner without warning them.
2. Mission: Execute reconnaissances in Middleland in order to create situational understanding to support the attack of EU Force without warning the extremists.
3. Commander's intent: This is a Covert operation with moderate-high time pressure. Try to use all assets at once, moderate-high risk is acceptable.
4. Appendix G (Combat engineer advisor) summary: The roads in the south west are too wet to enter for UGV2.
5. Restrictions: Do not execute Person Recce SW with UGV2.

Commander 72INFBAT 36INFCOY

Captain
R. Jones

□

PARTIAL OPERATION ORDER Nr 2016-0307 of Commander 72 INFBAT 36INFCOY

Timezone: B

1. Situation: The EU Force successfully arrested several important Redland extremists. The prisoners gave information about several strongholds they hold in Red village. Since this village is located within the Zone of Separation, we want more information about these strongholds.
2. Mission: Execute reconnaissances on suspected strongholds and suspected getaway cars in order to prepare another intervention by EU Force.
3. Commander's intent: This is an Overt operation with average-low time pressure. Use as many assets as needed, but do not take much risk (so average-low).
4. Appendix G (Combat engineer advisor) summary: The terrain around building G11 is surrounded by canals, which are impassable for UGV2.
5. Restrictions: Do not execute Recce Building G11 with UGV2.

Commander 72INFBAT 36INFCOY

Captain
R. Jones

□

PARTIAL OPERATION ORDER Nr 2016-0308 of Commander 72 INFBAT 36INFCOY

Timezone: B

1. Situation: Redland' extremists are present in large numbers in Red village, EU Force is preparing an attack on the village to arrest 2 persons who are amongst their top 10 leaders.
2. Mission: Identify the presence of enemy in several buildings and identify several indicated vehicles.
3. Commander's intent: This is an Overt operation with very high time pressure. Use as many assets as possible, some risk is allowed to support the attack of the EU forces.
4. Military advisor: Using UAV2 for Recce Building E8 could lead to enemy fire in the direction of civilians, do not use UAV2 for these tasks.
5. Restrictions: Do not execute Recce Building E8 with UAV2.

Commander 72INFBAT 36INFCOY

Captain
R. Jones

H

Specification

H.1. Use Cases

Use cases define the structure of the actors' actions and provide the (formal) contextualization (who, when, where). The activities defined in the design scenario can be described in four use cases: *initial information*, *constraints on task allocation*, *first schedule proposal* and *second schedule proposal*.

H.1.1. Scheduling process without additional transparency

The four use cases for the scenario without additional transparency are described below.

Phase 2a (activities #1-2): initial information

Description	eSupport shows all relevant information about the operation.
Actors	The operator and eSupport.
Preconditions:	1. The contextual information is available to eSupport, including enemy, missions, the tasks for the UV-team and the available assets.
Post conditions:	1. The operator understands the mission and the context. 2. eSupport knows the intent of the operator.
Action sequence:	#1. The operator sees the mission, the tasks for the UV-team and an overview of the available assets on the screen. #2. The operator specifies their intent for the operation.
Requirements:	RQ0001, RQ0002, RQ0003, RQ0005, RQ0006, RQ0007, RQ0011

Phase 2a (activities #3-4): Constraints on task allocation

Description	Operator sets constraints for the allocation of tasks to UVs.
Actors	The operator and eSupport.
Preconditions:	1. The operator understands the mission and the context. 2. eSupport knows the intent of the operator.
Post conditions:	1. eSupport knows the operator's constraints for the task allocation to UVs.
Action sequence:	#3. The operator starts the planning tool and does not want the person identification to be done by a certain UV, so indicates this constraint to eSupport. #4. The operator requests a schedule.
Requirements:	RQ0001, RQ0002, RQ0005, RQ0006, RQ0007, RQ0011

Phase 2a (activities #5-6): First schedule proposal

Description	Operator sets constraints about order of tasks.
Actors	The operator and eSupport.
Preconditions:	1. eSupport knows the operator's constraints for the task allocation to UVs. 2. The operator requests a schedule.
Post conditions:	1. eSupport shows a schedule that satisfies allocation constraints. 2. The operator set constraints for the order of tasks.
Action sequence:	#5. eSupport computes the most optimal task allocation that satisfies the constraint and shows the schedule to the operator in a time line. In this time line all assets are shown with their dedicated tasks and the travel time. eSupport also shows the task allocation on the map. #6. The operator sets a constraint on the building search to take place not earlier than the area reconnaissance has finished.
Requirements:	RQ0001, RQ0002, RQ0003, RQ0004, RQ0005, RQ0006, RQ0007, RQ0011

Phase 2a (activities #7-8): Second schedule proposal

Description	Schedule satisfies task order constraints
Actors	The operator and eSupport.
Preconditions:	1. The operator set the constraints for the order of tasks.
Post conditions:	1. The operator is satisfied with the proposed schedule that satisfies the constraints.
Action sequence:	#7. eSupport shows an adjusted schedule that satisfies the constraints. #8. The operator accepts the new schedule and eSupport sends the orders to execute tasks to the agent in control of the UVs.
Requirements:	RQ0001, RQ0002, RQ0003, RQ0004, RQ0005, RQ0006, RQ0007, RQ0008, RQ0011

H.1.2. Scheduling process with additional transparency

The four use cases for the scenario with additional transparency are described below.

Phase 2b (activities #1-2): initial information (with additional transparency)

Description	eSupport shows all relevant information about the operation.
Actors	The operator and eSupport.
Preconditions:	1. The contextual information is available to eSupport, including enemy, missions, the tasks for the UV-team and the available assets.
Post conditions:	1. The operator understands the mission and the context. 2. eSupport knows the intent of the operator.
Action sequence:	#1. The operator sees the mission, the tasks for the UV-team and an overview of the available assets on the screen. #2. The operator specifies their intent for the operation.
Requirements:	RQ0001, RQ0002, RQ0003, RQ0005, RQ0006, RQ0007, RQ0011

Phase 2b (activities #3-4): Constraints on task allocation (with additional transparency)

Description	Operator sets constraints for the allocation of tasks to UVs.
Actors	The operator and eSupport.
Preconditions:	1. The operator understands the mission and the context. 2. eSupport knows the intent of the operator.
Post conditions:	1. eSupport knows the operator's constraints for the task allocation to UVs.
Action sequence:	#3. The operator starts the planning tool and does not want the person identification to be done by a certain UV, so indicates this constraint to eSupport. #4. The operator requests a schedule.
Requirements:	RQ0001, RQ0002, RQ0005, RQ0006, RQ0007, RQ0011

Phase 2b (activities #5-6): First schedule proposal (with additional transparency)

Description	Operator sets constraints about order of tasks.
Actors	The operator and eSupport.
Preconditions:	1. eSupport knows the operator's constraints for the task allocation to UVs. 2. The operator requests a schedule.
Post conditions:	1. The operator understands the proposed optimal schedule from eSupport that satisfies allocation constraints. 2. The operator set the constraints for the order of tasks.
Action sequence:	#5. eSupport computes the most optimal task allocation that satisfies the constraint and shows the schedule to the operator in a time line. In this time line all assets are shown with their dedicated tasks and the travel time. eSupport also shows the task allocation on the map and the schedule's compliance with the operator's intent. #6. The operator sets a constraint on the building search to take place not earlier than the area reconnaissance has finished.
Requirements:	RQ0001, RQ0002, RQ0003, RQ0004, RQ0005, RQ0006, RQ0007, RQ0009, RQ0010, RQ0011

Phase 2b (activities #7-8): Second schedule proposal (with additional transparency)

Description	Schedule satisfies task order constraints
Actors	The operator and eSupport.
Preconditions:	1. The operator set the constraints for the order of tasks.
Post conditions:	1. The operator understands the consequences of the constraints on the order of tasks. 2. The operator has sufficient understanding of the reasoning behind the proposed schedule, that satisfies the constraints on task allocation and the order of tasks. 3. The operator is satisfied with the proposed schedule that satisfies the constraints.
Action sequence:	#7. eSupport shows an adjusted schedule that satisfies the new constraints and shows the schedule's compliance with the operator's intent. #8. The operator accepts the new schedule and eSupport sends the schedule to the agent in control of the UVs.
Requirements:	RQ0001, RQ0002, RQ0003, RQ0004, RQ0005, RQ0006, RQ0007, RQ0008, RQ0009, RQ0010, RQ0011

H.2. Claims

1. RQ0001 results in **effectiveness**: eSupport produces a schedule that complies with the user's objectives, therefore it enables the user to work more effectively
   Effect of additional transparency: It does not affect the schedule and therefore does not affect effectiveness.
   Positive: The user gets the expected result and does not have to correct the result.
   Negative: The user might want to force the automation too much towards its expected result.

2. RQ0002, RQ0003, RQ0004, RQ0005, RQ0006, RQ0007, RQ0009, RQ0010 result in **efficiency**: eSupport provides appropriate information at the right level of abstraction to minimize the required scheduling time and effort while maintaining the performance of the user.
   Effect of additional transparency: It improves the user's efficiency by providing appropriate information about the compliance of their objectives at the right level of abstraction to decrease the required time.
   Positive: The user requires less time and effort.
   Negative: The user might not see important information.

3. RQ0003, RQ0004, RQ0005, RQ0006, RQ0009, RQ0010 result in **situational awareness**: The user has sufficient knowledge and understanding about the environmental elements with respect to time or space.
   Effect of additional transparency: It improves the user's SA by providing information about the consequences of the schedule.
   Positive: The user is able to make fast decisions based on the most up-to-date information of the environment.
   Negative: The user might experience cognitive overload when too much information is provided, resulting in worse performance.

4. RQ0002, RQ0003, RQ0004, RQ0007, RQ0009, RQ0010, RQ0011 result in balanced **cognitive task load**: eSupport provides the user with the right amount of information at the appropriate abstraction level. This leads to a balanced amount of mental processes that does not lead to under- or overload of the user.
   Effect of additional transparency: It does decrease cognitive task load in order to reduce the scheduling time and support the user's focus on the task to improve appropriate (non-)compliance, by providing the right information that supports inspecting the automation decisions.
   Positive: The user experiences an optimal cognitive task load, this supports concentration and performance.
   Negative: eSupport might not perform optimally due to the limitation of cognitive capabilities of humans.

5. RQ0008 results in **decision authority**: The user has the final right to decide and can always overrule eSupport. eSupport does not start execution of the schedule before the user gave permission.
   Effect of additional transparency: It does not affect decision authority.
   Positive: This supports safety and the responsibility is clear.
   Negative: eSupport might not perform optimally when it has to wait for approval.

6. RQ0001, RQ0004, RQ0010, RQ0011 result in calibrated **trust**: eSupport gives the right information about the reliance of a suggested schedule to the user to calibrate the user's trust. This should prevent overtrust, which results in too much compliance of the user to the suggestions of eSupport, and undertrust, which results in overly controlling eSupport by the user.
   Effect of additional transparency: It improves calibrated trust by improving the user's understanding of the consequences of the schedule related to the user's objectives.
   Positive: An appropriate level of trust supports the user to actually use the automation with appropriate reliance in eSupport's reasoning. This leads to more efficient human-automation cooperation since both human's and automation's skills are utilized.
   Negative: Calibrated trust requires provision of information to the user about the automation's reasoning and reliability, this might have negative effects such as to cognitive task load and efficiency.

7. RQ0001, RQ0002, RQ0004, RQ0009, RQ0010, RQ0011 result in **appropriate (non-)compliance**: eSupport gives the right information about the performance of the schedule to the user and about the weak spots of eSupport. This leads to an increase of appropriate (non-)compliance.
   Effect of additional transparency: It improves the understanding of the user about the consequences of the schedule and the weak spots of eSupport, this should result in better estimation about the correctness of a schedule.
   Positive: The user does know the weak spots of eSupport and can estimate when it makes a mistake.
   Negative: The user does only check for weak spots of eSupport and misses other important information.

Documents for participants

I.1. Questionnaire

I.1.1. Demographic survey

1. Participant number:
2. Age:
3. Gender:
4. What region do you come from (circle one):

   Europe *North America* *Central/South America* *Sub-Saharan Africa*

   Middle East and North Africa *Asia and Pacific* *Not provided*
5. How do you rate your english reading skills (circle one):

   beginner *average* *good* *very good* *excellent*
6. Occupation:

   if student (circle one): Bachelor Master PhD
7. Military experience (circle one): Yes No

   if yes, what service:

   Years of service:
8. How often do you watch TV/series (circle one)?

   Never *Rarely* *Once a month* *Weekly* *A few times per week* *Daily*
9. How often do you play video games (circle one)?

   Never *Rarely* *Once a month* *Weekly* *A few times per week* *Daily*

   Types of games played:
10. Rate your comfort level with using computers (circle one):

    Not comfortable *Somewhat comfortable* *Comfortable* *Very comfortable*
11. What is your perception towards Unmanned Vehicles? (circle one):

    Intense dislike *Dislike* *Neutral* *Like* *Really like*
12. What is your perception towards the use of unarmed Unmanned Vehicles for military operations? (circle one):

    Intense dislike *Dislike* *Neutral* *Like* *Really like*

I.1.2. Attentional control

Below is a list of statements for evaluating attentional control (Derryberry and Reed, 2002). Please select for each question the value that best indicates how you feel about the statement.

	Almost never	Some-times	Often	Always	No opinion
It's very hard for me to concentrate on a difficult task when there are noises around.	O	O	O	O	O
When I need to concentrate and solve a problem, I have trouble focusing my attention.	O	O	O	O	O
When I am working hard on something, I still get distracted by events around me.	O	O	O	O	O
My concentration is good even if there is music in the room around me.	O	O	O	O	O
When concentrating, I can focus my attention so that I become unaware of what's going on in the room around me.	O	O	O	O	O
When I am reading or studying, I am easily distracted if there are people talking in the same room.	O	O	O	O	O
When trying to focus my attention on something, I have difficulty blocking out distracting thoughts.	O	O	O	O	O
I have a hard time concentrating when I'm excited about something.	O	O	O	O	O
When concentrating I ignore feelings of hunger or thirst.	O	O	O	O	O
I can quickly switch from one task to another.	O	O	O	O	O
It takes me a while to get really involved in a new task.	O	O	O	O	O
It is difficult for me to coordinate my attention between the listening and writing required when taking notes during lectures.	O	O	O	O	O
I can become interested in a new topic very quickly when I need to.	O	O	O	O	O
It is easy for me to read or write while I'm also talking on the phone.	O	O	O	O	O
I have trouble carrying on two conversations at once.	O	O	O	O	O
I have a hard time coming up with new ideas quickly.	O	O	O	O	O
After being interrupted or distracted, I can easily shift my attention back to what I was doing before.	O	O	O	O	O
When a distracting thought comes to mind, it is easy for me to shift my attention away from it.	O	O	O	O	O
It is easy for me to alternate between two different tasks.	O	O	O	O	O
It is hard for me to break from one way of thinking about something and look at it from another point of view.	O	O	O	O	O

I.1.3. Trust

Below is a list of statement for evaluating trust between people and automation. Please select for each question the intensity that best describes your feeling or your impression. If you desire you can leave additional comments per question.

1. **I am suspicious of the system's output**

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Extremely

Comments:

2. **The execution of the schedule will have a harmful or injurious outcome for the environment**

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Extremely

Comments:

3. **I am confident in the system**

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Extremely

Comments:

4. **The system is reliable**

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Extremely

Comments:

5. **I can trust the system**

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Extremely

Comments:

6. **I am familiar with the system**

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Extremely

Comments:

I.1.4. Situational awareness

For each of the questions below, please encircle the letter in front of the correct answer.

1. Task # at the example map represents:

a A building reconnaissance

b A person reconnaissance

c A vehicle reconnaissance

2. Task # at the example map represents:

a A building reconnaissance

b A person reconnaissance

c A vehicle reconnaissance

3. Why was task # not assigned to asset #?:

a Due to low risk utility

b Due to low time utility

c Due to low sensor utility

d It would lead to low massiveness utility

4. Task # is executed by asset:

a UAV1

b UAV2

c UGV1

d UGV2

e UGV3

5. What asset will enter building #?:

a UAV1

b UAV2

c UGV1

d UGV2

e UGV3

6. Is there risk of a loss of an asset due to the schedule?:

- yes, for 3 assets
- yes, for 2 assets
- yes, for 1 asset
- no

7. **What threat is located at task #?:**

a Rifle

b VOIED

c Sniper

d RPG

8. **What asset is exposed to the rifle threat?:**

a UAV1

b UAV2

c UGV1

d UGV2

e UGV3

9. **How fast does asset # finish its tasks?:**

a In 2 hours

b In 1,5 hours

c In 1 hour

d In 30 minutes

e In < 20 minutes

10. **Indicate the location of task # (see map)?:**

- A
- D
- F
- G
- H

11. **How fast are all tasks completed?:**

a In 2 hours

b In 1,5 hours

c In 1 hour

d In 30 minutes

e In < 20 minutes

12. **How many assets are used for the mission?:**

a 5

b 4

c 3

d 2

e 1

13. At what location does asset # start (see map)?:

- A
- B
- C
- E
- G

14. What threat is present at location # (see map)?:

a Rifle

b VOIED

c Sniper

d RPG

e None

I.1.5. Cognitive task load

The measurement technique and the description below is derived from (NASA, 1986).

The evaluation you are about to perform is a technique that has been developed by NASA to assess the relative importance of six factors determining how much workload you experienced. The following variables are used:

Table I.1: Rating Scale Definitions (NASA, 1986)

Title	Description
Mental Demand	How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?
Physical Demand	How much physical activity was required? Was the task easy or demanding, slack or strenuous?
Temporal Demand	How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?
Performance	How successful were you in performing the task? How satisfied were you with your performance?
Frustration Level	How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?

Please select for each question the value that best indicates your experience of the task.

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure I.1: NASA Task Load Index Questions

I.1.6. Usability

This questionnaire is derived from (Brooke et al., 1996).

Please select for each question the value that best indicates your experience, try to fill in your immediate response and not to think about an item for a long time.

1. I think that I would like to use this system frequently for the task

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

2. I found the system unnecessarily complex

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

3. I thought the system was easy to use

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

4. I think that I would need the support of a technical person to be able to use this system

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

5. I found the various functions in this system were well integrated

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

6. I thought there was too much inconsistency in this system

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

7. I would imagine that most people would learn to use this system very quickly

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

8. I found the system very cumbersome to use

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

9. I felt very confident using the system

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

10. I needed to learn a lot of things before I could get going with this system

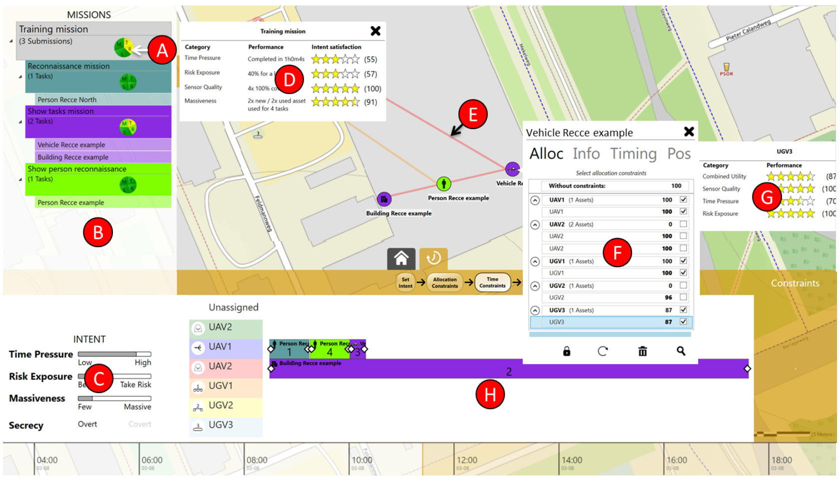
	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

11. I found the system was fast enough

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:



12. Write the letters corresponding to each of the following interface elements (see also the figure above) in order of **most helpful** for the scheduling **to least helpful**:

    A The utility icon (pie chart)
    B The missions overview
    C The intent overview
    D The mission utility details
    E The utility overview of assets per task
    F The task-asset utility details
    G The schedule overview

    Answer:

    Comments:

13. I understood better why this schedule was proposed due to the pie chart icons (A) and the mission utility details (D)

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

14. The pie chart icons (A) and mission utility details (D) made the system too complex for me

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Comments:

15. Were there interface elements that were not useful? If yes, which?
    Answer:

    Comment:

16. Were there interface elements that were crucial? If yes, which?
    Answer:

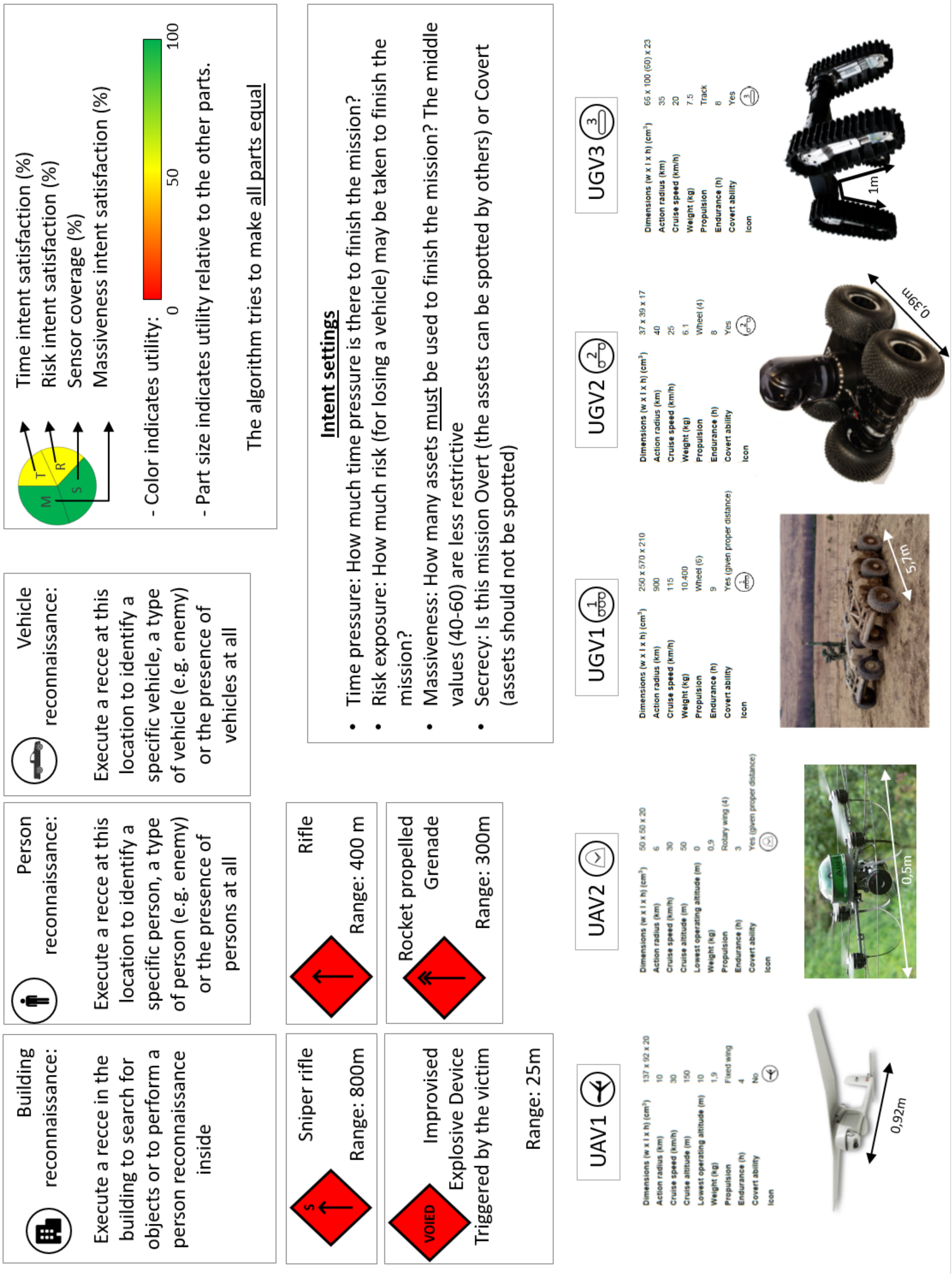
    Comment:

17. Were there times that your trust in the system increased or decreased? Why?
    Answer:

    Comment:

18. Do you have other comments?

I.2. Supportive sheet



All military symbols are based on standard NATO Military symbols (Landmachtstaf. Directie Beleid en Plannen. Afdeling Doctrine, Opleiding en Training, 2007).

J Overview of correlations

The correlations between variables are investigated to get more understanding of the relations between variables. Some correlations are described using Pearson's r when we deal with interval data. However, when the values contain a large number of tied ranks, so a specific value is present in large numbers, or when we deal with ordinal data, Kendall's τ is used.

The tables below show all the significant and almost significant correlations we found between independent and dependent variables. The almost significant correlations are included, since these might give additional understanding of partial relations between variables. **Table J.1** shows the correlations between demographic information and correlated variables. **Table J.2** shows the correlations between the other variables. A schematic overview of the correlations is provided in Figure J.1 to support better understanding. Note that we cannot draw conclusion about causality from the correlations.

Correlations with demographic information and experimental conditions

Table J.1: Correlations with $p < .09$ between demographic information and other measures. The correlation coefficient (τ/r), the BCa 95% Confidence Interval and the significance are shown. Double entries are removed

	Video game experience	Gender	Attitude to UVs	Attitude to military Uvs	Appropriate compliance	Attentional control
Video game experience		$\tau = -.61$ [-.799, -.337] $p = .001$	$\tau = .37$ [-.032, .672] $p = .037$	$\tau = .38$ [-.031, .654] $p = .050$		
Gender			$\tau = -.59$ [-.786, -.348] $p = .003$	$\tau = -.46$ [-.715, -.119] $p = .029$	$\tau = -.34$ [-.599, -.026] $p = .026$	
Attitude towards UVs				$\tau = .32$ [.077, .682] $p = .036$		
Duration	$\tau = -.21$ [-.459, .033] $p = .061$					
Cognitive task load		$\tau = .23$ [-.048, .496] $p = .07$	$\tau = -.37$ [-.568, -.138] $p = .002$			$\tau = .28$ [.120, .440] $p = .008$
Trust		$\tau = -.41$ [-.585, -.211] $p = .002$	$\tau = .28$ [.028, .510] $p = .024$			

Significant correlations between measures

Table J.2: Correlations with p < .09 between measures. The correlation coefficient (τ/r), the BCa 95% confidence interval and the significance are shown. Double entries are removed

	Cognitive task load	Trust	Situational awareness	Usability of eSupport	Usefulness of transparency	Attitude to UVs
Trust	r = -.36 [-.602, -.049] p = .015					
Mission Duration	r = .39 [.128, .642] p = .010		r = .39 [.114, .617] p = .009			
Appropriate compliance	τ = -.29 [-.504, .048] p = .024	τ = .30 [.032, .523] p = .021		τ = .24 [-.013, .467] p = .064		τ = .28 [-.036, .561] p = .053
Appropriate non compliance	τ = .33 [.125, .504] p = .011			τ = -.28 [-.474, -.067] p = .030		
Usability of eSupport (transformed)	r = -.352 [-.699, .024] p = .019	r = .392 [.115, .629] p = .009				
Usage of transparency information				τ = .34 [.021, .643] p = .033	τ = .66 [.423, .856] p < .001	
Mission block with transparency					τ = .369 [.124, .577] p = .008	

Schematic overview of correlations

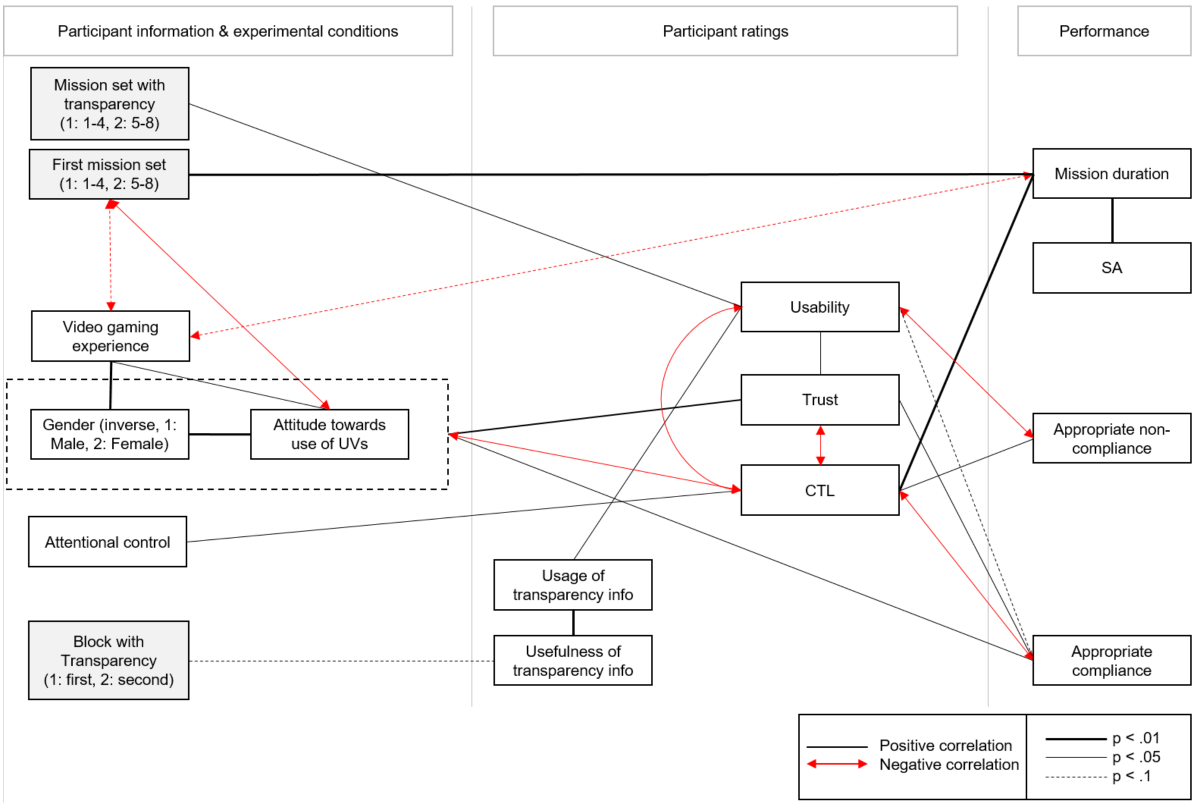


Figure J.1: Overview of the correlations

Bibliography

Baker, C., Naikar, N., and Neerincx, M. Engineering planetary exploration systems: integrating novel technologies and the human element using work domain analysis. In *Proceedings of the 59th International Astronautical Congress (IAC2008)*, 2008.

Barnes, M. J., Chen, J. Y., Jentsch, F., Oron-Gilad, T., Redden, E., Elliott, L., and Evans III, A. W. Designing for humans in autonomous systems: Military applications. Technical report, DTIC Document, 2014.

Barnes, M. and Evans III, A. Soldier-robot teams in future battlefields: An overview. *Human-robot interactions in future military operations*, pages 9–29, 2010.

Bennett, K. B., Posey, S. M., and Shattuck, L. G. Ecological interface design for military command and control. *Journal of Cognitive Engineering and Decision Making*, 2(4):349–385, 2008.

Benyon, D. Designing interactive systems: a comprehensive guide to hci and interaction design . cf: 8 envisonment, 2010.

Brooke, J. et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

Brouwer, P. and Scholten, M., M. *Joint Doctrine Publication 2 Intelligence*. Doctrine Division of the Netherlands Defense Staff, 2012.

Carey, S., Kleiner, M., Hieb, M., and Brown, R. Standardizing battle management language–a vital move towards the army transformation. In *IEEE Fall Simulation Interoperability Workshop, Orlando, FL*, 2001.

Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., and Barnes, M. Situation awareness-based agent transparency. Technical report, DTIC Document, 2014.

Chen, J. Y. and Barnes, M. J. Supervisory control of multiple robots effects of imperfect automation and individual differences. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(2):157–174, 2012a.

Chen, J. Y. and Barnes, M. J. Supervisory control of multiple robots in dynamic tasking environments. *Ergonomics*, 55(9):1043–1058, 2012b.

Chen, J. Y., Haas, E. C., and Barnes, M. J. Human performance issues and user interface design for teleoperated robots. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1231–1245, 2007.

Chen, J. Y., Barnes, M. J., and Harper-Sciarini, M. Supervisory control of multiple robots: Human-performance issues and user-interface design. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 41(4):435–454, 2011.

Clare, A. S. Modeling real-time human-automation collaborative scheduling of unmanned vehicles. Technical report, DTIC Document, 2013.

Clare, A. S., Cummings, M. L., and Bertuccelli, L. F. Identifying suitable algorithms for human-computer collaborative scheduling of multiple unmanned vehicles. In *AIAA Aerospace Sciences Meeting, Nashville, TN*, 2012a.

Clare, A. S., Cummings, M. L., How, J. P., Whitten, A. K., and Toupet, O. Operator object function guidance for a real-time unmanned vehicle scheduling algorithm. *Journal of Aerospace Computing, Information, and Communication*, 9(4):161–173, 2012b.

Clare, A. S., Macbeth, J. C., and Cummings, M. L. Mixed-initiative strategies for real-time scheduling of multiple unmanned vehicles. In *American Control Conference (ACC), 2012*, pages 676–682. IEEE, 2012c.

Clare, A. S., Cummings, M. L., and Repenning, N. P. Influencing trust for human-automation collaborative scheduling of multiple unmanned vehicles. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(7):1208–1218, 2015.

Cummings, M. L. and Bruni, S. Human-automation collaboration in complex multivariate resource allocation decision support systems. 2010.

Cummings, M. L., Clare, A., and Hart, C. The role of human-automation consensus in multiple unmanned vehicle scheduling. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 2010.

Cummings, M. L., Marquez, J., and Roy, N. Human-automated path planning optimization and decision support. *International Journal of Human-Computer Studies*, 70(2):116–128, 2012.

Cummings, M., Brzezinski, A., and Lee, J. D. The impact of intelligent aiding for multiple unmanned aerial vehicle schedule management. 2007.

de Greef, T., Mohabir, A., van der Poel, I., and Neerincx, M. scethics: embedding ethical values in cognitive engineering. In *proceedings of the 31st European conference on cognitive ergonomics*, page 4. ACM, 2013.

De Winter, J. Controversy in human factors constructs and the explosive use of the nasa-tlx: a measurement perspective. *Cognition, technology & work*, 16(3):289–297, 2014.

Derryberry, D. and Reed, M. A. Anxiety-related attentional biases and their regulation by attentional control. *Journal of abnormal psychology*, 111(2):225, 2002.

Dutch Royal Army Doctrine Commitee. *Defense Doctrine*. 2013.

Dutch Royal Army Doctrine Commitee. *Doctrine Publication 3.2 Land Operations*. 2014.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., and Anderson, B. W. Predicting misuse and disuse of combat identification systems. *Military Psychology*, 13(3):147, 2001.

Ekstrom, R. B., French, J. W., Harman, H. H., and Dermen, D. *Manual for kit of factor referenced cognitive tests*. Educational Testing Service Princeton, NJ, 1976.

Endsley, M. R. Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):65–84, 1995.

Endsley, M. R. Direct measurement of situation awareness: Validity and use of sagat. *Situation awareness analysis and measurement*, 10, 2000.

Endsley, M. R. *Designing for situation awareness: An approach to user-centered design*. CRC Press, 2011.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., and Croft, D. G. A comparative analysis of sagat and sart for evaluations of situation awareness. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 42, pages 82–86. SAGE Publications, 1998.

Endsley, M. R., Bolstad, C. A., Jones, D. G., and Riley, J. M. Situation awareness oriented design: from user's cognitive requirements to creating effective supporting technologies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 47, pages 268–272. SAGE Publications, 2003.

Friedman, B., Kahn Jr, P. H., Borning, A., and Huldtgren, A. Value sensitive design and information systems. In *Early engagement and new technologies: Opening up the laboratory*, pages 55–95. Springer, 2013.

Ghallab, M., Nau, D., and Traverso, P. *Automated planning: theory & practice*. Elsevier, 2004.

Gmelich Meijling, J. Battlefield Management Systeem. `https://www.rijksoverheid.nl/documenten/kamerstukken/1997/12/09/battlefield-management-systeem`, december 1997. Accessed: November 12, 2015.

Harbers, M., De Greeff, J., Kruijff-Korbayová, I., Neerincx, M., and Hindriks, K. Exploring the ethical landscape of robot-assisted search and rescue. 2015.

Hart, S. G. and Staveland, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

Hayes, A. F. An analytical primer and computational tool for observed variable moderation, mediation, and conditional process modeling. *Manuscript submitted for publication*, 2012.

Hazen, M. G. and Randall, T. E. Using modelling and simulation to represent commander's intent in the maritime context. In *Proceedings of the 2008 Summer Computer Simulation Conference*, page 22. Society for Modeling & Simulation International, 2008.

Hieb, M. and Schade, U. Formalizing command intent through development of a command and control grammar (i-069). In *12th International Command and Control Research and Technology Symposium*, 2007.

Jenkins, D. P. Using cognitive work analysis to describe the role of uavs in military operations. *Theoretical Issues in Ergonomics Science*, 13(3):335–357, 2012.

Jentsch, F. and Fincannon, T. Multiple robot-multiple operator control and teamwork. In *Proceedings of the Information Technology Conference of the American Institute for Aeronautics and Astronautics ("Infotech AIAA")*, 2012.

Jian, J.-Y., Bisantz, A. M., and Drury, C. G. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1):53–71, 2000.

Landmachtstaf. Directie Beleid en Plannen. Afdeling Doctrine, Opleiding en Training. *Handboek Staftechniek*. Landmachtstaf. Directie Beleid en Plannen. Afdeling Doctrine, Opleiding en Training, 2007.

Linegang, M. P., Stoner, H. A., Patterson, M. J., Seppelt, B. D., Hoffman, J. D., Crittendon, Z. B., and Lee, J. D. Human-automation collaboration in dynamic mission planning: A challenge requiring an ecological approach. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 50, pages 2482–2486. SAGE Publications, 2006.

Martinez, S. G., Talcott, C., Bennett, K. B., Stansifer, C., and Shattuck, L. G. Cognitive systems engineering analyses for army tactical operations. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 45, pages 523–526. SAGE Publications, 2001.

Mercado, J. E., Rupp, M. A., Chen, J. Y., Barnes, M. J., Barber, D., and Procci, K. Intelligent agent transparency in human–agent teaming for multi-uxv management. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58(3):401–415, 2016.

Miller, C., Funk, H., Wu, P., Goldman, R., Meisner, J., and Chapman, M. The playbook™ approach to adaptive automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 15–19. SAGE Publications, 2005.

Murphy, R. and Burke, J. The safe human-robot ratio. *Human-Robot Interactions in Future Military Operations, Human Factors in Defence*, pages 31–49, 2010.

NASA. Nasa task load index (tlx) v. 1.0 manual. `humansystems.arc.nasa.gov/groups/TLX/downloads/TLX.pdf`, 1986. Accessed: May 26, 2016.

Neerincx, M. A. Cognitive task load analysis: allocating tasks and designing support. *Handbook of cognitive task design*, 2003:283–305, 2003.

Neerincx, M. A. and Lindenberg, J. Situated cognitive engineering for complex task environments. *Naturalistic decision making and macrocognition*, pages 373–390, 2008.

Neerincx, M. A., Kennedie, S., Grootjen, M., and Grootjen, F. Modeling the cognitive task load and performance of naval operators. In *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*, pages 260–269. Springer, 2009.

Noël, C., J. A conceptual basis for the land operations of finabel countries - doctrine land operations manual. www.finabel.org/studies, june 2009. Accessed: July 7, 2015.

Office Secretary of Defense. unmanned aircraft systems roadmap 2005-2030. https://fas.org/irp/program/collect/uav_roadmap2005.pdf, 2005. Accessed: October 15, 2015.

Parasuraman, R. and Riley, V. Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2):230–253, 1997.

Parasuraman, R., Galster, S., and Miller, C. Human control of multiple robots in the roboflag simulation environment. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 4, pages 3232–3237. IEEE, 2003.

Rasmussen, J. and Lind, M. Coping with complexity. Technical report, 1981.

Ridder de, Cap. Iedereen hetzelfde plaatje. https://magazines.defensie.nl/landmacht/2015/08/elias, augustus 2015. Accessed: November 12, 2015.

Salmon, P. M., Stanton, N. A., Walker, G. H., Baber, C., Jenkins, D. P., McMaster, R., and Young, M. S. What really is going on? review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science*, 9(4):297–323, 2008.

Shattuck, L. G. and Woods, D. D. Communication of intent in military command and control systems. In *The Human in Command*, pages 279–291. Springer, 2000.

Stam, M. Presentation: Robots beslissen in het gevecht, bouwen aan veiligheid?! www2.wisernl.com/KIVI-DV.../Bert_Stam_-_Robots_beslissen_het_gevecht.pdf, 2009. Accessed: October 15, 2015.

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M. S., et al. Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics*, 49(12-13):1288–1311, 2006.

Streefkerk, J. W., Smets, N., Varkevisser, M., and Hiemstra-van Mastrigt, S. Future command and control systems should combine decision support and personalization interface features. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, pages 266–275. ACM, 2014.

TNO. Applying the situated cognitive engineering method - a comprehensive guide. http://scetool.nl/sce_manual_v1.0.pdf, 2012. Accessed: August 18, 2016.

US Department of Defense. Dictionary of Military Terms. http://www.dtic.mil/doctrine/dod_dictionary, november 2010. Accessed: November 17, 2015.

Whelan, R. Effective analysis of reaction time data. *The Psychological Record*, 58(3):475, 2008.