Master's Thesis in Cognitive Science

Performance and Shared Understanding in Mixed C^2-systems

Department of Computer and Information Science

Linköping University

Author: Erik Prytz

Advisors: Peter Berggren, Swedish Defense Research Agency
Björn Johansson, Linköping University

ISRN: LIU-IDA/KOGVET--10/009--SE

Abstract

OBJECTIVE: This thesis had two purposes. The main one was to examine how mixed conditions affect a Command & Control (C^2) system, particularly in terms of shared understanding, situation awareness (SA), performance and workload. Mixed conditions refer here to when subsystems of a larger C^2-system differ in terms of capabilities, particularly those capabilities influencing the understanding of a situation e.g. sensors or communication, which could affect the C^2-capabilities when working toward a common goal. The second purpose of this thesis was to investigate a newly developed tool for measuring shared understanding, Shared Priorities, in terms of validity and usefulness.

METHOD: A number of hypotheses were constructed and investigated by a controlled experiment using a microworld, C3Fire, where two-man teams fought a simulated forest fire. The independent variable manipulated was the type of support system used. One condition used one computer interface per participant, the second was mixed conditions where one participant used the computer interface and one used a paper map, and the last condition was both participants using a paper map. Different questionnaires developed to measure SA, workload etc. was used to measure the dependent variables.

RESULTS: The statistical analysis performed on the collected data showed that the performance and SA was comparatively better when both participants used the computer interface than the mixed condition, which in turn was better than when both participants used a paper map. For workload and teamwork, no differences between the mixed condition and the dual map condition were found. As for the Shared Priorities measurement, no differences were found between any of the conditions.

CONCLUSION: A C^2-system in which some additional capabilities are introduced for some but not all subsystems may not benefit in some regards, e.g. workload and teamwork, but could improve in others, e.g. SA and performance. A Structural Equation Model (SEM) shows that the theoretical constructs of SA, workload, teamwork and performance are related and affect each other, so that the workload of the system negatively affects the teamwork and SA, while the teamwork may affect SA positively and a high SA enables high performance.

Preface

During my work with this thesis I have had the fortune to collaborate with many talented individuals, some whom I owe many thanks. First and foremost my two advisors, Peter Berggren and Björn Johansson, whose experience and knowledge have been of tremendous help in steering this thesis on its course. Peter's practical it-should-work approach coupled with Björn's more theoretical reasoning has more than anything proved to me that it can be hard to synthesize different operational pictures from such mixed conditions!

Staffan Nählinder supported me with much needed advice on statistics and data management (and who taught me the truth of the saying "have data, need results"), and Erland Svensson gladly jumped in to share his substantial experience and expertise on the dark wizardry that is Structural Equation Modelling. Per-Anders Oskarsson helpfully filled in the gaps in my knowledge and patiently answered my questions.

C3Fire is a complex microworld with substantial potential. Rego Granlund is the world's foremost expert (and also sole developer) of this microworld, so his aid in configuring and teaching me how to set up and calibrate the experiment was very welcome. I am also grateful for all the participants who signed up for the experiment and gave of their time. Of course, the experiments could not have been run without the two operators, Fredrik Höglund and Sandra Jonsson, nor would they have been as fun to run with anyone else. They did a, so to speak, fanta-stic job.

I would also like to thank my friends who have, sometimes forcefully, parted me from my work to reconnect with reality and life alike.

Last, Jill for her patience and never ending support. And proofreading.

Table of contents

Abstract I

Preface III

Table of contents V

Tables & Figures VIII

Terminology and Abbreviations X

1 *Introduction* 1

1.1 Purpose 2

1.1.1 Research Questions 2

1.1.2 Hypotheses 2

1.2 Delimitations 3

2 *Theory* 5

2.1 Command & Control 5

2.1.1 The OODA and the DOODA Loop 6

2.2 Cognitive Systems Engineering 8

2.2.1 JCS 9

2.2.2 COCOM & ECOM 10

2.3 Microworlds 12

2.3.1 Methodological Issues 12

2.3.2 C3Fire 13

2.4 Situation Awareness 14

2.4.1 Critique of Situation Awareness 15

2.4.2 Endsley's Three-Step Model 16

2.4.2.1 Three-Step Model & DOODA 17

2.4.3 Situation Awareness Measurement 19

2.4.3.1 CARS 20

2.4.4 Methodological Issues 20

2.5 Mental Workload 21

2.6 Shared Understanding in Teams 22

2.6.1 Shared Priorities 22

2.6.2 Questionnaires for Distributed Assessment of Team Mutual Awareness 23

2.6.2.1 Taskwork Awareness 23

2.6.2.2 Workload Awareness 24

2.6.2.3 Teamwork Awareness 24

2.6.3 Methodological Issues 25

2.7 Summary 25

3 *Method* 27

3.1 Experimental design 27

3.1.1 Scenario Setup 29

3.1.1.1 Condition 1 30

3.1.1.2 Condition 2 32
3.1.1.3 Condition 3 33
3.1.2 Pilot Study 33
3.1.3 Participants 34
3.1.4 Apparatus 35
3.1.4.1 Hardware 35
3.1.4.2 C3Fire Version 35
3.1.4.3 Shared Priorities program 35
3.1.4.4 Material 35
3.1.4.5 Communication Scheme 35
3.1.5 Procedure 37

3.2 Dependent Measures 41
3.2.1 Performance Measure 41
3.2.2 Shared Priorities 41
3.2.3 Situation Awareness: Modifications to and Measures in CARS 41
3.2.4 Workload, Teamwork & Shared Understanding: Modifications to and Measures in DATMA 42

4 Results 45

4.1 Data Overview 45
4.1.1 Missing values 45
4.1.2 Fire, Task and Position Awareness 46
4.1.3 Individual Workload 49
4.1.4 Team Workload 51
4.1.5 Teamwork 53
4.1.6 CARS 56
4.1.7 C3Fire Performance 58

4.2 Shared Understanding Results 59
4.2.1 Hypothesis 1: Shared Understanding by Shared Priorities 59
4.2.2 Hypothesis 2: Shared Understanding by DATMA 60

4.3 Effects of Mixed Conditions 61
4.3.1 The PCA 61
4.3.2 Factorial Split-Plot ANOVAs on Factors 63
4.3.2.1 Hypothesis 3: Workload 64
4.3.2.2 Hypothesis 4: SA 65
4.3.2.3 Hypothesis 5: Teamwork 66
4.3.2.4 Hypothesis 6: Performance 67
4.3.3 External Factors 68

4.4 Relations between Factors 69
4.4.1 What is SEM? 69
4.4.2 The SEM Analysis 71

5 Analysis/Discussion 74

5.1 Method Discussion 74
5.1.1 Task, Fire and Position Awareness 74
5.1.2 The PCA 75
5.1.3 The effect of roles 76
5.1.4 Paper Map Training Effects 76

5.2 Result Analysis 77
5.2.1 Research Question 1 77

5.2.2 Research Question 2 78
5.2.3 Research Question 3 79
5.2.4 Research Question 4 80

5.3 Conclusions & Future Works 82

6 Bibliography 83

Appendix A – Questionnaires 1

Appendix B – Correlation Matrix 7

Tables & Figures

Figure 2-1 The OODA-Loop....7
Figure 2-2 The DOODA-Loop (adapted from Brehmer 2006) 8
Figure 2-3 Joint Cognitive System (From Hollnagel & Woods 2005) 9
Figure 2-4 COCOM (from Hollnagel & Woods 2005)....10
Figure 2-5 ECOM (from Hollnagel & Woods 2005)....11
Figure 2-6 Endsley's Three Step Model (Adapted from Endsley 1995)....17
Figure 2-7 DOODA and the Three Step Model....18
Figure 2-8 CARS Questions (adapted from McGuinness 1999) 20
Figure 3-1 C3Fire Map 29
Figure 3-2 GUI/GUI Setup....30
Figure 3-3 C3Fire GUI Layout....31
Figure 3-4 Information Window....31
Figure 3-5 GUI/Map Setup 32
Figure 3-6 Map/Map Setup 33
Figure 3-7 Physical Setup 38
Figure 3-8 Two Images of the two trained operators. 39
Figure 3-9 The interface of the operators.. 39
Figure 3-10 Images of participants. 40
Figure 4-1 Fire Awareness Result 47
Figure 4-2 Task Score....48
Figure 4-3 Position Score....49
Figure 4-4 Individual Workload Example Question....49
Figure 4-5 Individual Workload Result Overview 50
Figure 4-6 Individual Total Workload....51
Figure 4-7 Team Workload Example Question....51
Figure 4-8 Team Workload Result Overview 52
Figure 4-9 Team Total Workload....53
Figure 4-10 Teamwork Example Question 54
Figure 4-11 Teamwork Result Overview....55
Figure 4-12 Teamwork....56
Figure 4-13 CARS Example Question....56
Figure 4-14 CARS Result Overview 57
Figure 4-15 CARS....58
Figure 4-16 Performance Overview....59
Figure 4-17 Shared Priorities 60
Figure 4-18 PCA Factors 62
Figure 4-19 PCA Factors Overview 64
Figure 4-20 Workload Factor....65
Figure 4-21 SA Factor 66
Figure 4-22 Teamwork Factor....67
Figure 4-23 Performance Factor....68
Figure 4-24 External Factors Factor 69
Figure 4-25 Structural Equation Model 73
Figure 5-1 Experiment setup as DOODA with multiple JCSs 81

Table I Abbreviations X
Table 1-1 Hypotheses 3
Table 2-1 Inclusion of Objects in a JCS (Adapted from Hollnagel & Woods 2005) 10
Table 3-1 Balancing over Time 28
Table 3-2 Communication Scheme: Command 36
Table 3-3 Communication Scheme: Events 36
Table 3-4 Communication Scheme: Questions 36
Table 3-5 Task Categories 43
Table 3-6 Combined Task Categories 44
Table 4-1 Fire Awareness Scoring 46
Table 4-2 Individual Workload Questions 50
Table 4-3 Team Workload Questions 52
Table 4-4 Teamwork Questions 54
Table 4-5 CARS Questions 57
Table 4-6 Hypothesis 1 Answered 60
Table 4-7 Hypothesis 2 Answered 61
Table 4-8 Hypothesis 3 Answered 64
Table 4-9 Hypothesis 4 Answered 65
Table 4-10 Hypothesis 5 Answered 66
Table 4-11 Hypothesis 6 Answered 67
Table 4-12 SEM Fit Indexes 72

Terminology and Abbreviations

Table I Abbreviations

Abbreviation	Meaning	Explanation
AGFI	Adjusted Goodness of Fit	A statistical term
ANOVA	Analysis of Variance	A statistical test
C^2	Command & Control	A theoretical concept and research domain
C3	Communication, Command and Control	A theoretical concept
C3FIRE	Communication, Command and Control FIRE	A Computer-based simulation software
C3I	Communication, Command, Control and Intelligence	A theoretical concept
CARS	Crew Awareness Rating Scale	A method for measuring SA
CFI	Comparative Fit Index	A statistical term
COCOM	Contextual Control Model	A CSE model
COP	Common Operational Picture	A C2-term, in this thesis also a variable in CARS
CSE	Cognitive Systems Engineering	A research domain
CV90	Combat Vehicle 90	A vehicle type in the Swedish army
DATMA	Questionnaires for Distributed Assessment of Team Mutual Awareness	A method for measuring Mutual Awareness
DOODA	Dynamic OODA	A C^2-model
DSA	Distributed Situation Awareness	A theoretical concept
ECOM	Extended Control Model	A CSE model
FOW	Fog of War	A C^2 term
FRAM	Functional Resonance Accident Model	A method for accident analysis
GFI	Goodness of Fit	A statistical term
GIS	Geographic Information System	A system that analyzes and presents geographic data
GUI	Graphical User Interface	A user interface for e.g. computers
HF	Human Factors	A research domain
HQ	Headquarters	A military term for the location from which military systems are controlled
IFI	Incremental Fit Index	A statistical term
IQR	Inter-Quartile Range	A statistical term
JCS	Joint Cognitive system	A CSE concept
KMO	Keisser-Meyer-Olkin	A statistical test
LISREL	Analysis of Linear Structural Relationships	A statistical analysis software for SEM
LOS	Line of Sight	A C^2 term
N	Number of scores in data	A statistical term
NASA	National Aeronautics and Space Administration	A U.S. government agency

NASA TLX	NASA Task Load Index	A method for measuring workload
NATO	North Atlantic Treaty Organisation	A military alliance
NFI	Normed Fit Index	A statistical term
NNFI	Non-Normed Fit Index	A statistical term
OODA	Observe, Orient, Decide, Act	A theoretical model
PCA	Principal Component Analysis	A statistical term
RMSEA	Root Mean Square error of Approximation	A statistical term
SA	Situation Awareness	A theoretical concept and research domain within HF
RADAR	Radio Detection and Ranging	A type of sensor
SABARS	Situation Awareness Behavioural Rating Scale	A method for measuring SA
SAGAT	Situation Awareness Global Assessment Technique	A method for measuring SA
SASHA	Situation Awareness for SHAPE	A method for measuring SA
SEM	Structural Equation Modelling or Structural Equation Model	A statistical test
SHAPE	Solutions for Human-Automation Partnerships in European ATM (Air Traffic Management)	A European research project for ATM
SME	Subject Matter Expert	A person with expert knowledge in a certain domain
SP	Shared Priorities	A method for measuring shared understanding
SRMR	Standardized Root Mean Square Residual	A statistical term

1 Introduction

This thesis will investigate the effects of mixed conditions on the functioning of a Command & Control (C^2) system. Mixed conditions mean that the different parts of the system have different abilities and possibilities to work toward the system goal. Why is this important for C^2-systems? This is a situation that is common in different military settings today. It could be differences between different military organizations that are trying to cooperate on foreign missions, different branches of the armed forces, or even within the same branch.

Take, for example, the case of the Swedish army's main battle tank Leopard 122 that is currently in operation and the Combat Vehicle 90, CV90. The Leopard 122 has an upgraded, electronic fire control system for target marking, etc. that also can be linked to a rear HQ where a commander can coordinate the units based on the information retrieved from the system. The CV90 however, lacks this modern equipment and the operators have to use a paper map, a pen and a radio to communicate similar information to the higher command, sometimes while performing in the same operations as the Leopard 122s. The higher command then, has to coordinate and synthesize these different operational pictures received from the units with electronic map and those reported via radio and subsequently marked down on the commander's own paper map.

The units and the commander are parts of the same system, trying to work toward a common goal, but differ in their ability, in terms of C^2, to do so. Each type of vehicle has advantages and drawbacks, and they are made to perform different parts of a mission, but one concern is the coordination between them. The commander in the example is blind to the actual situation and receives different feedback of the same situation from his different units. Will this difference negatively affect the system performance to the degree that introducing certain novel functions or abilities in only parts of the system (e.g. only equipping the Leopard 122 with the electronic system and not also the CV90) actually might do more harm than good? Will the mixed conditions create a problem when forming the common operating picture or situation awareness necessary for a cohesive and coordinated effort by the system?

In order to explore this further, this thesis used a simulated C^2-system consisting of two human operators, or commanders, fighting a fire in the microworld C3Fire. The system was manipulated so that in one condition both commanders used a computer interface (Graphical User Interface; GUI) to control their units, but in a second condition only one of the commanders used the computer interface while the other had a paper map. A third condition had both commanders using a paper map. The thesis will explore how these manipulations are reflected in system performance and other measures. The other measures selected were situation awareness and workload, as these two concepts have previously been shown to covariate with performance (Nählinder, Berggren & Svensson 2004). Further, this thesis will also attempt to validate a new measurement for shared understanding in teams, called Shared Priorities. This measurement and other similar shared understanding, or mutual awareness[1],

[1] The terms shared understanding and mutual awareness will be used interchangeably in this thesis.

measures will be used when looking at the differences between systems with same or mixed conditions.

This thesis will be structured in the following way: A more formal definition of the purpose and both the hypotheses and research questions of this thesis will follow in the next section. In the next chapter, theories that are relevant for this purpose and study are presented, explaining more precisely what a C^2-system is and how this can be investigated in terms of situation awareness and shared understanding using a microworld. The method chapter will provide a detailed description of how the study was done and what the dependent measures were. In the following result section a brief overview of the data is first presented which is complemented with more in depth statistical analyses. Finally, there is a discussion of the relevance and meaning of the findings, summarized conclusions as well as potential future works.

1.1 Purpose

The purpose of this thesis is twofold. The first and main purpose aims to examine how mixed conditions affect a C2-system, particularly in terms of shared understanding, situation awareness, performance and workload. Secondly, this thesis also attempts to validate a recently developed shared awareness measurement called Shared Priorities by examining how well it measures differences in conditions and how it relates to the other measures used in this study.

1.1.1 Research Questions

There are some larger, overarching research questions that are relevant for this thesis. Some will be specified further into narrower hypotheses. The questions are:

1. How will mixed conditions affect the shared understanding in the system?
2. How well will Shared Priorities be able to measure shared understanding?
3. How will mixed conditions affect the mental workload, SA, teamwork, and performance of the system?
4. How are the concepts used to investigate the effects of mixed conditions (SA workload, performance, shared understanding) related, and how do they affect each other?

These research questions are not as structured as the hypotheses and they will be investigated in a more exploratory way. They are guiding questions for interpreting the results from the study and for placing the conclusions from the hypotheses in a larger picture.

1.1.2 Hypotheses

6 formal hypotheses are formed, based on some of the research questions, to investigate the purposes of this study. Hypotheses 1 and 2 are based on research question 1, while hypotheses 3-6 are based on research question 3. The hypotheses are presented in table 1-1.

Table 1-1 Hypotheses

Hypothesis 1: Mixed conditions will have an impact on the shared understanding of the team as measured by Shared Priorities
H_0: There is no difference in the Shared Priorities measure between conditions.
H_a: There is a difference in the Shared Priorities measure between all three conditions.
H_{a2}: There is a difference in the Shared Priorities measure between some but not all three conditions.
Hypothesis 2: Mixed conditions will affect team mutual awareness as measured by DATMA.
H_0: There is no difference in team mutual awareness as measured by DATMA.
H_a: There is a difference in team mutual awareness as measured by DATMA in all three conditions.
H_{a2}: There is a difference in team mutual awareness as measured by DATMA in some but not all of the three conditions.
Hypothesis 3: Mixed conditions will affect mental workload.
H_0: There is no difference in subject and team mental workload between any of the conditions.
H_a: There is a difference in subject and team mental workload in all three conditions.
H_{a2}: There is a difference in subject and team mental workload in some but not all of the three conditions.
Hypothesis 4: Mixed conditions will affect Situation Awareness (SA).
H_0: There is no difference in SA between any of the three conditions.
H_a: There is a difference in SA between all the three conditions.
H_{a2}: There is a difference in SA between some but not all of the three conditions.
Hypothesis 5: Mixed conditions will affect teamwork.
H_0: There is no difference in teamwork between any of the conditions.
H_a: There is a difference in teamwork in all three conditions.
H_{a2}: There is a difference in teamwork in some but not all of the three conditions.
Hypothesis 6: Mixed conditions will affect performance in C3Fire.
H_0: There is no difference in performance between any of the three conditions.
H_a: There is a difference in performance between all three conditions.
H_{a2}: There is a difference in performance between some but not all of the three conditions.

1.2 Delimitations

A microworld study has the advantage in terms of ease of data collection: log files with time stamped messages between participants, replayable scenarios with every single action done by all participants recorded, etc. While this may provide a rich and valuable source of data, it may also be overwhelming and detract from the actual purpose of the study. Therefore the data analysis in this thesis work will be limited to the data collected from distributed questionnaires and one measure of performance from the microworld. While I acknowledge the importance of communication analysis for a fuller understanding of participants' motives and teamwork, it is beyond the scope of this thesis.

This study is operating within a command and control paradigm, more specifically a small team (2 members) cooperating via either a graphical user interface, a paper map or a combination thereof with the common goal of extinguishing forest fires in a microworld. Studies in the command and control paradigm are notoriously hard to generalize over larger populations due to the differences in context, culture and other factors, even more so when conducting microworld studies using student participants (as in this one) without formal training in command and control situations and procedures. As such, the generalization of conclusions from this study must be made with an awareness of the limits in ecological validity and pursued with care and caution.

2 Theory

This chapter provides a theoretical background to the thesis. The chapter explores the theoretical frameworks, concepts, and measurements that are used to investigate the purpose of this study (the effect of mixed conditions on military systems).

First in this chapter Command & Control (C^2) is introduced and the definition(s) of the term is discussed. The main focus is military systems, and it is argued that C^2 is a function of such a system. A cybernetic model, the DOODA-loop (Dynamic Observe-Orient-Decide-Act loop) is introduced to further elaborate on this point. As C^2 is viewed as a function performed by a military system, a brief section on Cognitive Systems Engineering (CSE) as a theoretical framework for reasoning around functions and complex systems is provided following the C^2-section.

The section after CSE introduces the research tool called “microworlds” as a way of simulating these complex systems in a controlled way. A microworld study allows the researcher to control and replicate the experiment so that the variables of interest can be measured reliably. With Nählinder, Berggren & Svensson (2004) in mind, the methods used for measuring in this study are based on the SA, workload and shared understanding theoretical frameworks. Therefore SA is introduced as a concept following the section on microworlds. A well-known SA model, Endsley’s three-step model, is explained and it is shown how this model can be applied to C^2-systems, using the DOODA-loop, theories and terminology from the C^2-domain.

After the section on SA, workload is introduced as a related concept that is one influencing factor on system performance and SA, which makes it relevant for the purpose of this study. Shared understanding is more than simply SA and workload, so a section on shared understanding in teams further discuss what this term entails. This section also introduces the Shared Priorities measure that this thesis attempts to validate. Last in this chapter is a summary that further ties the different concepts and theories introduced together.

2.1 Command & Control

> “The Organisation, Process, Procedures and Systems necessary to allow timely political and military decisionmaking and to enable military commanders to direct and control military forces”
> (A definition of C^2, in NATO 1996 quoted in NATO Code of Best Practice for C^2 Assessment, 2002, p.2)

Command & Control (C^2) is an ambiguous term of many meanings and definitions. Above is (one of) NATO’s definition(s) of what C^2 is, but it is far from the only one. So what is C^2?Andriole & Halpin (1986) state that there are as many answers as there are questioners but that it for many it means “weapons control via communications technology”. Lawson (1981) however, believes that the terms “command control” and “command control system” mean “whatever the speaker wants them to mean” but most commonly “some form of computer complex which presumably ‘processess’ information and presents it to a ‘decision maker’ for his use” (Lawson 1981, p. 5).

NATO's definition is more recent, and it is reflected in the definition which includes four different concepts (organisation, process, procedures and systems) rather than simply "weapons control" or similar, and that these should "allow for timely decisions" as well as "enable the direction and control of military forces" which goes beyond simply presenting information to a decision maker as per Lawson's definition. Even still, Pigeau & McCann (2002) try to re-conceptualize C^2 due to what they call the "confusing complexity" with which the term is used; they also state the NATO definition is both redundant and circular.

Brehmer, citing Van Creveld, states that C^2 is a "*function* of the military system" (Van Creveld 1985, in Brehmer 2007, p. 212, italics in original) that is essential for creating military effects. He also notes that C^2 is always performed within a C4ISR system (Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance). This somewhat cumbersome and all-encompassing acronym has been gradually developed for many years; Andriole & Halpin (1986) remarked dryly on the development from C^2 to C3 (Command, Control and Communications) and then C3I (Command, Control, Communications and Intelligence) and evidently three more letters have been added to the acronym since 1986, perhaps reflecting advances in C^2 research (or perhaps only the C^2 science community's fascination with acronyms).

What most definitions do agree on is that C^2 concerns the process or function of directing or controlling (military) systems to achieve some specific effects, and this is what is important. A C^2-system would then be the system that can perform this function, or process. C^2-systems are today often defined as complex socio-technical systems (Riley et al. 2006) rather than earlier views of a C^2-system as a "computer complex" as Lawson (1981) described it. This change in definition likely reflects an acknowledgement of decisionmaking and cognition as something that occurs in a system (see e.g. Hansberger et al. 2008), also known as distributed cognition (Hutchins 1995). This connection and theoretical framework is further discussed in the summary section (2.7 Summary) at the end of this chapter.

2.1.1 The OODA and the DOODA Loop

One of the most famous C^2-models that have influenced the domain is the OODA-loop (Boyd, 1987). It is interesting to note that Boyd never published a book, journal article or any of the sorts explaining his reasoning, yet the OODA-loop has received much attention. The OODA stands for Observe – Orient – Decide – Act (see figure 2-1) and the most commonly used metaphor to describe it is that of a fighter pilot engaging in a dog fight with a hostile aircraft (Boyd himself was Colonel in the US Air force and an experienced fighter pilot). In fact, the model was originally created by Boyd to explain the success of American fighter pilots against Korean fighter pilots during the Korean War (Brehmer 2005).

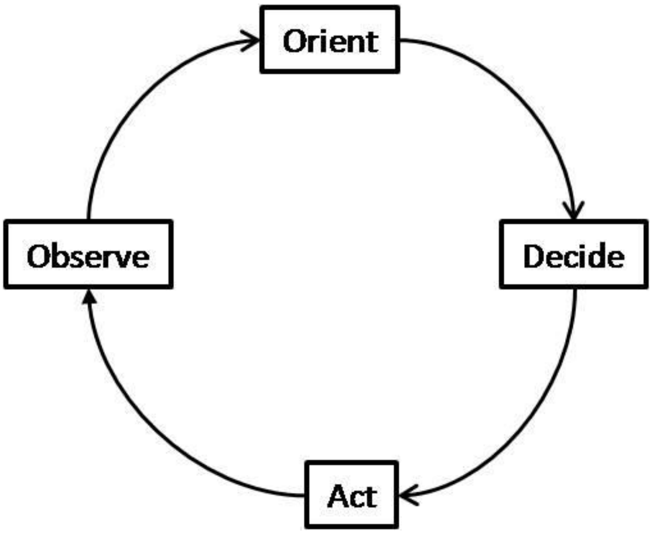


Figure 2-1 The OODA-Loop

In order for the pilot to win the dog fight, he would have to “get inside” the enemy’s OODA-loop, meaning that he would have to observe faster than the adversary, i.e. spot the enemy aircraft before they spotted him, orient the aircraft toward the enemy in an attack position before the enemy could align his, decide what action to take and then act, which would be the act of engaging the enemy aircraft. This model was later extended by Boyd, and others, to include more general situations, mostly by re-purposing the “Orient” stage to a mental orientation rather than a physical (Brehmer 2005). The OODA-loop has since been applied in many domains and in many forms, often with little or no actual connection to what Boyd originally intended with the loop (Brehmer 2008).

The DOODA-loop (Brehmer 2005) uses a cybernetic approach and takes a functional perspective on C^2-systems (Brehmer 2006) to provide a more robust and general model of C^2-systems. There are three basic functions that are seen as key to the overall functioning, and definition, of the system. Those three are information collection, sensemaking and planning. Figure 2-2 shows the relationship between these functions as they relate to the OODA-loop.

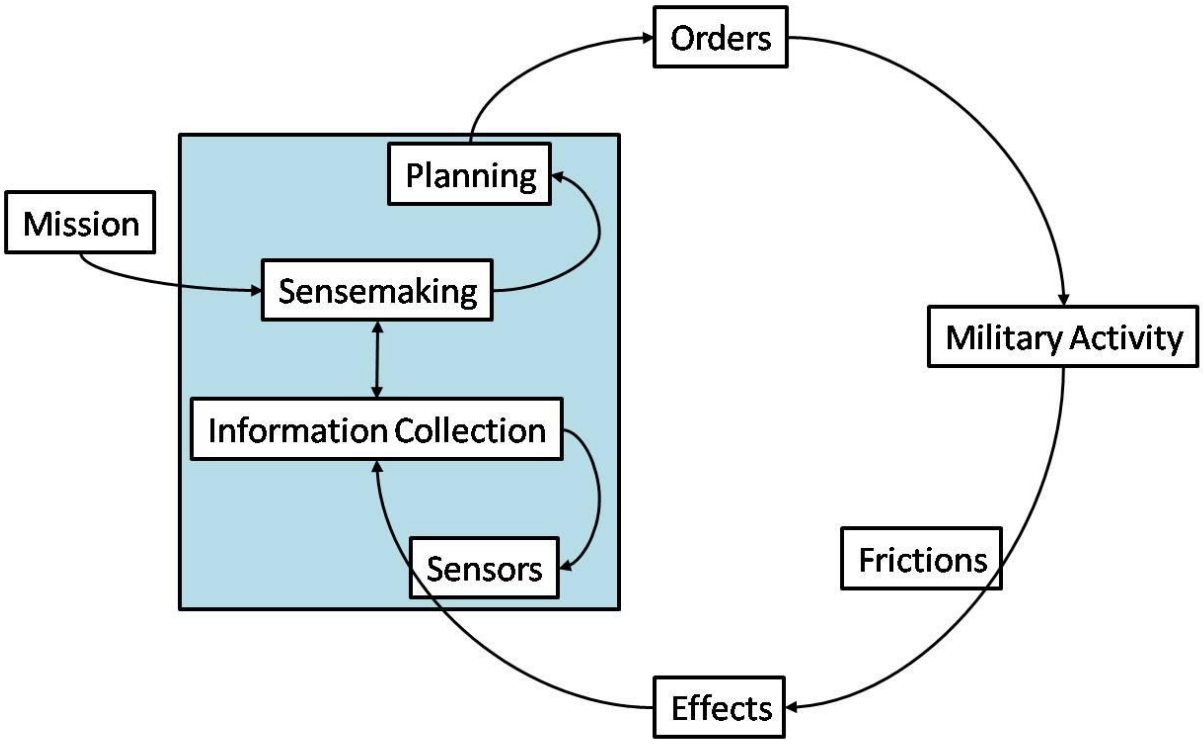


Figure 2-2 The DOODA-Loop (adapted from Brehmer 2006)

The area in blue in image 2-2 is the C^2-system in question. The information collection function receives feedback from the sensemaking function, which directs the search for information. Sensemaking in turn receives input from the mission. The sensemaking function aims to produce an understanding of the situation, in the form of "what should be done" in the current situation (Brehmer 2006). The planning function takes this understanding and produces orders, which will result in military activity. The effects are separated from the activity itself since frictions will always arise, i.e. things will not always go as planned. This is a well-known phenomenon in military theory, as evidenced by the famous quote by Helmuth Graf von Moltke that "no campaign plan survives first contact with the enemy" and the term "friction" has been commonly used since Carl von Clausewitz's "On War" was originally published in 1832.

In essence, the DOODA-loop takes the OODA-loop, which in its more general form is about individual command and control functioning, to a system level with functions as well as expanding it with more detailed functions. I will return to the DOODA-loop later in this chapter and tie it to Endsley's three-step model of Situation Awareness as a way of illustrating how one can re-conceptualize the three-step model to a system level.

2.2 Cognitive Systems Engineering

Cognitive Systems Engineering (CSE; Hollnagel & Woods 1983, 2005) is an approach related to distributed cognition (Hutchins 1995) and macro-cognition (Klein et al. 2003) for analyzing complex socio-technical systems. The perspective taken is functional, that a system is defined by the functions it performs rather than the actual construct of the system or how it performs these functions.

Two basic CSE models are the Contextual Control Model (COCOM) and the Extended Contextual Control Model (ECOM), which are concerned with the different control levels and modes at which a system operates (Hollnagel & Woods 2005), as well as more practically oriented methods such as the Functional Resonance Accident Model (FRAM), which can be used to explain how seemingly normal activity within a system may cause disturbances or collapses (Woltjer & Hollnagel 2007).

2.2.1 JCS

One often-used term and unit of analysis in CSE is Joint Cognitive Systems (JCS). The term JCS will in this thesis mostly be used when explaining different theoretical models, such as COCOM, and as a common term that connects the different theories. It will play a lesser role in the later analysis of the results of this thesis but is nonetheless seen by the author as an important concept in the CSE field as a whole.

A JCS consists of a cognitive system, which is a goal-oriented system that can adapt its behavior based on experience, together with one or more other cognitive systems or some form of artefact, either physical or social. A common example is that of a driver (a cognitive system) together with a car (an artefact) as one JCS. The boundaries of the system does not necessarily end there, but can, depending on the purpose of the analysis, be extended to include the roads, traffic infrastructure, topography and so on, se figure 2-3.

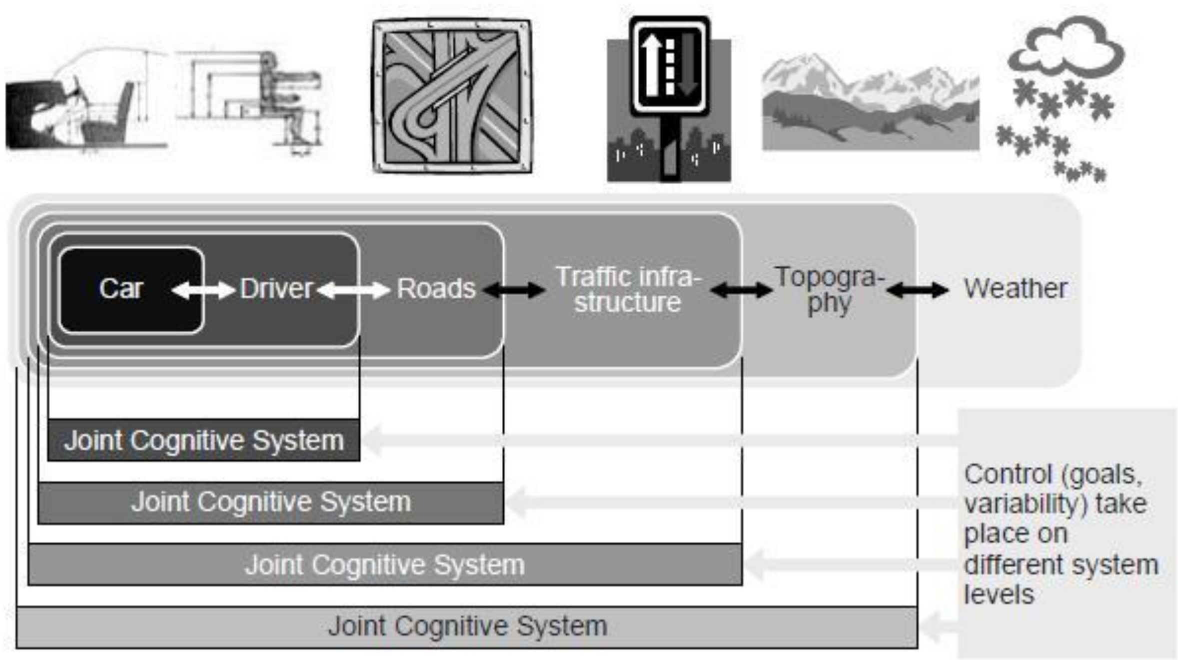


Figure 2-3 Joint Cognitive System (From Hollnagel & Woods 2005)

The boundary between what is included in the JCS and what is part of the environment is commonly focused on where the system stops being in control, but that still affects the JCS, see table 2-1 below.

Table 2-1 Inclusion of Objects in a JCS (Adapted from Hollnagel & Woods 2005)

	Objects whose functions are important for the ability of the JCS to maintain control.	Objects whose functions are of no consequence for the ability of the JCS to maintain control.
Objects that can be effectively controlled by the JCS.	1. Objects are included in the JCS	2. Objects may be included in the JCS
Objects that cannot be effectively controlled by the JCS	3. Objects are not included in the JCS	4. Objects are excluded from the description as a whole.

2.2.2 COCOM & ECOM

The Contextual Control Model, COCOM, is a cyclical model of how a JCS maintains control (figure 2-4). It is inspired by Neisser's Perceptual Cycle Model (Neisser 1976). Hollnagel and Woods (2005) argue that a cyclical model has several advantages over sequential models for studying complex systems, for instance seeing the users as parts of the whole process, the model being functional rather than structural and combining feedback with feedforward.

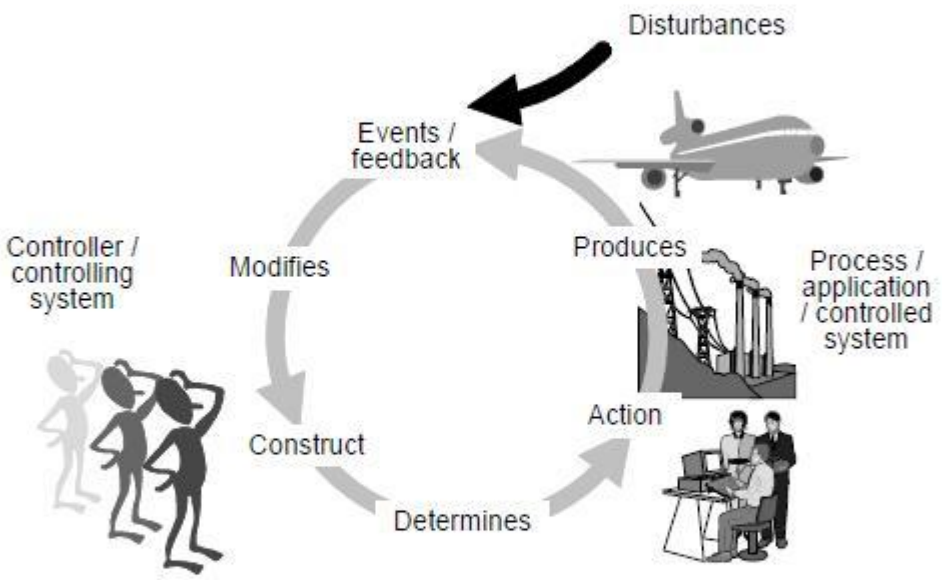


Figure 2-4 COCOM (from Hollnagel & Woods 2005)

An important distinction is made in the name, the "Contextual" control model. A contextual model is based largely on the notice that the context determines the next action, unlike a procedural model where the next action is determined based on a pre-defined pattern. Three important components of this model are *competence*, the possible actions that can be carried out by the JCS, *control*, the application of actions, and *constructs*, the understanding of the situation by the JCS (Hollnagel & Woods 2005).

The control concept concerns how orderly the application of competence (actions) is. Hollnagel & Woods (2005) divide this into four discrete steps on a continuum from disorderly to orderly. The most disorderly mode is *scrambled*, where there is essentially no control. Actions are determined mostly at random and without regards to the context. The next mode, *opportunistic*, takes some aspects of the context into account when determining the next action. It is however mostly a trial-and-error approach where many actions will be unsuccessful or useless. The *tactical* mode is when the system has more time than is needed to make a decision and the next action can be planned with the context fully in mind. The last mode is *strategic*, where higher level goals, secondary goals and long-term effects can be taken into account and actions planned out appropriately.

A system will transition between these different modes depending on many factors. Most JCSs will be going between opportunistic and tactical (ibid.), as the strategic mode is very demanding in terms of resources (for instance time). One of the key points to this model is how it explains the performance of a system. Performance, in general, can be seen as dependent on the actions of a system, and COCOM states that the construct (i.e. understanding of the current situation) is the base on which a decision is made. The understanding in turn comes from feedback from the environment where both the actions of the system and disturbances are taken into account.

Another level of complexity can be added by looking at different levels of control and performance in a system. This is the purpose of the Extended Control Model (ECOM) which is shown below in figure 2-5.

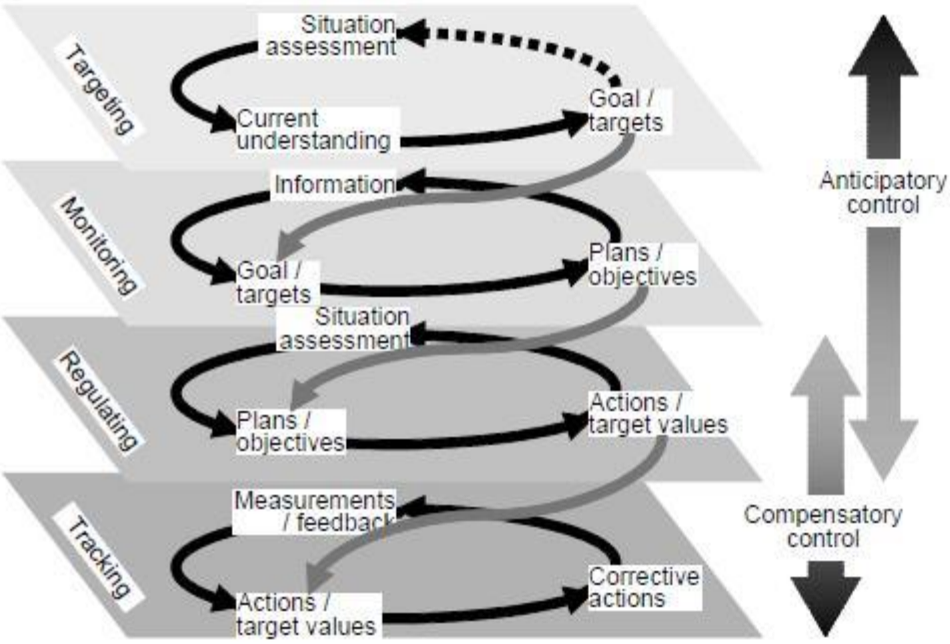


Figure 2-5 ECOM (from Hollnagel & Woods 2005)

ECOM is an extension of COCOM, and introduces parallel processes on different levels to explain how a system may have high performance in some regards and low performance in others. The illustration above is simplified and does not show all potential couplings between levels. ECOM has previously been used, among other things, to describe and understand the coordination of an emergency response in a simulated forest fire (Aminoff, Johansson & Trnka 2007).

The lowest level is *tracking*, which is on the level of basic actions of the JCS. In the example of driving a car, this would be adjusting speed and position to stay within the appropriate lane without going too fast or slow. *Regulating* on the other hand is a slightly higher level of control and can consist of several different tracking sub-loops. The regulating process is still concerned with actions, and provides the tracking layers with input in terms of actions to be performed. *Monitoring* is not directly concerned with actions but operates on a higher level. Monitoring level control refers to setting higher level goals, so while the regulating level may be concerned with avoiding obstacles in the traffic situation, the monitoring level would be planning a route to the intended destination. The highest level of control is *targeting*, and would be the level that determines the destination. This level will set goals that affect all lower level loops in a cascade effect (Hollnagel & Woods 2005; Aminoff, Johansson & Trnka 2007).

2.3 Microworlds

Studying a complex socio-technical system can be difficult. One way of simulating a complex system in a controlled way is to use microworlds. Microworlds are an experimental tool that aims to bridge the gap between real-world field studies and controlled laboratory experiments (Brehmer & Dörner 1993). It is a computer simulation of a system that is *complex*, *dynamic* and *opaque* (ibid.). It provides participants with a rich environment in which to act and react. While so called "real-world" studies, studies done in a field setting, often, if not always, lack the control needed to investigate certain issues, laboratory studies are often the direct opposite, with so many variables controlled that the result cannot be generalized outside of the laboratory setting. This trade-off between so-called internal and external validity is well-known. Brehmer & Dörner (1993) state that there is too much complexity in the real world for definite conclusions, while there is too little complexity in the laboratory for any interesting conclusions. Microworlds, while not a perfect solution, are one way of bringing complexity and a dynamic environment to the participant while still allowing the researcher to remain in control and giving the ability to replicate the experiment.

The complexity and dynamicity aside, the opaque quality of the microworlds refers to the fact that the participants cannot see directly how the microworld works, but have to create a mental model of how the microworld works. This includes, just like in the real world, guesswork, hypothesizing, heuristics and trial-and-error testing. Microworlds have also led to a new approach in the study of decision-making (Brehmer 2005a), even though microworlds normally engage many of the participants' psychological functions such as problem solving, decision making and even emotions (ibid.).

2.3.1 Methodological Issues

It is said that the best simulation of a cat would be another cat. This refers to a problem with microworlds that tries to simulate existing complex systems, which most do. While it may be desirable to raise the fidelity of the microworld to being equal or nearly equal to that of the system it is meant to simulate, this would not provide much benefit over just using the actual system in the first place. A microworld will therefore always be simpler than what it tries to simulate. So while a microworld is indeed closer to the real world than most other controlled environments, it is not the real thing. This means that there will be differences in the behavior of the participants, for example that they are willing to take greater risks than they would normally. In practice, the microworld is a world without real-life consequences.

Another problem can arise from using a microworld that simulates a system with participants who are not trained operators or regular users of the simulated system. The participants will be without the experience, training and knowledge that a skilled operator would have. This is however not unique to microworlds, but rather part of the tradition in many social sciences to use students, for example, rather than participants drawn from the intended population. To complicate things further, the reverse can also be a problem. A skilled or trained operator may already have a fixed mental model of how the system works, a model developed through training and years of experience using the real system. When confronted with a microworld that is almost but not quite the real thing, this mental model may cause difficulties in handling the microworld and make their performance on par with a naïve participant.

2.3.2 C3Fire

C3Fire is a microworld in which the user assumes the role of a firefighter fighting a forest fire. More specifically, the participant normally has a role of directing fire, water and fuel trucks to the simulated forest fires. The C3Fire microworld can be used in large teams utilizing complex command structures or be played by just a single participant. It provides a rich task environment where different types of terrain, structures and computer-simulated agents can be utilized. The graphical user interface (GUI) can be customized to great extent, commonly using a geographical information system (GIS, an interactive electronic map), a communications tool similar to email or chat-clients and different kinds of information displays, e.g. time, wind direction, fuel level of the units etc.

The first version of C3Fire was called DESSY (Brehmer 1987), which turned into NEWFIRE (Lövborg & Brehmer 1991) and later D3Fire (Brehmer & Svenmarck 1994) until landing more firmly in its current form, C3Fire (Granlund 1997) although in an early version. The three C's stand for Communication, Command, and Control. Since it has been in use, in many forms, for such a long time (for a microworld), it has also been used on many studies relating to C^2, SA, teamwork and similar concepts (e.g. Artman & Granlund 1998; Granlund 2003; Johansson et al. 2003). Artman & Granlund (1998) in particular is worth mentioning further. They investigated the difference in shared SA, using the three levels of Endsley's (1995) model, when using either a text based or a graphically based interface. The teams in their experiment consisted of two staff members who controlled two fire chiefs that interacted with C3Fire. They did not find any significant differences in performance between the two conditions, but this may have been due to the measures used than an absence of effects.

The ability to customize scenarios and GUIs in C3Fire, as well as the straight forward task (to extinguish a fire) makes it a good choice for the study in this thesis. It is also a microworld that has been used for similar purposes before, and has proven to be a reliable and useful research tool.

2.4 Situation Awareness

Situation Awareness, sometimes Situational Awareness, or SA for short, is a commonly used (some would say overused) term to describe, essentially, an operator's notion of "what is going on". It has been a hot topic ever since the term first emerged in the 1980s as something critical for pilots during the First World War (Press 1986, cited in Endsley 1995). During the '90s it received much attention from the Human Factors (HF) community and research intensified. In fact, it has been described as something of a "buzzword of the '90s" (Wiener 1993).

There are many different definitions of what SA precisely is; Salmon et al. (2009) found over 30 definitions when writing their literature analysis. Dominquez (1994) synthesized 15 definitions and arrived to the definition of SA as an individual's

> "continuous extraction of environmental information, and integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in directing future perception and anticipating future events" (Dominquez 1994, p. 11, in Salmon et al. 2009, p. 8)

In this definition, SA is seen as something inherently individual and cognitive, which has long been the predominant view. For example Hartmen & Secrist (1991) state that "situational awareness is principally (though not exclusively) cognitive, enriched by experience".

One of the most common SA definitions is that postulated by Endsley in 1995 (Wickens 2008). Endsley first defined SA as "the pilot's internal model of the world around him at any point in time" (Endsley 1988) and later clarified and extended this to:

> "Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." (Endsley 1995, p. 36)

While the definition by Dominquez (1994) can be seen as a process, Endsley (1995) defines SA as a product separate from the process. A completely different approach is taken by Smith & Hancock (1995a, 1995b) who view SA as "adaptive, externally directed consciousness that has as its products knowledge about a dynamic task environment and directed action within that environment" (Smith & Hancock, 1995b) and more formally defined as,

> "the invariant in the agent-environment system that generates the momentary knowledge and behaviour required to attain the goals specified by an arbiter of performance in the environment" (Smith & Hancock 1995a, p. 145)

This view has SA not only as a process (as Dominquez 1994) or a product (as Endsley 1995) but rather a combination of the two. In fact, this difference in viewing SA as a product or a process is one of the major differences in definitions and models (Salmon et al. 2009). Another main difference in definitions and models is how SA is scaled to more than a single individual. Just as there are many definitions of SA, there are definitions about Team SA, Shared SA and Distributed SA. Some of these issues will be discussed more under 2.6 Shared

Understanding in Teams. One thing that most definitions of SA do agree on, however, is that it concerns the dynamic awareness (of an individual) of an ongoing external situation (Salmon et al. 2009).

2.4.1 Critique of Situation Awareness

SA has a mixed reputation in the research community. As previously mentioned it was dubbed the "buzzword of the '90s" by Wiener (1993), and not in a positive way. The plethora of, often conflicting, definitions, theories and models have led some to conclude that SA is more of a folk psychology term than a scientific one, which Dekker & Hollnagel (2004) touch upon. Sarter & Woods (1995, quoted in Endsley 1995) went as far as to say that "developing a definition of SA is futile and not constructive" even though they four years earlier themselves had tried to define and use SA in system design (Sarter & Woods 1991). However, Wickens (2008) argues that the very fact that SA is so commonly used is proof of its viability and that the scientific community progresses by these exact kinds of debates.

Researchers have noted that there is a danger in using SA as a casual term, and that the scientific community should be wary of such use of the terminology (Billings 1996; Flach 1995). The problem, they argue, would be a confounding of the real causes and an easily accessible, but ultimately wrong, explanation that for example an accident was "caused" by "a loss of SA" or similar notices that ultimately do not provide any real insight to the underlying causes and problems. This is mostly a critique against how SA is used as a concept rather than a critique of the term itself.

Further critique has been leveraged against SA regarding its psychological, and biological, foundations. SA is, depending on the definition used, to a greater or lesser extent partly based on other psychological constructs, such as long term memory and mental models. Sometimes it is hard to separate these different concepts, and the lines of what should be included in the SA term and what should not are diffuse at best (Wickens 2008). Further, the theoretical underpinnings vary greatly, from Endsley (1995) who bases her model on an information processing paradigm, while Smith & Hancock (1995a) base their model on Neisser's perceptual cycle model (Neisser 1976) and others still on other underpinnings, see for example Bedney and Meister's (1999) model which uses activity theory as a base.

One last issue has created a lot of debate: ways of measuring SA. There are many methods that claim to measure SA, but few have had more than a few validation studies (Salmon et al. 2009). Problems with replication of previous studies and showing differences between experimental conditions are two other problems with some of the methods that claim to measure SA.

These different issues will be dealt with in various ways in this thesis. First of all SA will not be directly used as a sole casual term to explain performance, acknowledging the points brought up by Billings (1996) and Flach (1995). SA will be a part of many in the model later created to reason around the potential effects of mixed conditions on system functioning, and thus used cautiously with the critique levied in this section in mind. Care will be taken to use validated methods of measuring SA, as well as complementing this data with methods measuring other psychological constructs, such as workload and teamwork, to put the results into perspective. The theoretical problems of diffuse boundaries between SA and other psychological phenomena will be avoided by casting SA in the light of system functions rather than individual mental constructs, and using the frameworks of Cognitive Systems Engineering (Hollnagel & Woods 2005), distributed cognition (Hutchins 1995) and distributed SA (Salmon et al. 2009).

2.4.2 Endsley's Three-Step Model

One of the most common SA models is the three-step model by Endsley (Endsley 1995, Salmon et al. 2009). Remember Endsley's definition of SA as the perception of elements, the comprehension of their meaning and the projection of their future status. In this view, SA is not a process but a "state of knowledge" (Endsley 1995, p. 36) while the process of acquiring SA is called "situation assessment". SA should also be viewed as something separate from decision making and performance, even though it is related to both. A person with poor SA will, according to Endsley (1995) make poor decisions and thus have a poor performance, while a person with good SA still might make erroneous decisions or fail to execute actions which will lead to poor performance. This model relies heavily on the "mental models" of the operator to map onto features in the environment, which is consistent with the view of Klein (1989) and the Naturalistic Decision Making approach to decision making (Klein 1998).

SA is also something separate from, but related to, individual factors such as workload, attention and stress. Endsley (1993) has shown that workload may vary independently from SA, even though many other studies have been made, some with a positive correlation between workload and SA, some with a negative, and some showing no effect or correlation, see e.g. Hart (2006) or Hansman (2004).

The three steps in Endsley's model follow from the definition of SA. The first step is thus the perception of elements, the second the comprehension of said elements and the third and last step is the projection of future states based on the comprehension. This results in SA, which in turn influences decision making and the performance of actions which leads to modifications in the state of the environment, which again is perceived. Figure 2-6 is a common illustration used to explain this model.

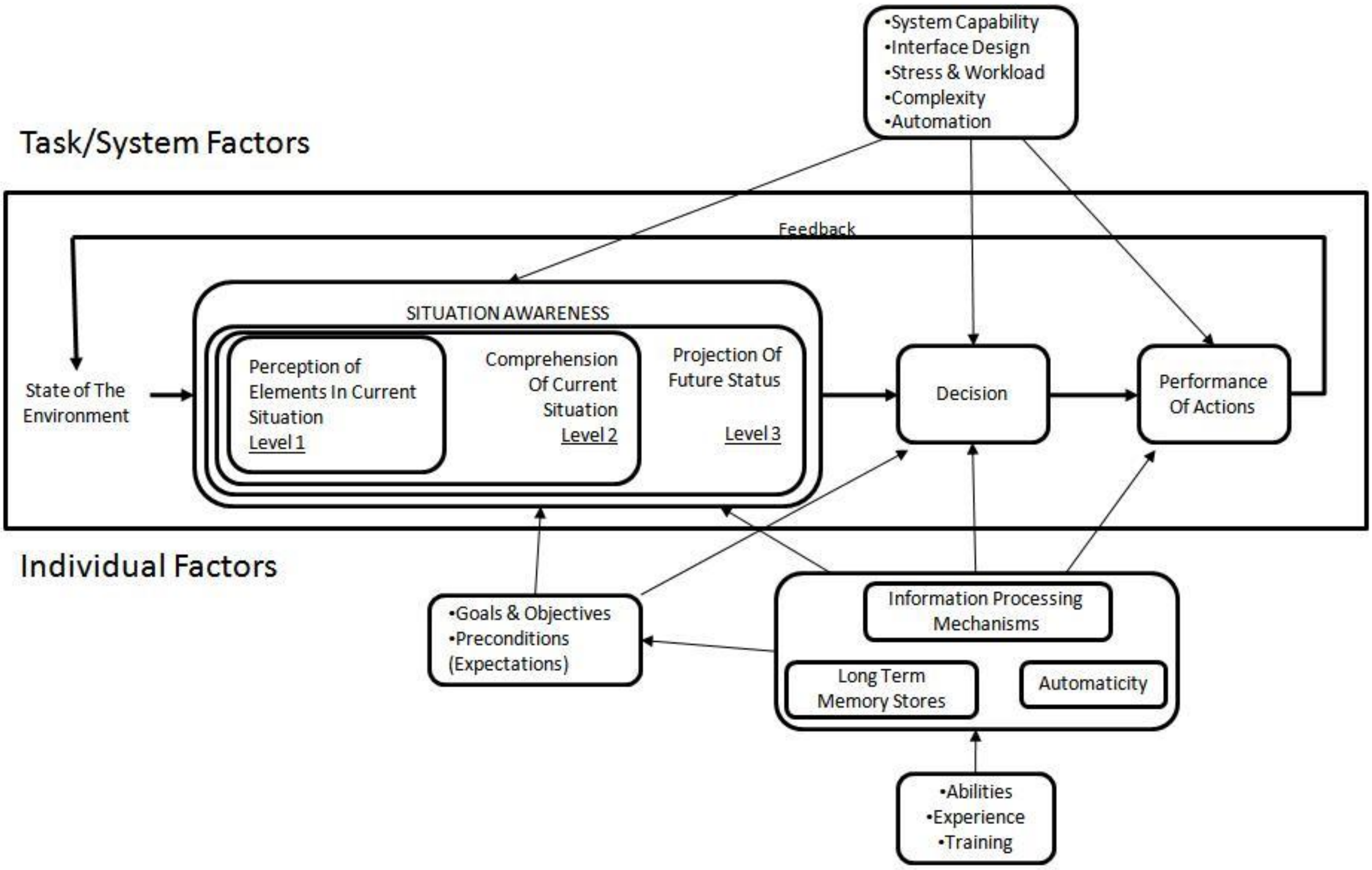


Figure 2-6 Endsley's Three Step Model (Adapted from Endsley 1995)

The model in the figure above shows how SA relates to other constructs such as decision making. The three steps of SA precede the decision. It also shows how SA is influenced by system and task factors, such as stress, workload, complexity and interface design, as well as individual factors, such as goals and objectives, long term memory and training.

2.4.2.1 Three-Step Model & DOODA

The three-step model is a general SA model for an individual. However, by applying the same line of reasoning that supported the creation of the DOODA-loop from the OODA-loop, it can easily be described and applied to a system operating in a C^2 environment, thus establishing a clear link between SA and C^2.

The OODA-loop and the three-step model are both constructed on the premise that an individual first needs to perceive elements in the environment, then comprehend these elements, before a decision can be made. This is a classic behavioral view of the human as an input-output machine, and, as previously stated, Endsley constructed the model on an information processing paradigm (Salmon et al. 2009).

The DOODA-loop then takes the OODA-loop and uses a cybernetic approach to create a system-level model. This could also be done on Endsley's three-step model. In fact, the three-step model and the DOODA-loop are already strikingly similar. The first function of the C^2-system in the DOODA-loop is "information gathering" (via sensors) while the first step in SA in Endsley's model is "perception of elements in current situation". Then comes "sensemaking" and "comprehension of current situation" respectively, and last "planning" or "projection of future status". This is then followed by a "decision", or "orders", then "military activity" or "performance of actions" which has "effects" on the "state of the environment". The similarities between sensemaking and Endsley's Level 2 SA is further discussed in Endsley (2004) and also touched upon in Salmon et al. (2009, p. 30-31).

To further demonstrate the similarities between the DOODA-loop and the three-step model an illustration of a synthesis between the two is provided below in figure 2-7.

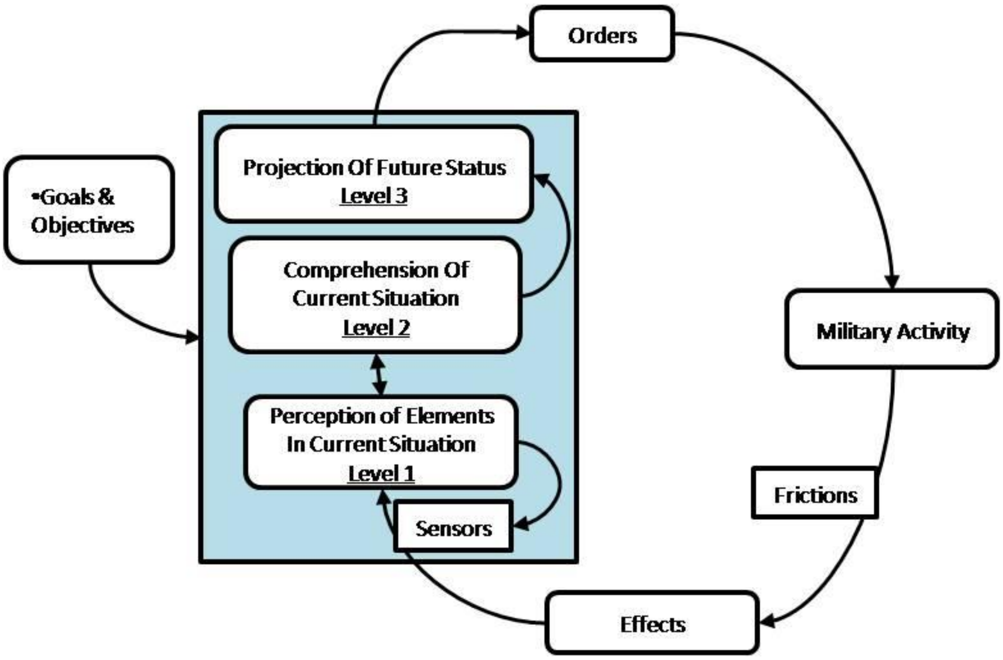


Figure 2-7 DOODA and the Three Step Model

This is a functional perspective inspired by CSE where the structure of the system and how it performs these functions are of secondary concern. It is easy to make additional connections to COCOM and how the blue C^2-system could be seen as a JCS. This connection will be explored further in section 2.7 Summary.

Brehmer (2006) states that there are three levels to a C^2-system; the Purpose (Why?), the Functions (What?) and the Form (How?). This concept illustration in Figure 2-7 concerns the functions, the "what?", and not the form or the "how?", while Endsley's original three-step model tries to take both the "what?" and the "how?" into account. Therefore the system and individual factors shown in the three-step model is not included in this illustration. However, we will return to Brehmer's (2006) three different levels of analysis in section 2.7 Summary as a framework for showing how mixed conditions potentially could affect the system functions.

Brehmer (2007) has also commented on the use of SA as a function in the C^2-domain, and also stating that sensemaking "is the current buzz word in discussions of C^2, having succeeded situational awareness (SA) as everybody's favorite concept" (Brehmer 2007, p. 223) and cautions against confusing a *function* with a *process*. Although he is also admitting that the concept of sensemaking in the DOODA-loop relies on the definition of Weick (1995, in Brehmer 2007) and that it is not clear whether Weick intended sensemaking to be a function or a process (Brehmer 2007, p. 224), but that considering it a function is a good fit for the model. The notion of SA as a (distributed) function is not new however, (see for example Artman & Garbis 1998) and research in that direction is currently going strong (Salmon et al. 2009).

In this functional view, the perception of elements would be more akin to the information gathering function of the DOODA-loop, where one or more parts of the system perceive elements via sensors, and not necessarily senses. This means that we are moving from viewing SA as an internal product but rather as a set of interlinked processes or functions that can be performed by the system as a whole or by parts of it. An example would be a traditional military hierarchy where soldiers on the fields, aided by artefacts (or sensors) such as RADAR, signal detection and binoculars, give reports to higher commanders. These commanders are assembling reports from many different sensors and try to create an operative image of what is happening. Their need for more information is communicated back to the appropriate sensors who direct their efforts on gathering specific information. Once the commanders have achieved an understanding of the situation, they plan ahead (projection) and issue orders (makes a decision) that are communicated to lower ranking units that then put these orders into effect. This means that the entire system as a whole has a quality of SA. This notion of SA as a system property will be further discussed in section 2.7 Summary.

2.4.3 Situation Awareness Measurement

As mentioned, there are many models and definitions of SA. It is therefore only natural that there are also a plethora of methods available that claim to measure some form of SA. However, there are few validated methods for measuring SA (some would claim that there are no validated methods). Salmon et al. (2009) reviewed several methods in terms of type, individual or team, number of validation studies, main strengths and weaknesses, etc. Of those, the method deemed to be the best fit for this study was the Crew Awareness Rating Scale (CARS) developed by McGuinness & Foy (2000) at the British Aerospace's Sowerby Research Centre (BAe SRC). CARS was originally developed for generic use, such as SA assessment in army commanders or for evaluating control room operators (McGuinness 1999), which made it easy to adapt to the C3Fire environment. Salmon et al. (2009) lists the strength of the method as;

> "1) Developed for use in infantry environments
> 2) Less intrusive than on-line techniques
> 3) Quick, easy to use requiring little training"
> (Salmon et al. 2009, p. 53)

In addition, McGuinness (1999) and Stanton et al. (2005) note that it is a method easily adapted to other environments. Since CARS is a post-trial questionnaire it is not as intrusive as freeze-probes and similar techniques, such as SAGAT (Endsley, 2000). One of the main advantages is that it is a quick and easy method that does not require prior training of the participants. Nor does it call for subject matter experts (SMEs) for analytical purposes as other methods do, such as SABARS (Matthews & Beal 2002) or SASHA (Jeanott, Kelly & Thompson 2003). In previous studies (for example McGuinnes & Ebbage 2002), CARS has been used together with NASA Task Load Index (TLX), which is also the basis of one of the other measurement used in this study, and as a tool for looking at differences in SA when different systems (analogue vs. digital) were used during simulated command and control exercises (ibid.).

2.4.3.1 CARS

CARS is based on Endsley's three-step model (Salmon et al. 2009). It uses 8 questions to assess individual SA, as shown below:

1: Perception – Content	2: Perception - Process
3: Comprehension – Content	4: Comprehension - Process
5: Projection – Content	6: Projection - Process
7: Integration – Content	8: Integration - Process

Figure 2-8 CARS Questions (adapted from McGuinness 1999)

For each of these the participant is asked to rate themselves on a 1-4 scale where 1 is the best case scenario and 4 the worst case (McGuinness & Ebbage 2002). The reason for using only 4 steps in the scale is that research has previously showed that individuals have trouble rating their own SA on larger scales (McGuinness 1999). The four-point scale for the "content" ratings therefore range from a 1 which is a definitive and certain "Yes, I have a good SA" to 4 which is a definite and certain "No, I do not have good SA" (ibid.). The two options in the middle are "Probably" and "Probably not", which are the uncertain affirmative and negative answers. For the "process" part, the four point scale goes from "Easy" to "Unmanageable" (ibid.).

The first three levels are easily deductible from Endsley's three-step model (Endsley 1995) and the fourth, "integration", is meant to assess how well the participants can synthesize their SA and their future actions, for instance how well they can deduct what they should do next based on their current understanding of the situation (Salmon et al. 2009).

McGuinness (1999) also suggests using more than one CARS-questionnaire focused at different subtasks to further explore the subject SA with regards to different aspects of the environment and task.

2.4.4 Methodological Issues

As with most methods pertaining to SA there are drawbacks and trade-offs. Salmon et al. (2009) claims that only two validation studies have been made using CARS. However, only 4 out of the 18 methods Salmon et al. (2009) investigated have more than 2 validation studies while half (9 out of 18) have fewer than 2 validation studies.

CARS is a subjective measurement that relies on the participant to accurately and correctly recall and judge their own SA on a number of subtasks. As mentioned before, this is a serious question in SA research; "can individuals with low SA be aware of the fact that they have low

SA?" Another issue is that the task must allow for the distribution of the CARS measurement, and its distribution may alter the level of realism in the experiment (Breton, Tremblay, & Banbury, 2007).

In regards to the issue of looking at system level SA rather than SA as something inside the individual operator's head, it might seem strange to use a method that relies on self-rating. There are however no validated methods for analyzing distributed situation awareness. The methodology suggested by Stanton et al. (2006) and Salmon et al. (2009) would be ill-fitted with regards to the other measurements in this study, as well as require too much in terms of subject-matter and methodology expertise and is thus not suitable. CARS is based on Endsley's three-step model, and just as that model can be cast in a functional system perspective, so could the results of CARS. CARS is also going beyond the three-step model and is also asking for the process of acquiring SA, what Endsley would call *situation assessment*, which further eases an analytical perspective where "functions" is the focus. It is also important to keep in mind that CARS is but one of the methods used for data collection in this study, and that it together with Shared Priorities (see section 2.6.1) and DATMA (see section 2.6.2) should provide a rich enough image of the system workings for the purpose of this thesis.

2.5 Mental Workload

Mental workload (also referred to only as "workload" in this thesis), much like C^2 and SA, is a well-used term with no one standard definition. Hart & Staveland (1988) define workload as "a hypothetical construct that represents the cost incurred by a human operator to achieve a particular level of performance". This definition reflects the view of workload as something internal to a human operator. As with virtually all hypothetical constructs that exist solely in the mind of an individual, the most common way to measure workload has been subjective ratings, such as NASA TLX (Hart & Staveland 1988). The "human operator" part is still in use today for workload definitions, as seen in a later definition by Hart defining workload as "a term that represents the cost of accomplishing mission requirements for the human operator" (Hart 2006) and Parasuraman, Sheridan & Wickens (2008) who describe workload as "the relation between the function relating the mental resources demanded by a task and those resources available to be supplied by the human operator".

Workload is often seen as something that is influencing performance or SA (see for example Endsley's model described previously), but is not always clearly defined. It has also been bandied around with little care among certain HF practitioners and used as a catch-all term for explaining accidents much like SA has. This led to criticism from, among others, Dekker & Hollnagel (2004) and Dekker & Woods (2002) similar to their critique on SA; that it is a folk model without empirical base and has been overused in a way that conflate catchphrases such as "mental overload caused..." and the actual underlying cause.

Parasuraman et al. (2008) meets this critique and argues that there is a solid empirical base and general consensus (if not one specific definition) in the scientific community regarding what workload is, and perhaps more importantly what it is not. By separating workload from things like performance, they can be investigated separately which is useful in many cases.

Several studies have shown that a high workload can be maintained with little or no loss in performance compared to situations with low workload (see Parasuraman et al. 2008 for a review). This is mainly done through adaptations and strategies on behalf of the human operator.

Mental workload will in this thesis be used as an indicator of differences between the different conditions as well as a way of looking at discrepancies between operators in the mixed condition. The method for measuring it will be further described in section 2.6.2 Questionnaires for the Distributed Assessment of Team Mutual Awareness.

2.6 Shared Understanding in Teams

Shared understanding, or mutual awareness, is an important aspect of team performance. Team members need to coordinate their own actions to achieve the common goal, and in order to adapt and coordinate actions they need to be aware of the behaviors of the other members. Previous research (such as Stout, Cannon-Bowers, Salas, and Milanovich, 1999) has shown that high awareness, or "sharedness", can positively influence team performance.

One theoretical framework to use when looking at team shared understanding is SA, but rather than looking at individual SA, the whole team's (or system's) SA is analyzed. This type of SA goes under many names: Shared SA, Team SA, Distributed SA and so on. However, there are problems in scaling models and theories originally created for single individuals in an information processing paradigm (for example Endsley 1995). Some attempts have been made to create Team SA theories, for instance Endsley & Jones (1997) and Salas et al. (1995), but there are few, if any, validated measures that go with these theories (Salmon et al. 2009). This indicates that there is still a need for a simple and reliable method for investigating shared understanding, be it within the SA framework or based on other theoretical frameworks.

2.6.1 Shared Priorities

Shared Priorities (SP) is a recently developed measurement method for looking at shared understanding in teams.

The first step of the SP method is to ask the members of the team to either write down a number of priorities for the team, commonly 5, that are important for success in meeting the common goal, or to rank a number of pre-defined priorities with the same criterion (importance for team success) in mind. These 5 priorities may either be ranked on a 1-5 scale or a larger scale, for instance 1-25, to enable the team members to weigh the importance of the priorities with greater resolution. Next, the priorities of each member are scrambled and sent to another member in the team, who are asked to rank them on the same scale used previously. This is either repeated until all team members have ranked each other's priorities, or the procedure simply ends in larger teams where this is not feasible. A measure of concordance, for instance Kendalls' W (Kendall et al. 1939), can be used to determine to which degree the team members are in agreement of the next course of action. The hypothesis is that the greater the shared understanding of the situation (or the sharedness of a common operational picture) is in the team, the more in agreement they would be (i.e. the greater the concordance measure would be) concerning the priorities of what need to be done to achieve their common goal.

The SP-measure has been used in a few studies (e.g. Höglund et al. 2009; Berggren & Johansson 2010) so far, and none that has provided a reliable and statistically significant correlation with any other measure, e.g. performance. As with all new methods, several problems with the original design have been found such as the lack of a well defined goal for the teams in Höglund et al. (2009). Therefore one of the aims of this thesis is to make another attempt at cross-validating the SP with other common C^2- and team-measures, such as SA, workload and performance. This study has therefore been designed with the notion of giving the participants a clear, concrete and common goal in mind. In addition, the previous validation studies used methods with low validity (e.g. using only a single question with a 1-7 Likert scale for measuring SA, and similarly for workload, in Höglund et al. 2009) which might have interfered with cross-validating the SP-measure. The methods for measuring SA and workload used in this study are more robust and have had previous validation studies done, which hopefully will counteract problems encountered in previous studies using SP.

2.6.2 Questionnaires for Distributed Assessment of Team Mutual Awareness

The "Questionnaires for the Distributed Assessment of Team Mutual Awareness" (DATMA) method is described by MacMillan et al. (2005) as a way to measure the mutual awareness in teams. They define "mutual awareness" as "the extent to which team members are informed of other team members' behaviors" (ibid.).

DATMA utilizes three different aspects of "mutual awareness" as measurements; "Taskwork Awareness", "Workload Awareness" and "Teamwork Awareness". Different questionnaires are used for each of these measurements. Each of these questionnaires is administered post-trial and are subjective in nature.

The DATMA method was chosen because it is a battery of questionnaires, partly based on well-validated measures (NASA TLX), which, together with the CARS questionnaire, will provide a rich image of the system workings. DATMA is administered post-trial and thus not as intrusive as on-line or real-time questionnaires or measurements, nor does it require any training on the part of the subjects, and the use of Subject Matter Experts (SMEs) is limited to creating the first questionnaire.

2.6.2.1 Taskwork Awareness

For Taskwork Awareness the team members are individually asked to remember a salient event during the trial. They are then asked to recall and mark what they themselves were doing at the time as well as what the other team members were doing. MacMillan et al. (2005) proposes two different ways of constructing the questionnaire. One is by having a SME pre-trial create categories of actions. These categories are then included on the questionnaire and used by the team members when classifying each other's actions. The other option is to have the team members themselves freely write down what they themselves and the others were doing. A SME would then create categories from these free-form answers. The scoring for this section uses a congruence measurement of agreement between the different team members. The more accurate the team members are in reporting each other's actions at that specific time (i.e. using the correct category for classifying the others actions), the higher the score. The score is percent of agreement between team members.

2.6.2.2 Workload Awareness

There are three parts to the Workload Awareness measurement. It builds on the widely used Task Load Index (TLX; Hart & Staveland, 1988) that provides measurement of individual workload. TLX uses 6 subscales on which the subject rates him/herself between 0 and 100 in unlabeled 5-step increments. The 6 subscales are: mental demand, temporal demand, performance, effort, frustration, and physical workload.

NASA TLX has previously often been used together with SA measurements of various kinds, among them CARS (McGuinnes & Ebbage 2002), with varying results. Some, such as Farley et al. (1998), have not found any correlations with TLX and SA, while others found correlations between SA and TLX (Hansman 2004). Hart (2006) mentions that SA is one of the most popular covariates with TLX with 7% of all papers using TLX also using SA but that the results are indeed varied to the point where it has been suggested that SA is actually an effect of workload rather than an independent construct (Hendy 1995).

The first part uses five of the TLX subscales with physical workload omitted, as it rarely is useful in simulator studies (MacMillan et al., 2005). Each team member rates themselves from 0 (Very Low) to 100 (Very High) in 5-step increments. The second part has each member rate the other team members' *overall* workload from 0 (Very Low) to 100 (Very High). The subscales are not used in this part. For the third and last part of the Workload Awareness measurement each team member rates the overall workload of the team as a whole using the aforementioned five subscales.

The "sharedness"-scoring for this part consists of using the differences between the different team workload ratings. The difference in two team mates' ratings of the workload of the team is squared and summed separately for each of the five subscales. The square root of this is then used as a measure of discrepancy in perceived team workload.

2.6.2.3 Teamwork Awareness

The third questionnaire concerns how each individual experienced the teamwork. This part is a development from earlier SME observer rating schemes. However, rather than having independent SMEs perform the rating, the team members themselves rate their teamwork performance on four subscales, each capturing different aspects of teamwork. The four scales are: communication, back-up, coordination and information management, and leadership/team orientation. MacMillan et al. (2005) define the communication scale as concerning "team members' ability to provide important information to others" while coordination and information management is "team members' ability to pass critical information to the other members". Back-up is the awareness of others' workload and the adjustment of others action to compensate for a team member that is overloaded. Finally, the leadership and team orientation score aims to measure the members' ability to agree on goals and tasks.

Again, agreement score can be calculated between team members to assess the differences in the perceived level of teamwork along the four subscales, which are measured by 7-point Likert-scales.

2.6.3 Methodological Issues

There are a number of recognized problems with this type of assessment (Stanton et al. 2005). For instance, all the data collected is based on the subjective opinions of the participants, each likely to have a different frame of reference than the others. Participants are also more likely to remember salient events, such as times when they lost control or had extremely high workload or teamwork, and a tendency to forget the periods of low workload/teamwork. Participants may also have a problem with assessing the workload for the entire scenario since the workload is likely to fluctuate throughout the scenario. Stanton et al. (2005) also lists two other disadvantages to this assessment method; the lack of evidence of usage in the literature as well as the lack of validation studies (even though some parts, for instance the NASA TLX, has been validated separately). They also point out that participants are likely to correlate their workload with performance (ibid.).

The advantages on the other hand are that it is a quick method which requires minimal training for the participants. It is also easy to score and inexpensive. Further, it assesses the workload on several aspects (subscales) for both team and individual workload (ibid.). It is also based on NASA TLX which has been validated in numerous studies (Hart 2006). The method is also looking at workload, teamwork and partly, via taskwork awareness, a kind of shared understanding.

2.7 Summary

In the section 4.2.1 Three-step Model & DOODA, the notion of using a functional systems perspective when looking at SA is brought up. This is not a novel idea, one of the first suggestions of using distributed cognition, a closely related perspective, when analyzing SA was made in 1998 (Artman & Garbis 1998) and it has also received some more attention lately (Salmon et al. 2009). While the three-step model by Endsley has an information processing approach (Salmon et al. 2009), CSE looks at constraints that are external to the system, i.e. present in the environment, rather than on internal constraints, i.e. memory limitations, attention span, etc., that are the foci of the information processing perspective (Woltjer, Prytz & Smith 2009). For instance, the Contextual Control Model (COCOM) described earlier specifically states that it is the context that, together with the competence of the system, determines the actions that lead to a change in the environment. Salmon et al. (2009) argues that the idea of distributed SA (or DSA) is in direct conflict with Endsley's model due to the difference in theoretical views, but I would rather argue that the essential functions of Endsley's three-step model still can be used in a DSA paradigm to explain and reason around in.

The shift of analytical foci, among others, that a CSE perspective brings with it has already been discussed briefly in the "Three-step Model & DOODA" section, but some important points bear repeating. The first being that in adopting a functional perspective, the model can be carried outside of the mental workings of an individual and used to describe the workings of a system, for example a JCS. The scaling of SA from an individual to a team is thus made much easier. Weick and Roberts (1993, cited in Artman & Garbis 1998) state that "there can exist a collective state of mind which is not represented in any single locus" where the collective state in this case would be SA. But more than that, SA should be seen both as a system-wide state and as a system-wide function.

COCOM and ECOM are relatively abstract models that aim to describe how one can look at the performance of a JCS on several different "levels" and "modes" of control. Woltjer, Prytz & Smith (2009) took a more practical approach and found that a JCS, in a C^2 setting, could be analyzed on several different levels using FRAM, where subordinate functions formed larger functions corresponding to different layers of control in C^2; Operational, Strategic or Tactical. The baseline is that actions are determined by an understanding of the current situation, and that this understanding can exist at different levels with different implications. Applying this reasoning to SA would mean that functions in some JCSs that are part of a greater JCS could perform functions that would amass to for instance the information gathering functioning (or perception function) of a large-scale JCS, and thus provide DSA.

Brehmer (2006) uses three different levels (the "Why", "What" and "How" previously mentioned) for the purpose of analyzing C^2-systems. In the Figure 2-7 DOODA and the Three Step Model only the functions (the "What") of the synthesis between DOODA and Endsley's model were shown. The main focus of this thesis is, as mentioned in the introduction, to study the effect of mixed conditions on a number of factors or, to use CSE terminology, functions, including performance and shared understanding (operationalised partly as SA). Thus we come to the lowest level of analysis in Brehmer's analytical framework; the "How". Endsley's SA model already includes system and individual factors, such as interface design and system capabilities, and it is these factors that will be manipulated in the experiment described in the Method-section.

3 Method

This chapter consists of two larger sections. The first is a section on Experimental Design. Here the general design of the experiment, including the scenario design of C3Fire, will be described, as well as the participants, pilot studies and general experimental procedure. The second section will detail the various dependent measures used in this study, and also detail what changes were made to the measurement methods described in the previous theory chapter.

3.1 Experimental design

A within-group design with three conditions balanced over time was used to collect the data. For each condition two participants were tasked with fighting a fire in C3Fire. As mentioned in the theory chapter, the foundation of C3Fire is the forest fire simulation but it is also much more than that. Depending on the research design, C3Fire can be configured to accommodate almost any command structure and number of participants. Different roles can be assigned to different participants and customized GUIs can be designed for each role, as well as constraints and information flow (for example feedback). In this experiment the participants assumed the roles of commanders, one responsible for the fire fighting vehicles (Commander 1) and one responsible for the water and fuel trucks (Commander 2). Each commander communicated his/her orders to a trained operator via electronic messages and the operator carried out the orders (i.e. moved fire, water and fuel vehicles) in the micro-world.

Three different experimental conditions were used. One with both commanders using the C3Fire graphical user interface (GUI) to manage their units and send messages to each other and to the operators who controlled the units, called the GUI/GUI condition. A second condition had one commander (Commander 1) using the C3Fire GUI while the other (Commander 2) was using a regular paper map and an electronic messaging system to communicate with his/her operator and the other commander. This was named the GUI/Map condition. In the third condition both commanders used paper maps and the messaging system for communication, the Map/Map condition. Each scenario was functionally equivalent to allow comparisons to be made. This was achieved by rotating and mirroring the original scenario (the terrain, location of vehicles and fires) while using the same map. The GUI/GUI condition was played on the original map while the GUI/Map was played on the same map rotated 90 degrees clockwise and mirrored vertically. The Map/Map condition was rotated 90 degrees counterclockwise. The scenarios were balanced over time (see table 3-1).

Table 3-1 Balancing over Time

Experiment #	Trial 1	Trial 2	Trial 3
1	GUI/GUI	GUI/MAP	MAP/MAP
2	MAP/MAP	GUI/GUI	GUI/MAP
3	GUI/MAP	MAP/MAP	GUI/GUI
4	MAP/MAP	GUI/MAP	GUI/GUI
5	GUI/GUI	MAP/MAP	GUI/MAP
6	GUI/MAP	GUI/GUI	MAP/MAP
7	GUI/GUI	GUI/MAP	MAP/MAP
8	MAP/MAP	GUI/GUI	GUI/MAP
9	GUI/MAP	MAP/MAP	GUI/GUI
10	MAP/MAP	GUI/MAP	GUI/GUI
11	GUI/GUI	MAP/MAP	GUI/MAP
12	GUI/MAP	GUI/GUI	MAP/MAP
13	GUI/GUI	GUI/MAP	MAP/MAP
14	MAP/MAP	GUI/GUI	GUI/MAP
15	GUI/MAP	MAP/MAP	GUI/GUI
16	MAP/MAP	GUI/MAP	GUI/GUI
17	GUI/GUI	MAP/MAP	GUI/MAP
18	GUI/MAP	GUI/GUI	MAP/MAP

The role of the operators in the experiment was to control the units in the microworld and, in the GUI/Map and Map/Map conditions, relay information back to the participant(s) using a map. The operators were needed in order for the commanders to control the units in the microworld especially when they were restricted to a paper map, else they would not have any way of interacting with the microworld that did not also provide feedback that would invalidate the experimental setup. A communication scheme was set up in order to ensure that the communication was coherent over trials and groups, as well as controlling what the operators could and could not do. This communication scheme is further described in 3.1.4.5 Communication Scheme.

In all conditions, the goal was to extinguish the fires in the microworld. More specifically, the goal was to prevent the fire from spreading and in particular spreading to buildings. Each square in the grid with a fire (burned out or active) was worth one point at the end of the scenario. If the square contained a building (a house or pump station) it was worth an additional two points. The goal was of course to have as low score as possible. The participants were given two sets of questionnaires, CARS and DATMA, after each scenario as well as the shared priorities measure to measure team SA, mental workload and mutual awareness respectively.

3.1.1 Scenario Setup

Figure 3-1 shows the basic map that was used in the GUI/GUI scenario, and that was rotated and mirrored for the other scenarios. The fire, in this scenario, starts in the top right corner (red squares). The participants knew from the briefing that the fire was in the northeastern corner and had to use their units to scout for the fire. The fuel station at coordinate AE4 is within reach for the fire, if it is allowed to grow in a direct line from AJ4. If at any point a fire was extinguished during this way, the fire would not reach the fuel pump. This was essentially an extreme performance marker; if the participants failed to extinguish any of the five squares between the original fire and the fuel pump it would add 10-12 points to the total score when it exploded. This also served to direct the participants focus for the first part of the scenario, since it was natural for the participants to start by closing off the fire in the direction of the fuel pump and then start working their way around the rest of the forest.

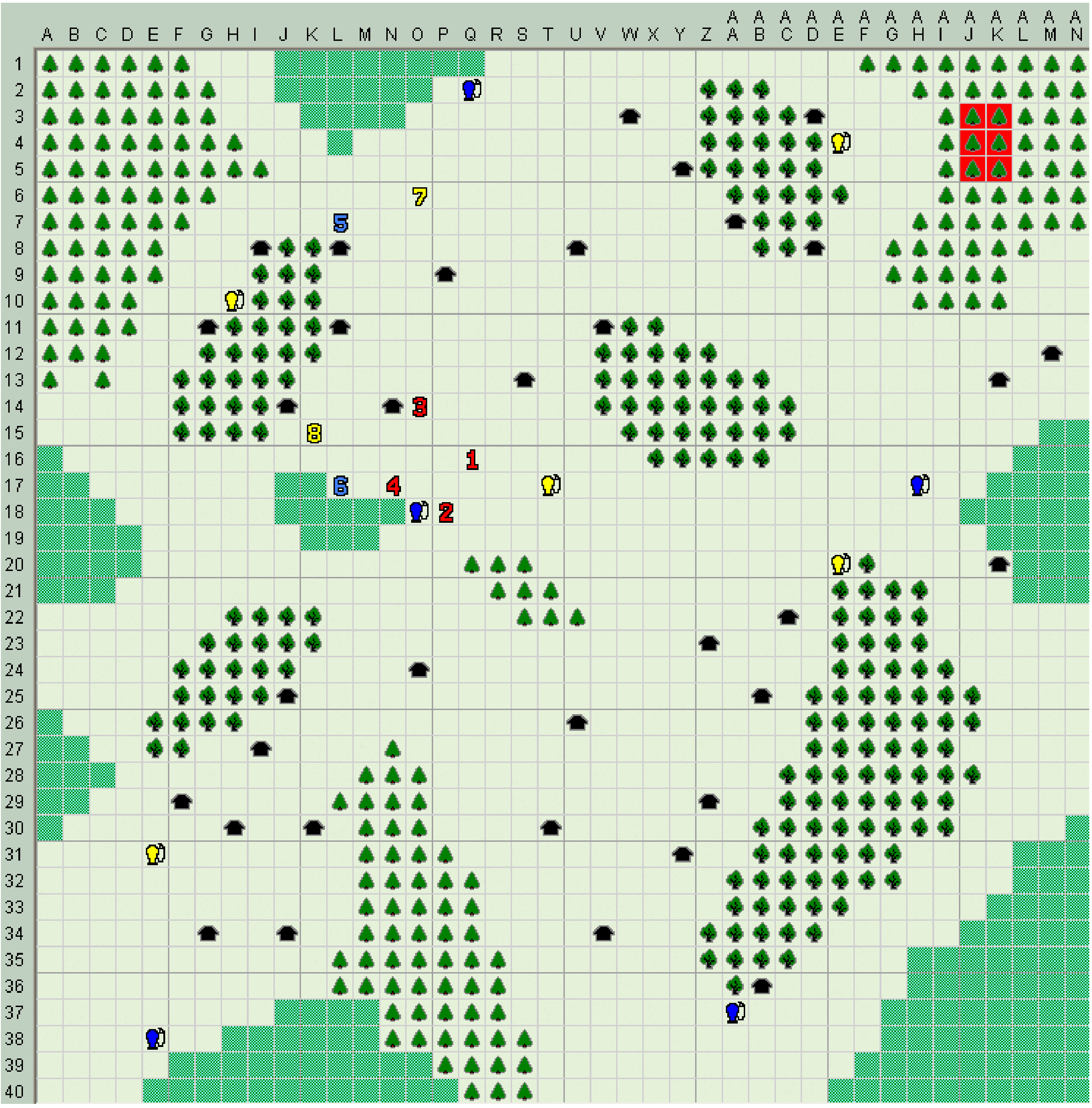


Figure 3-1 C3Fire Map

3.1.1.1 Condition 1

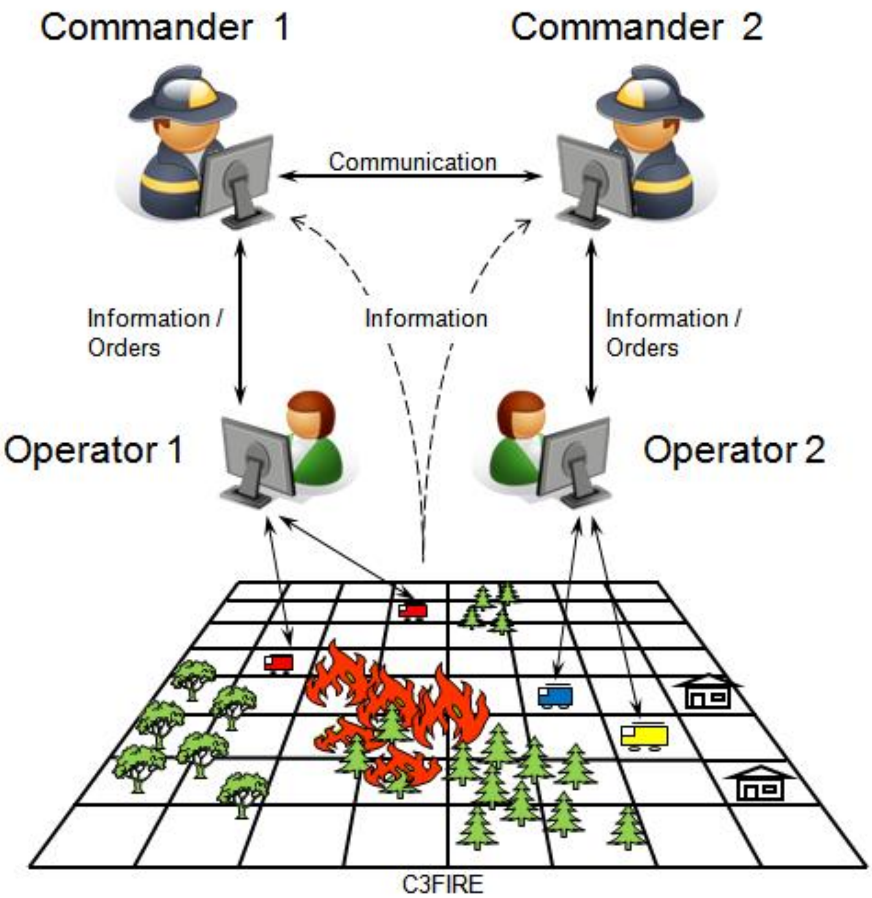


Figure 3-2 GUI/GUI Setup

In condition 1 (GUI/GUI) both commanders had access to the graphical user interface (GUI) of C3Fire, see the illustration above (fig. 3-2). That meant that they could see the position of each vehicle and also get a moment-to-moment update on what the vehicles could "see", such as fire. Each vehicle had a 3x3 line of sight (LOS) while the rest of the map was covered in a "fog of war" (FOW) where only the static map, i.e. terrain and structures, could be seen. The electronic map used in C3Fire had a coordinate system consisting of 40x40 squares. The maximum score for the scenario if no action was taken (that is, no fires extinguished) was 93, the same as in the other two conditions.

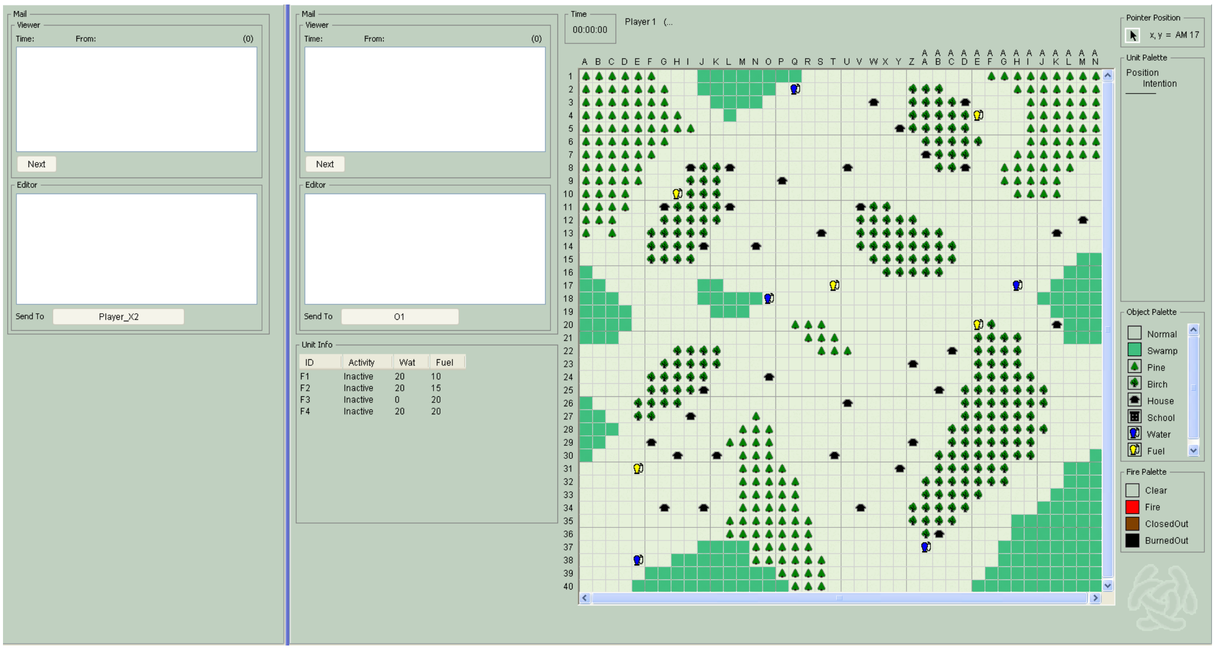


Figure 3-3 C3Fire GUI Layout

The GUI interface is shown in figure 3-3 above. There are two chat/messenger windows, one for communication with the other participant and one for communicating with the operator (receiving information and giving orders). The major feature of the GUI is the large map shown in the middle.

Unit Info

ID	Activity	Wat	Fuel
F1	Inactive	20	10
F2	Inactive	20	15
F3	Inactive	0	20
F4	Inactive	20	20

Figure 3-4 Information Window

There is also an information display (figure 3-4) under one of the chat windows where the unit IDs of the units the participant controlled are shown, along with fuel and water level. It also showed what the unit was doing at that time, e.g. fighting fire, inactive or moving.

3.1.1.2 Condition 2

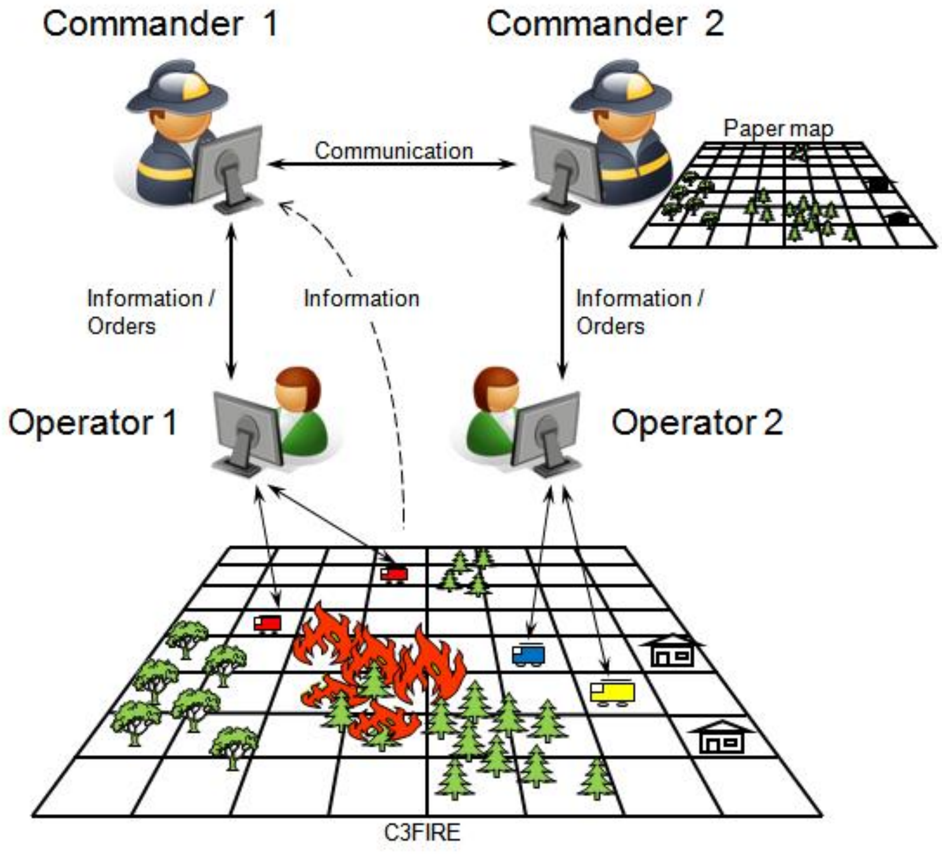


Figure 3-5 GUI/Map Setup

In condition 2 (GUI/Map), Commander 1 had access to the C3Fire GUI while Commander 2 had a regular paper map and a system for electronic messaging on the computer, illustrated in figure 3-5 above. The map used was the same as in C3Fire and using the same grid. Commander 2 was given markers in different colors (one green, one red, one blue, and one black) that he/she could use together with plastic Over Head-sheets (OH-sheets) to annotate on top of the map. The operator provided feedback to the commander with the map regarding what the vehicles could see. The commander using the C3Fire GUI could see his/her own vehicles and their LOS, but not the vehicles that were controlled by the other commander (unless they were within the LOS for his/her units). The commander using the paper map still had a similar computer interface for communicating with the operator and other team member, but it lacked the electronic map. The small information window, see fig 3-4 above, was still present, so the participants could still see fuel and water levels as well what activities the units were performing.

3.1.1.3 Condition 3

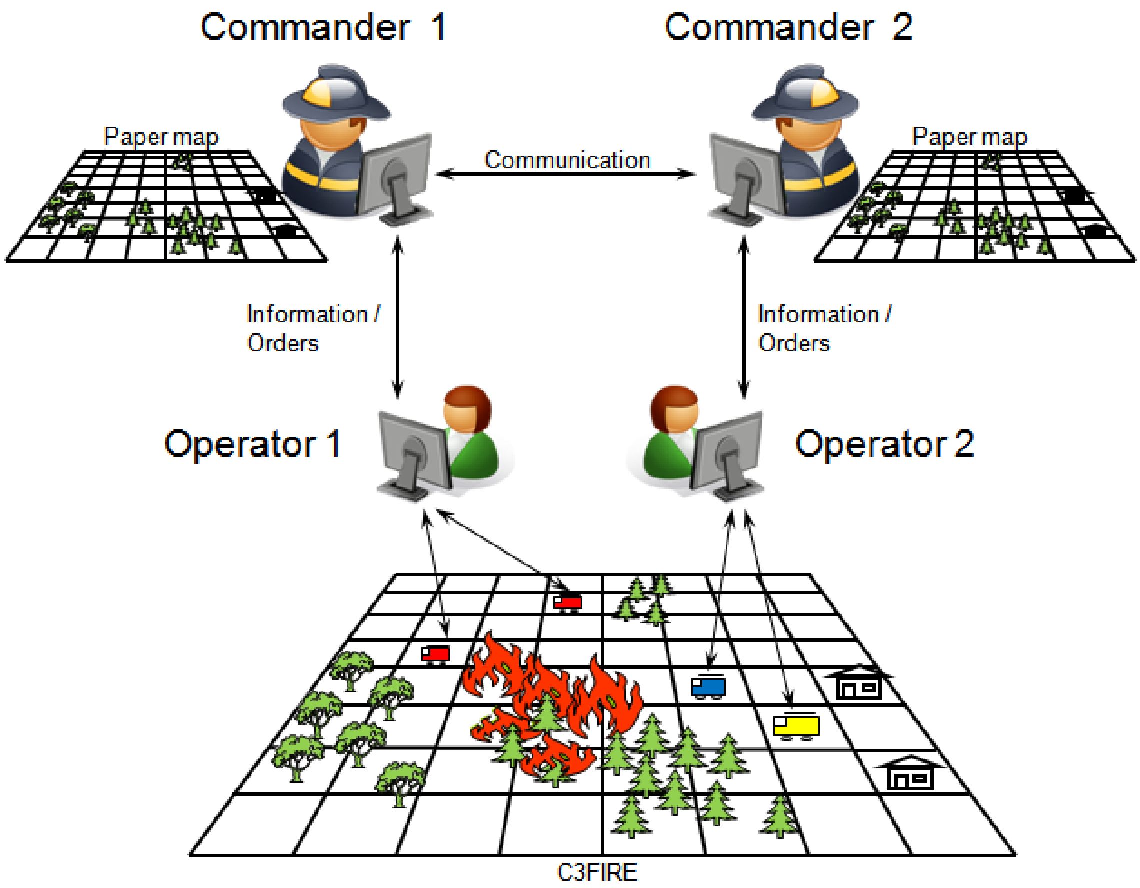


Figure 3-6 Map/Map Setup

For condition 3 (Map/Map) both commanders used a map that was an exact copy, and used the same coordinate system, as the C3Fire GUI map, illustrated above in figure 3-6. Both also had an electronic messaging system on their computer and markers for annotation. Feedback was provided from the operators with the location of fires that the vehicles could see. In this condition both commanders were essentially blind to the actual scenario and had to rely solely on second-hand information from the operators.

3.1.2 Pilot Study

Three runs of a pilot study were conducted. The first used two of the operators as they played through the scenarios for the first time. The questionnaires were also distributed to them at the appropriate times. This could of course only hint at the difficulty level of the scenarios using the GUI/GUI condition and could not be used to judge the difficulty of the GUI/Map or Map/Map conditions as the operators used their interfaces at all times and did not switch to a paper map. However, it was useful for establishing a baseline for optimal, or at least near-optimal, performance as they had direct access to feedback from and control over C3Fire and were aware of the experimental setup with rotated and mirrored maps. It was also useful to test the questionnaire and minor changes were done to all after this first pilot, correcting typographical errors and adapting them to fit the experimental setting.

The second pilot run was done using naïve participants drawn from the student population of Linköping University. This run was conducted as a normal test procedure would run. The participants were given the same briefing and played the practice sessions and trials as normal. An unstructured interview was conducted after the experiment to collect opinions and feedback. The most important lesson from this second pilot run was that the communication scheme had to be optimized further by trimming unnecessary words, so that both operators and participants could communicate faster and more efficiently. The scenarios were deemed to be at the right level of difficulty.

Last a third pilot run was done using two more naïve participants, both students from Linköping University. This pilot was done only to confirm the difficulty and that the new communication scheme worked as planned. Thus only one scenario was used (GUI/GUI).

3.1.3 Participants

36 students from Linköping University were recruited to participate in the experiment. Ages ranged from 20 to 30, with the average age 23.6 years. 23 were male and 13 were female. All participants were offered one movie ticket as compensation, regardless of whether they finished the experiment or not. All participants completed the experiment. 8 participants had previous experience of C3Fire, with 1 participant having used C3Fire twice before and the other 7 once. These 8 were randomly distributed (due to the random sampling) so that four of these participants worked with participants without experience and four (including the one participant who had used C3Fire twice) worked together in two different teams. The remaining 28 participants had never used C3Fire before. The self rated previous computer experience ranged between 3-5 (on a 5-point scale) with the average being 4.28, while the self rated previous computer gaming experience ranged between 1-5 (on a 1-5 scale) with an average of 3.42. Two participants rated themselves at 1, as having no previous experience with computer games. Three teams consisted of participants who did not know each other previously, while the other 15 teams had participants who knew each other from earlier. The participants in those 15 teams had known each other for an average of 19 months, with the max being 43 months and min 1 month.

10 participants had served in the Swedish armed forces for an average of 11.8 months, max 15 months and min 10 months. The other 26 participants had not served in any of the armed forces nor had any training from the armed forces (e.g. via reserve officer training or similar).

In general the participants rated themselves low on previous experience with command and control systems (average 1.5 on a 1-5 scale), and on experience with fire fighting (average 1.05 on a 1-5 scale). They rated themselves on average as having low experience using paper maps (2.17 on a 1-5 scale) but slightly higher on using electronic maps (2.75 on a 1-5 scale).

All participants were Swedish citizens without reading or writing disabilities and all had native-level proficiency in Swedish.

Three operators were recruited to run the C3Fire scenario, two regular and one back-up. All were students from Linköping University and were subsequently trained in operating C3Fire.

During each experiment two operators were located separately from the participants, interacting with them only via electronic chat messages. The roles of the operators were to act out the order given by the commanders and communicate specific information to them during the scenarios.

3.1.4 Apparatus

Four computers where used to run the simulation and Shared Priorities program, and additional material, specified below, were also distributed to the participants.

3.1.4.1 Hardware

Four Dell computers were used to run C3Fire. One was a server computer that ran the C3Fire server. The server computer had a 1.17 GHz processor and 2 GB RAM while the other three computers each had a 2.67 GHz processor and 3 GB RAM. All computers used Windows XP PRO as operating system and were equipped with a Dell SX2210b 21.5" widescreen monitor at 1920x1080 resolution, a Microsoft Digital Media keyboard and a Microsoft Intellimouse optical mouse.

3.1.4.2 C3Fire Version

The C3Fire version used in this experiment was version 3.2.7.

3.1.4.3 Shared Priorities program

A program for collection of the Shared Priorities measurement was developed for the purpose of this study. It followed the procedure described in 2.6.1 Shared Priorities and allowed the participants to enter 5 priorities and then rank them on a scale 1-25. Then their priorities were scrambled and sent to the other team member for ranking. The output of the program was two .txt-files for each experiment where the priorities and both sets of rankings were saved.

3.1.4.4 Material

The materials, aside from the computers, used in this experiment were paper maps (exact copies of the maps used in C3Fire), plastic OH-sheets and colored markers for annotation. Scrap paper and pens were also distributed to the participants. All questionnaires but the Shared Priorities were printed on paper. A printout of the communication scheme (more information below) was also given to each participant. Before each scenario the participants were given a printed briefing with information about the scenario and the conditions (for instance if their team mate used paper map or GUI).

3.1.4.5 Communication Scheme

While the operators were necessary from a practical standpoint, since a participant with only a paper map would have no ability to control his/her units otherwise, it was also necessary to ensure that they were restricted to perform only the actions specified by the commander. The communication scheme provided strict rules and format for the communication between the operator and his/her commander. This enabled the operators to behave consistently between groups and provide the same information, in the same format, at all times. Due to the simplistic nature of C3Fires' control interface, the only user manipulation possible is to move the units by dragging them with the mouse. This meant that the only straight command the participants could give the operators was to move a specific unit to a specific position. Below, in table 3-2, one part of the communication scheme given to both participants and the operators explaining the format of the order (translated from Swedish) is provided.

Table 3-2 Communication Scheme: Command

Command:	Explanation:	Example:
<unit> <pos>	Operator moves <unit> to <pos>	F1 AC23

If the commanders wanted a specific unit to extinguish a fire, they would specify the unit and the position of the fire in order for the operators to move the unit to the position of the fire, which automatically initiates the fire fighting in vehicles equipped to perform that action.

While the user input may be simple, the information feedback given by C3Fire is richer. In order to give the participant using a map an understanding of the development of the scenario, the operator had a fixed set of events to which he/she responded with a written message to the participant. Below is a table (3-3) with the events to which the operators responded (translated from Swedish).

Table 3-3 Communication Scheme: Events

Event:	Message from operator:	Example:
Unit has extinguished fire	<unit> extinguished <pos>	F1 extinguished J7
Units spots fire not previously reported	Fire <pos>	Fire AA23
Unit has arrived at destination	<unit> arrived <pos>	W7 arrived K12

The events cover the important features of the fire fighting and unit movements. Essentially, no further relevant information would be gained by looking at the electronic, dynamic, map instead of relying on these messages. Further, the participants could ask for more information in the format provided below in table 3-4 (translated from Swedish).

Table 3-4 Communication Scheme: Questions

Questions:	Reply:	Example replies:
<unit>?	<unit> <pos>	F3 L32 G7 A8
<pos>?	If fire: Fire If unit: <unit> Else: Empty / N/A	AB12 fire F1

These questions allowed the participants to inquire about two of the most important features of the microworld: they could ask where a unit was and how the situation for a specific coordinate was. Note that the order in which a question or order was phrased was important.

For the <unit> question, the reply was always phrased with the unit ID followed by the position. More than one unit could occupy the same square. The order was important in those rare cases where the unit ID and the coordinates could be confused, such as F1, F2 and so on.

All these commands, questions and information messages are very brief. This was to enhance the ability of the operators so that the messages could be typed and sent quickly as well as for the participants to read and reply with their own orders or questions. The operators were trained in using the communication scheme by practising several times before the first live experiment. Participants were also given careful instructions on how to communicate with the operators during the briefing and had a copy of the communication scheme with them at all times during the scenarios.

3.1.5 Procedure

First a briefing session with the two participants was held. The participants were given a brief presentation on the purpose of the experiment and the procedure. The purpose was intentionally vaguely worded, to avoid participant bias, and read (translated form Swedish) “The purpose of this study is to examine shared understanding and performance in command and control systems”. The participants were also given instructions and a demonstration of C3Fire and how to communicate with the operators. The scoring system was explained as well as the primary goal of the scenarios; to achieve as low score as possible. The shared priorities method was also demonstrated with made up examples. All questionnaires used in this experiment were explained. Each participant received one paper with specific instructions for how to communicate with the operators. One participant was assigned to take the role of Commander 1 (firefighting) and the other was assigned the role of Commander 2b(fuel and water). The role assignment was randomized by having the participants fill out a background questionnaire each and then assigning the participant with the odd participant number as Commander 1 and the even number as Commander 2.

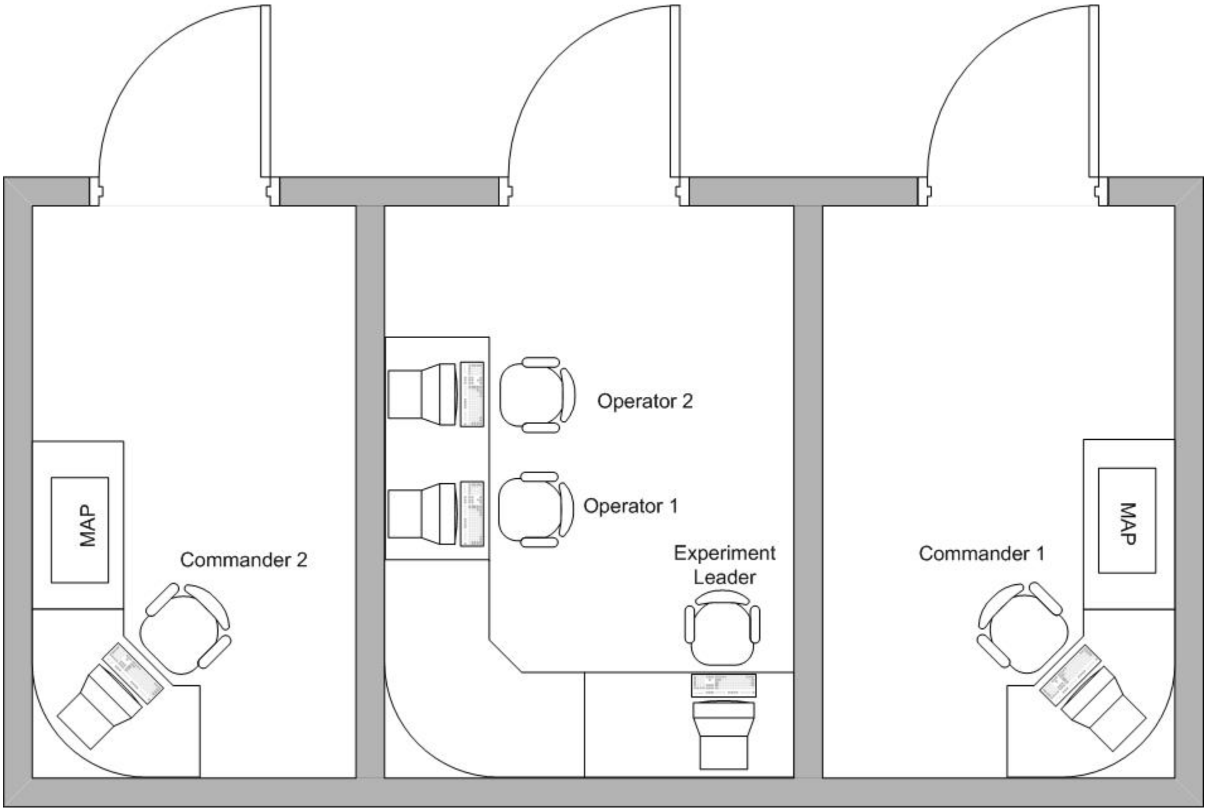


Figure 3-7 Physical Setup

The two participants were then seated in separate rooms with one computer each (see figure 3-7 for a schematic overview). The two operators were placed in a third room with one computer each. Each operator had contact with one commander. Before each scenario, both training and real, each participant were given a written briefing stating the location (coordinates) and fuel- and water status of each unit he/she controlled. Further it also informed them whether they and/or their team mate were using paper map(s) or GUI, how many points each terrain type was worth and finally also the approximate location of the first fire (for instance “A fire has been reported in the northeast corner”). The participants were informed during the briefing that one fire would start at the beginning of the scenario (and that the location was specified on the briefing) and that other fires may start during the scenario. In fact, in every scenario (including the training scenarios) a second fire would start at specific location. Participants were informed about this second fire in the same messaging window that they used for communication with the operators by specific coordinates.

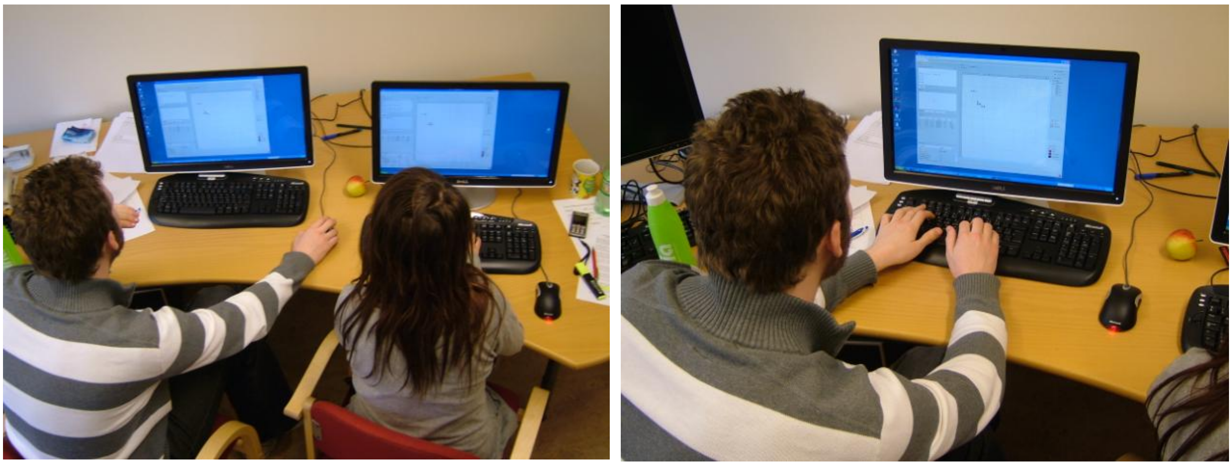


Figure 3-8 Two Images of the two trained operators.

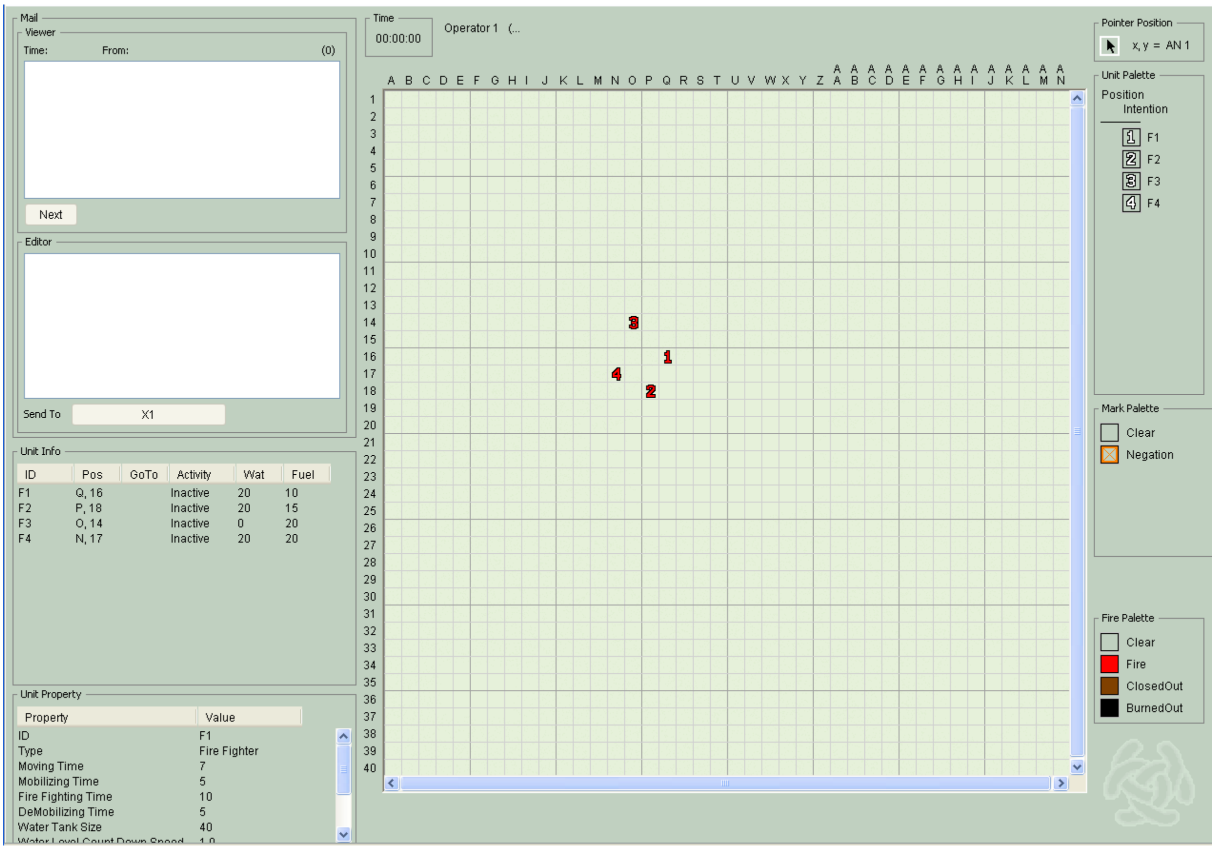


Figure 3-9 The interface of the operators. Notice that the map used by the operators lacked information on structures and terrain, to ease their task by eliminating, for them, useless information.

The first training session was run using a unique map. For the training session, both commanders used the C3Fire GUI. The training session lasted 10 minutes. During this first training scenario the commanders could also see the fire at all times, making the scenario easier and allowing them to see how the fire spread throughout different terrain types. Once the scenario was completed, the participants and the experiment leader gathered in one room to discuss the scenario. The experiment leader gave feedback to mistakes done by the participants (for instance when placing a water truck on the same coordinate as the fire truck instead of next to it or when attempting to move units that had no fuel) and also helpful tips to improve performance. The participants were also given time to discuss with each other their own communication and teamwork. Next a second training scenario was played. This scenario used the same map as in the first training scenario, the first fire started in the same place (the second fire started at another location) and the units were positioned in the same way as in the previous scenario. The participants could not see the fire at all times but had to scout for it using their units. When the scenario ended after 10 minutes the participants were again given a chance to discuss with each other and the experiment leader to further improve their performance and understanding of the microworld.

Figure 3-10 Images of participants. Top left is a participant playing in the GUI/GUI condition, top right is a participant in the Map/Map condition. Bottom left is a participant filling in the first DATMA questionnaire, and bottom right a participant with the Individual Workload questionnaire

Once these two training scenarios were done, the first of the three scenarios (varying to balance the order over time) was run. Each scenario took 15 minutes to complete. Each scenario started the clock at different times (GUI/GUI started at 00:10:00, GUI/Map at 01:25:00 and Map/Map at 02:55:00) in order to prevent the participants from easily “counting down” time of the scenario. Near the end of the scenario (after 14 minutes of play) it was “frozen” in time and the first questionnaire of DATMA was distributed. The monitors were turned off and the paper maps covered when this questionnaire was filled in. Once completed, the monitor was turned on, the paper map uncovered and the Shared Priorities data collection program was run. The reason for freezing the scenario rather than complete or abort it was in order to keep the current situation in the scenario active in the participants mind. In addition it made questions about possible future plans, actions and outcomes more relevant and motivated. Once these two questionnaires were done, the scenario was resumed for one last minute of play. At the end of this time the remaining questionnaires were distributed. This process was iterated for each of the three scenarios.

3.2 Dependent Measures

In this section the dependent measures used are described. This includes the CARS, DATMA and SP measures previously discussed in the theory chapter which are here further developed into concrete measurements. Adaptations and changes from the original design are also described.

3.2.1 Performance Measure

The performance measure from C3Fire was calculated in the following way: The maximum score that could be received from any of the scenarios (where one burned down square equaled 1 point and a burned down structure 3 points) was calculated to 93 by allowing the fire to spread without intervention. The performance score was then normalized to a scale of 1 to 0 where 1 was the ideal best performance, i.e. all squares were saved, and 0 the ideal worst performance, i.e. all squares that could burn down did so. The normalization, in short, was thus “1-score/max score” where “max score” = 93.

Further, a subjective performance measure is provided via the NASA TLX questionnaires as one question on the individual and team workload questionnaire respectively asks the participant to rate his/her own performance. This will provide subjective ratings of both individual and team performance.

3.2.2 Shared Priorities

Shared Priorities was collected using the computer program described earlier, following the same procedure as outlined in the theory chapter. The 5 priorities could be ranked 1-25 by the participants. Kendall’s W was then calculated to measure the level of agreement between participants.

3.2.3 Situation Awareness: Modifications to and Measures in CARS

The eight questions in CARS were translated into Swedish since that was the native language of the participants. The four specific answer alternatives (from “Yes, definitely” to “No, definitively not”) were abandoned for a more general four point Likert-scale with “Best possible case” in one end (1) and “Worst possible case” (4) in the other. This was done to space the alternatives in a more equidistant manner, similar to the other Likert-scales used on other questionnaires and simplify data processing.

The questions were also slightly modified to reflect this change in scales and asked (for “content”) how “reliable and accurate” (translated from Swedish) the information, understanding, predictions and the fusion of those three with their own actions are, respectively. The “process” questions were modified to ask how easy it is to maintain this level of information, understanding etc.

Since the questions were of a general nature, for example asking how reliable and accurate the information that the participant has right now is, the pilot study revealed that the participants found this abstract and without connection to the actual task. After the first pilot study (out of three), some examples were added in parenthesis after each question. The examples highlighted the connection between the general question and a specific instance of this in the actual task. During the briefing the participants were also instructed more carefully in what each category meant and that the example question in parenthesis were just there to make the connection between the general question and the C3Fire scenario clear and should not be considered as the question they are actually answering. There were no problems for the participants in the following two pilot studies or during any of the main trials.

3.2.4 Workload, Teamwork & Shared Understanding: Modifications to and Measures in DATMA

The DATMA method was adapted slightly to better fit this study. Firstly, the questionnaires were translated into Swedish, the native language of the participants in the study. The Workload questionnaires (NASA TLX) were not modified apart from the translation. Thus all five subscales (Mental Workload, Time Pressure, Performance, Frustration and Effort) were used.

The Teamwork questionnaire was also more or less unchanged apart from the translation, and all four scales mentioned in the theory chapter (2.6.2.3) were used. The specific terminology was simplified through the translation so that it asked for "common operational picture"[2] rather than "leadership/team orientation" (as there was no clear leadership structure) and "support" rather than "back-up". This change in terminology will be used consistently through the remaining thesis. The questions themselves asked for the same information as described in MacMillan et al. (2005), and the questionnaire can be found in Appendix A.

While these three questionnaires (individual workload, team workload and teamwork) was distributed at the end of the scenario, the first questionnaire in DATMA, Taskwork Awareness, was not. Rather it was distributed 14 minutes into the 15-minute scenario. The scenario was paused for the duration of this, and the Shared Priorities questionnaire which was given directly following the Taskwork Awareness questionnaire. The purpose of this was to change the nature of the questionnaire from a post-trial to a freeze-probe. While this is more intrusive than a post-trial, it was thought that the participants would be more likely to have a greater retention of the units' positions and tasks, which the questionnaire asks for. Another reason for introducing a freeze-probe was that the participants would likely be more motivated to complete the Taskwork and Shared Priorities questionnaires if they knew that the scenario was merely paused rather than finished, as that would have turned it into a rather pointless and academic exercise with a hypothetical question. To reflect this change, the Taskwork Awareness questionnaire did not ask the participants to remember a salient event but to mark the activities of the units as the current situation.

[2] COP will be used as shorthand for this variable.

11 categories were created by a subject matter expert and later condensed into 9 categories, see table 3-5. These were used by the participants to classify what each vehicle was doing at that specific time in the scenario. Translated from Swedish these were:

Table 3-5 Task Categories

Category	Task
1	Fighting fire
2	Moving to fight fire
3	Moving with other purpose than to fight fire
4	Inactive due to lack of fuel and/or water
5	Inactive due to any other reason
6	Refilling own fuel or water from vehicle or pump
7	Refilling fuel or water to other vehicle
8	Do not know
9	Other – please specify

One of these, #5, was added after the first pilot run. After the second pilot study the original 11 categories were condensed into 9 by combining previously separate categories of "refilling own fuel" and "refilling own water" into one, as well as "refilling fuel to other vehicle" and "refilling water to other vehicle" into one. The participants were asked to mark what each vehicle, including their team mate's, was doing at the time of the pause.

The questionnaire further asked the participants to give the location of each vehicle at the current time as well as mark down each square that was burning or had been burning during the scenario up to this point. This was asked for to further measure the level of SA that the participants had at that time. While not part of the original design of the questionnaire it was considered a minor edition that might prove useful. As it intended to measure the SA of the participant and not the level of shared understanding or mutual awareness, no congruency scores were calculated for these measures.

The questionnaire contained a map of the current scenario where participants marked the units and fire. The entire questionnaire had to be completed without aid from the digital or paper map used during the trial. The participants were instructed to turn off their monitors and, if applicable, cover their paper maps using a large white A3-paper sheet that was available (the experiment leader verified this when distributing the questionnaire) when completing the Taskwork Awareness questionnaire.

The scoring of this questionnaire consisted of several independent parts. First, the categories were grouped in "super categories", shown below in table 3-6.

Table 3-6 Combined Task Categories

Category	Task
1	Fighting fire
2 and 3	“Moving to fight fire” and “Moving with other purpose than to fight fire”
4 and 5	“Inactive due to lack of fuel and/or water” and “Inactive due to any other reason”
6	Refilling own fuel or water from vehicle or pump
7	Refilling fuel or water to other vehicle
8	Do not know
9	Other – please specify

This meant that if a participant had marked either a 2 or 3 and the unit was in fact moving when the scenario was paused, they were awarded one point for being correct. The grouping of categories together were done primarily for grading purposes; it is not possible for an outside observer to accurately judge, for instance, whether a unit is moving with the intent of fighting fire or some other intent. This meant that each participant could receive a maximum of 8 points for accurately marking what each unit was doing at the time. Category 8 was given 0 points, and a marking of 9 was judged on a case-to-case basis whether the specified action accurately described what the unit was doing at the time. This score is called the “Task score” and is, like the position score and fire awareness intended to measure SA.

Secondly, the participants received a “Congruence score”. Regardless of whether they were correct or not they were awarded 1 “congruence point” if they both had used the same category for a unit, which would measure shared understanding. If they both used category 8 for “Do not know” however, they were not awarded any points.

As for the map-part of the questionnaire, for each square that was accurately marked as having been or currently being on fire, the participant received 1 point. For each square wrongly marked 1 point was removed. The total number of squares that had been or was on fire was counted and a percentage of correct “Fire Awareness” was calculated.

Further, the participants received 3 points for marking a vehicle correctly on the map, 2 points if they had marked the vehicle within 1 square of its actual location or a vehicle of the same type on the actual position and 1 point if they marked a vehicle within one square of the location of a vehicle of the same type. This system was developed to allow certain leniency. For example, many participants knew the approximate location of a fire truck but without knowing which of the four trucks it was, and this scoring system still awarded some points for marking a fire truck in the direct vicinity of another. This meant that the participants could get a total of 24 points if they accurately marked all vehicles at their correct positions. This score is called the “Position score”.

4 Results

This chapter is divided into four sections. The first is the 4.1 Data Overview section, which is a walkthrough of the collected data. The purpose of this section is not to answer any hypotheses or research questions, but to present a subset of the collected data in an easily understandable way to facilitate the understanding of later statistical tests.

The structure of the Data Overview section will be modeled roughly after in which order the data was collected, i.e. first the first DATMA questionnaire will be presented (Fire, Position and Task Awareness), followed by the remaining of the DATMA questionnaires (Individual and Team Workload, and Teamwork), then CARS and last the performance measure from C3Fire.

After the Data Overview section there will be three sections that aim to answer research questions 1 through 3 respectively. The first of these is 4.2 Shared Understanding Results, where research question 1 and hypotheses 1 and 2 will be answered. Next is 4.3 Effects of Mixed Conditions, which will answer research question 2 and the remaining hypotheses (4-6). This will be done by a Principal Component Analysis (PCA) performed on a large number of the collected variables that were presented in the Data Overview section, in order to reduce and distill them into fewer factors[3].

The last section, 4.4 Relations Between Factors, explores research question 3 by presenting a Structural Equation Model (SEM) of the casual relationships between the factors that were obtained from the PCA.

4.1 Data Overview

In this section the collected data is explained and initially explored, mainly visually by using graphs and charts, with some statistical tests to verify differences between conditions. The purpose of this section is to familiarize the reader with the data material and to show an overview with some general findings, not to analyze every variable contrasted to all others. Each set of data (for example the different questionnaires or C3Fire performance measure) is shown with the variables collected via that method and what the numbers mean. Most of the variables described in this section will later be used in the Principal Component Analysis and thus explored further in that manner.

4.1.1 Missing values

There was one missing value from one of the questionnaires from one participant. It was the Perception Process question from the CARS questionnaire from participant #3, in the second experiment, who had the role of Commander 1. The value was replaced by the average of that participant’s other answers on the same questionnaire, which was 1.571. No other missing values were found in this study.

[3] The correct term would be ”component” and not ”factors” as it strictly speaking is not a Factor Analysis but a PCA, but the term factors will be used in this thesis as this distinction is not always made.

4.1.2 Fire, Task and Position Awareness

The first questionnaire in the DATMA-battery asked the participants to mark the position of the fire, the units and also to determine what task the units were carrying out at the moment. This yielded three separate scores. The first is a "Fire Awareness" score that indicates how well the participants managed to correctly mark where there had been or currently was burning.

Table 4-1 Fire Awareness Scoring

	Mark	No Mark
Fire	1	0
No Fire	-1	0

One point was given for correctly marking the location of a current of former fire, while one point was deducted for false alarms (marking a square that had no fire). No points were awarded or deducted for correct rejections (not marking a square that was not burning) or misses (not marking a square that had fire), see table 4-1. This meant that a participant who indiscriminately marked many correct but also many false alarms could get a lower score than a conservative participant who marked a few correct and no false alarms. It also meant that participant could get a negative score. The maximum score for each individual scenario was also calculated, and the amount of points was then divided by the max to obtain a score where the max was 1. The result is summarized in the box-and-whiskers plot, as percentage correct, below (figure 4-1). Not displayed is an outlier at -133 in the GUI/Map condition and one at −166 in the Map/Map condition (the outliers were able to achieve such a low score simply by marking a great number of squares as being on fire when they were in fact not).

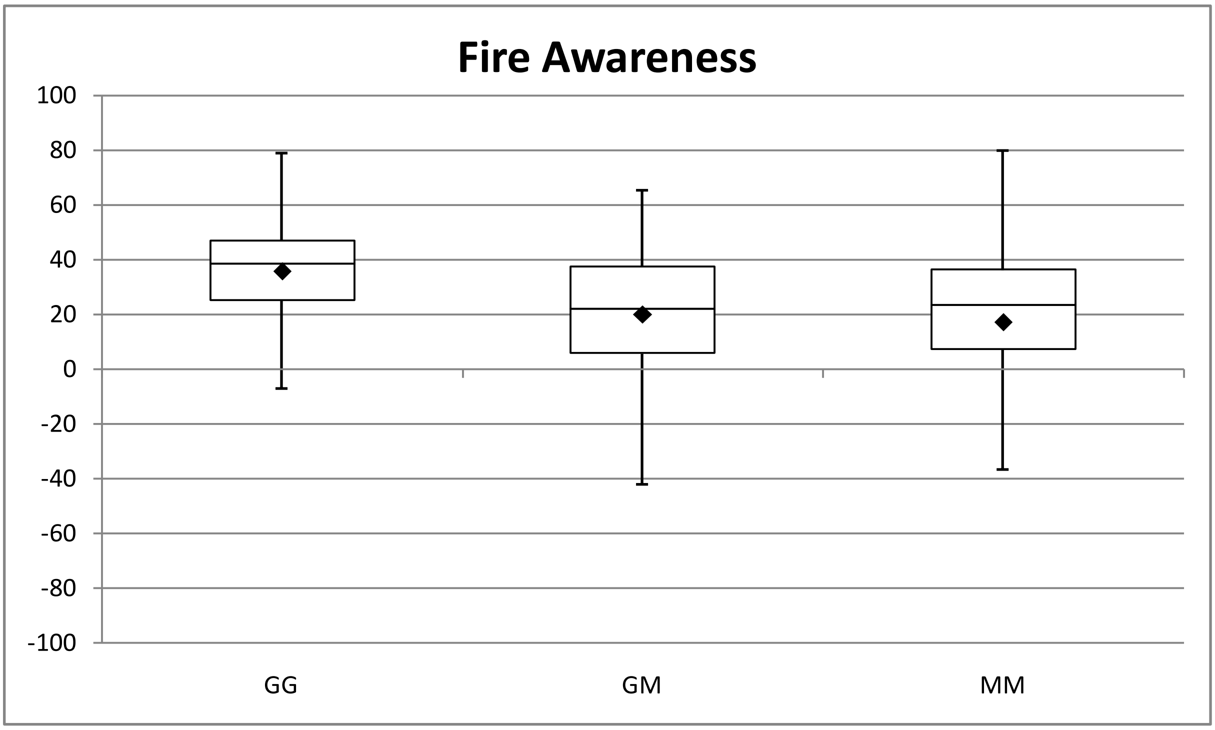


Figure 4-1 Fire Awareness Result

In this box-and-whiskers plot, and all others appearing in this thesis, both the average (black diamond; ♦) and the median (center bar in the box) as well as the first and third quartile (bottom and top bar of the box) are shown. The whiskers show the lowest and highest data point still within 1.5 of the IQR (inter-quartile length). If there are any outliers (i.e. a datum more than 1.5 IQR from the median) they are marked by a black dot and the whiskers show the maximum or minimum allowed value that is within 1.5 of the IQR.

A repeated measure ANOVA shows that there are significant differences between conditions. Greenhouse-Geisser correction was used since Mauchly's test of sphericity was significant (p=.044) which yielded $F(1.713, 59.965) = 7.712$, $p<0.01$. Sidak post-hoc test shows that GUI/GUI (mean 35.810, std. error 4.008) is higher than GUI/Map (mean 20.114, std. error 5.708), p=0.011, and Map/Map (mean 17.171, std. error 6.768), p=0.011. GUI/Map is not significantly different from Map/Map, p=0.860.

The second score was a task score that was based on the participant correctly marking the correct task that a unit was doing at that time. The max score was 8, one point for each unit, and the min 0. A box-and-whiskers plot is provided below (figure 4-2).

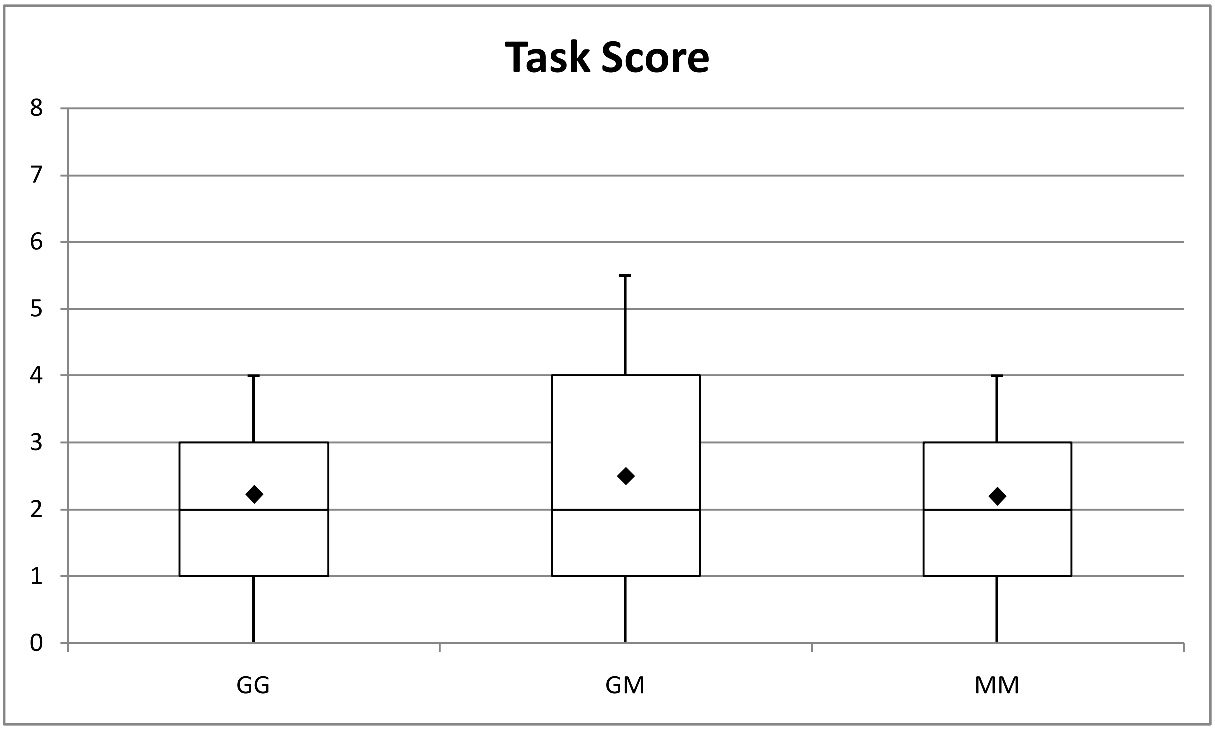


Figure 4-2 Task Score

In general the scores were low and the median for all conditions were 2 (out of 8 possible). There is no significant differences between any condition, $F(2, 70) = 0.391$, $p=0.68$.

The last score extracted from the questionnaire is a position score. The participants marked on the map the position of each vehicle and was given a score based on this. The participant was awarded one point for marking a vehicle within one square of another vehicle of the same type, an additional point if it was the same ID (i.e. the correct unit) and an additional point if the correct unit was marked on the correct spot. This meant that a maximum of 3 points per vehicle for a total of 24 points, see figure 4-3 for the distribution.

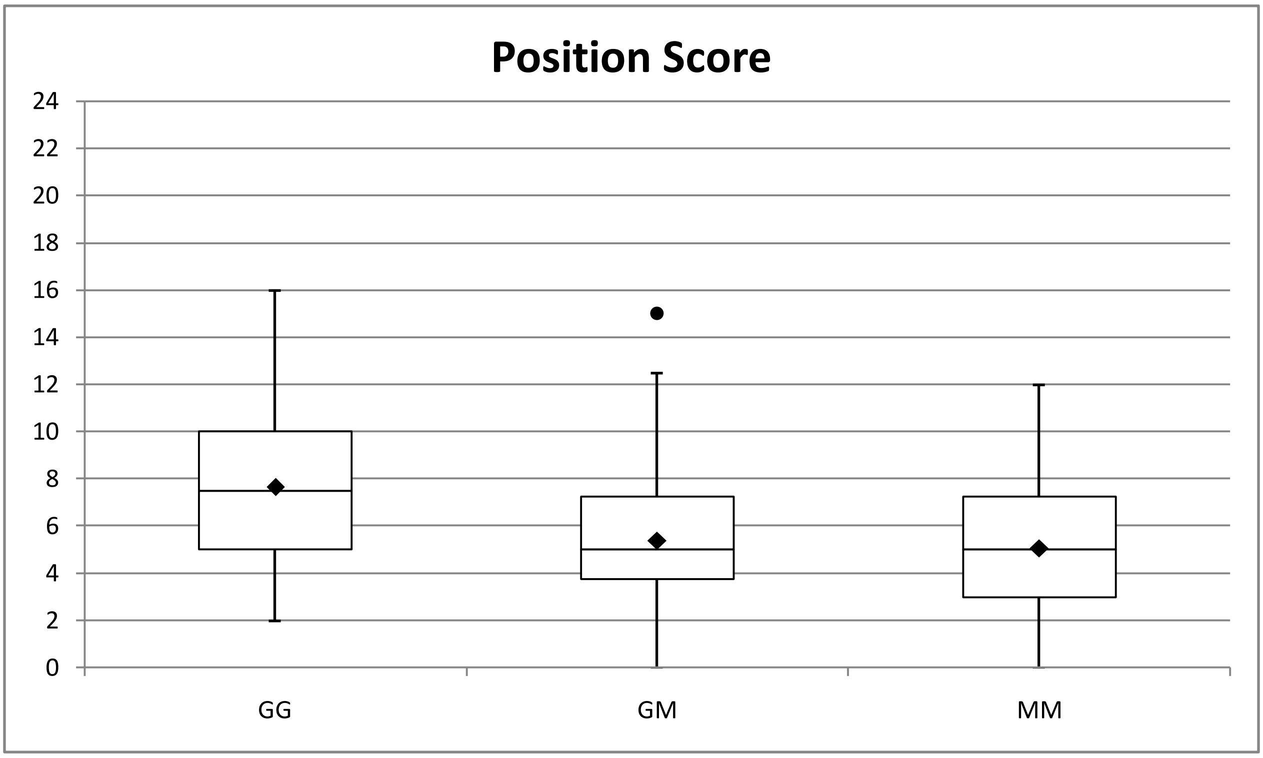


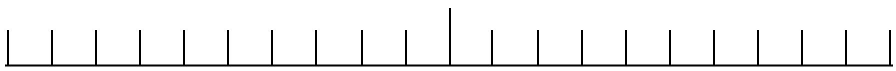
Figure 4-3 Position Score

A repeated measure ANOVA shows significant differences in the position scores between conditions, $F(2, 70) = 6.622$, $p<0.01$. Sidak post-hoc test shows that GUI/GUI (mean 7.667, std error 0.650) is higher than GUI/Map (mean 5.389, std. error 0.490), $p<0.05$, and Map/Map (mean 5.056, std. error 0.558), $p<0.01$. GUI/Map and Map/Map are not significantly different.

4.1.3 Individual Workload

Individual workload was measured by the first DATMA questionnaire distributed after the scenario. It consisted of five subscales between 0 and 100. An example question (translated from Swedish, original questionnaire can be found in appendix A) is shown below in figure 4-4 and all variables in this questionnaire is provided in table 4-2.

Mental Workload – How mentally taxing was the scenario ***for you personally***?



Very Low Very High

Figure 4-4 Individual Workload Example Question

Table 4-2 Individual Workload Questions

Variable	Purpose	Scale
Mental Workload	Subject rating of their own mental workload	0 (low workload) – 100 (high workload)
Time Pressure	Subject rating of their own time pressure	0 (low time pressure) – 100 (high time pressure)
Performance	Subject rating of their own performance	0 (poor performance) – 100 (high performance)
Effort	Subject rating of their own effort	0 (low effort) – 100 (high effort)
Frustration	Subject rating of their own frustration	0 (low frustration) – 100 (high frustration)

This data was collected from each participant after every scenario. It should be noticed that for all subscales but Performance, higher numbers indicate higher workload, time pressure etc. while the performance measure is inverted compared to the others (i.e. a low rating indicates poor performance). The ratings were then organized after the correct condition. A chart showing the average score for each rating is shown below (figure 4-5).

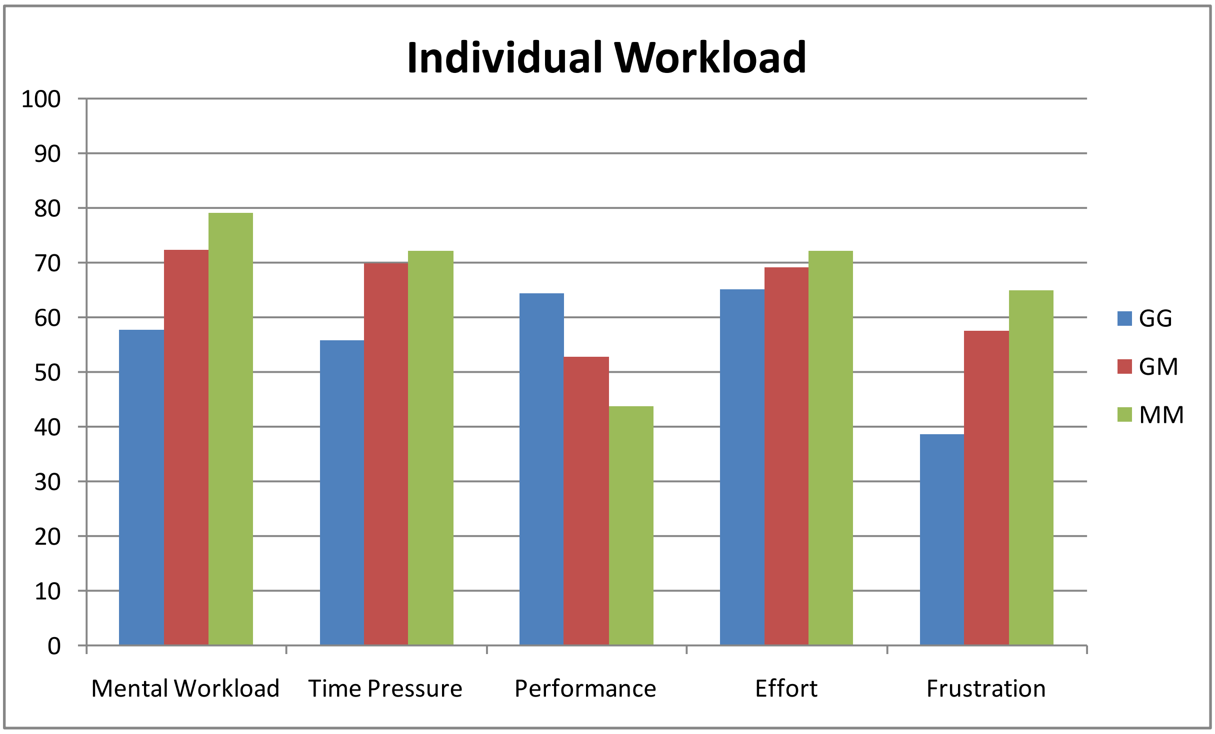


Figure 4-5 Individual Workload Result Overview

As can be seen in this chart, GUI/GUI (blue bar labeled GG) seem to lie below or well below GUI/Map (red bar labeled GM) and Map/Map (green bar labeled MM) in all cases but Performance. This indicates that the workload, time pressure etc. was lower in the GUI/GUI condition than the others, and that the performance was higher. There is also some difference between most GUI/Map and Map/Map ratings, with time pressure and effort being the ones with the least amount of difference. All variables but Performance were then summarized into an average for each participant to get a rough measure of the total workload. The result is shown in the box-and-whiskers plot, figure 4-6, below.

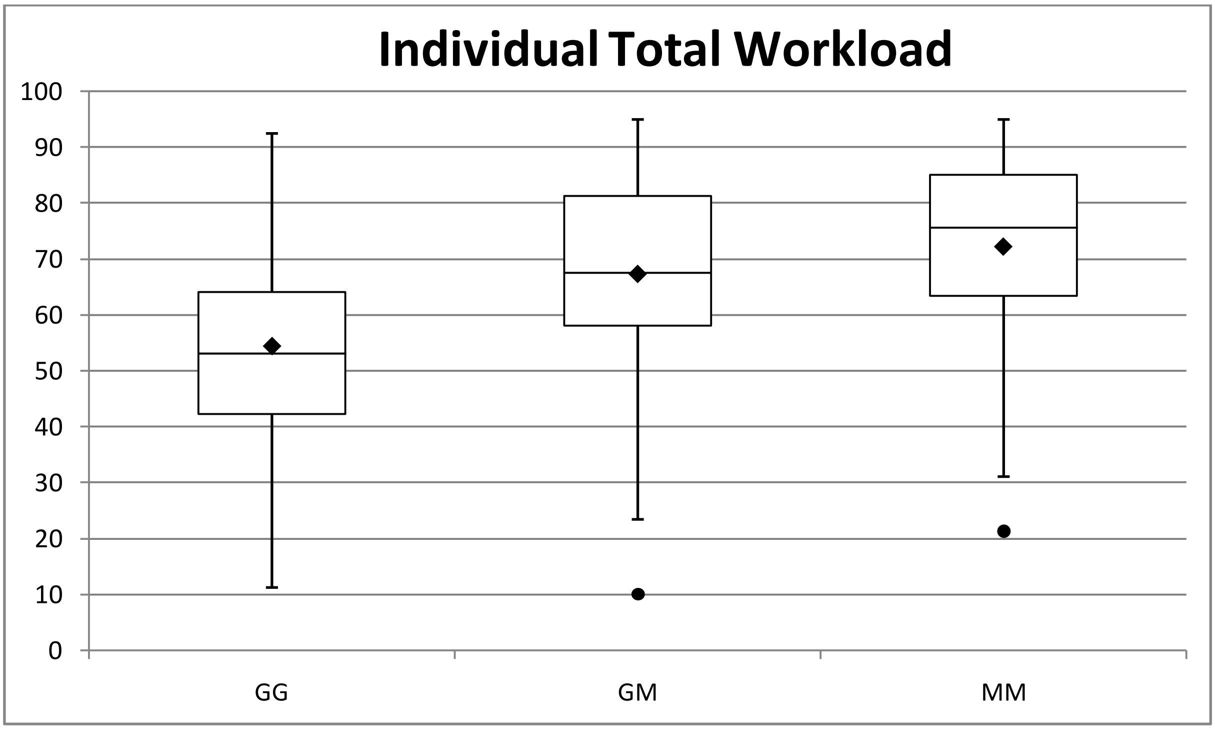


Figure 4-6 Individual Total Workload

A repeated measure ANOVA shows significant differences in average individual workload, $F(2, 70) = 13.174$, $p<0.001$. Sidak post-hoc test shows that GUI/GUI (mean 54.340, std error 3.207) is lower than GUI/Map (mean 67.257, std. error 2.897), $p<0.001$, and Map/Map (mean 72.153, std. error 2.851), $p<0.001$. GUI/Map and Map/Map are not significantly different.

4.1.4 Team Workload

Team workload was measured much in the same way as individual workload by a second DATMA questionnaire. The same five subscales were used and shown in a similar manner using the same scale as the individual workload questionnaire. An example subscale is shown below in figure 4-7 (translated from Swedish, original can be found in Appendix A) and all variables measured with the questionnaire are listed in table 4-3.

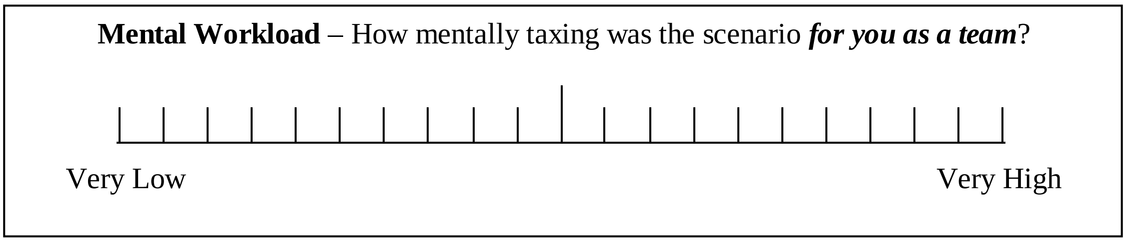


Figure 4-7 Team Workload Example Question

Table 4-3 Team Workload Questions

Variable	Purpose	Scale
Mental Workload	Subject rating of the mental workload of the team	0 (low workload) – 100 (high workload)
Time Pressure	Subject rating of the time pressure of the team	0 (low time pressure) – 100 (high time pressure)
Performance	Subject rating of the performance of the team	0 (poor performance) – 100 (high performance)
Effort	Subject rating of the effort of the team	0 (low effort) – 100 (high effort)
Frustration	Subject rating of the frustration of the team	0 (low frustration) – 100 (high frustration)

Each subject filled out the questionnaire by him/herself but was asked to rate the teams workload on the different subscales. Again, the data was collected and organized according to the conditions and the average rating on each subscale is shown in the chart, figure 4-8, below.

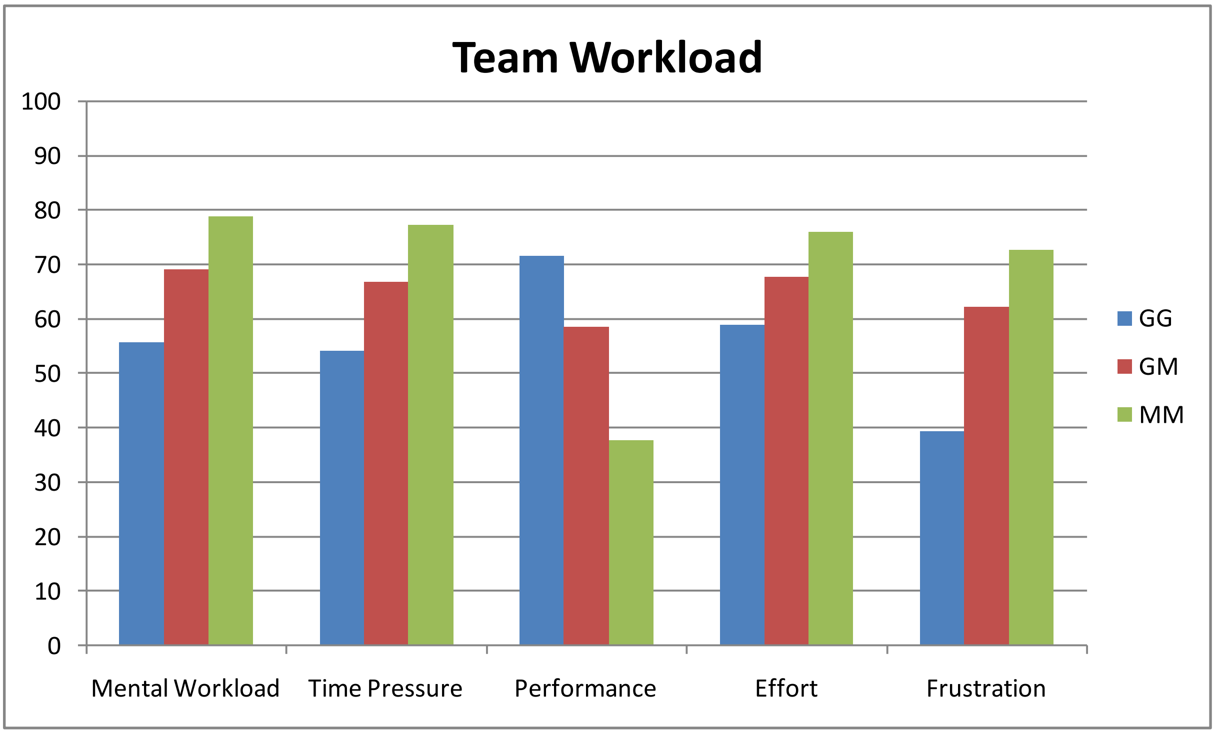


Figure 4-8 Team Workload Result Overview

The same trend that can be seen in the individual workload ratings can be found here. GUI/GUI is rated lower than GUI/Map and Map/Map on all scales but performance. The difference between GUI/Map and Map/Map is also clearer for the team workload ratings than the individual workload ratings. All subscales save Performance were then summarized into a box-and-whiskers plot, figure 4-9 shown below.

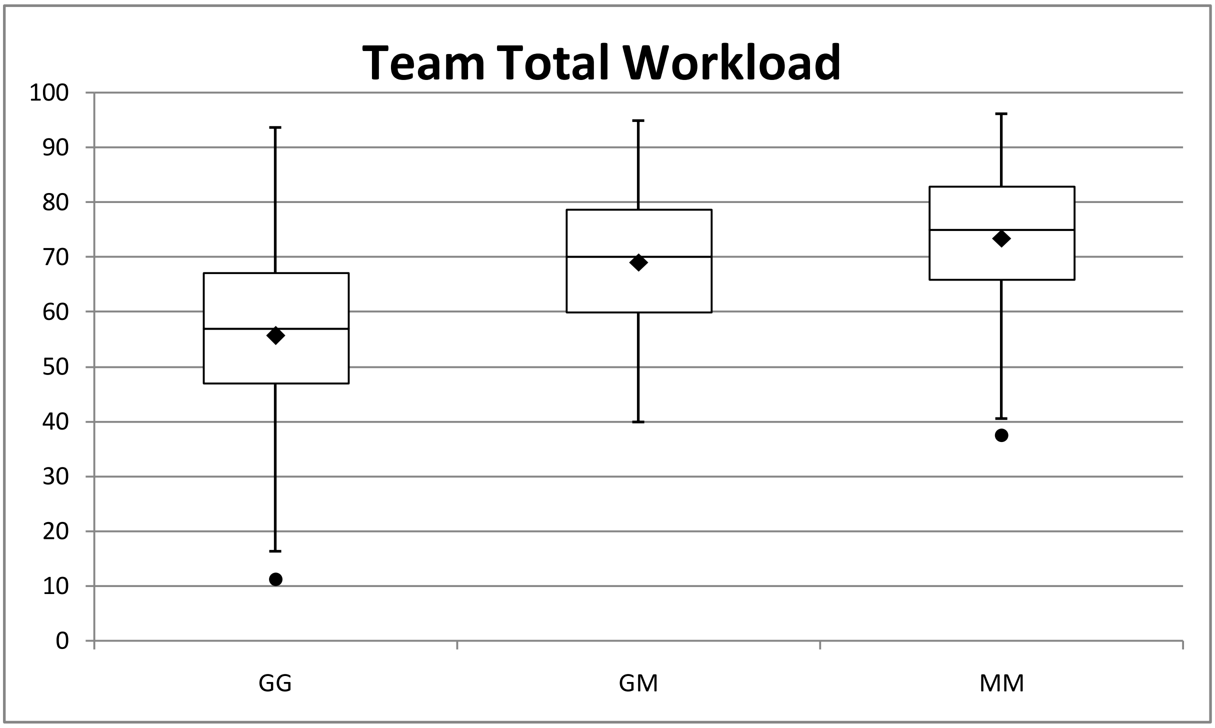


Figure 4-9 Team Total Workload

The box-and-whiskers plot shows that GUI/GUI has a larger spread than GUI/Map and Map/Map with a lower average.

A repeated measure ANOVA shows significant differences in average team workload, $F(2, 70) = 20.398$, $p<0.001$. Sidak post-hoc test shows that GUI/GUI (mean 55.625, std error 3.092) is lower than GUI/Map (mean 67.063, std. error 2.112), $p<0.001$, and Map/Map (mean 73.403, std. error 2.242), $p<0.001$. GUI/Map and Map/Map are not significantly different.

4.1.5 Teamwork

The teamwork questionnaire used four questions on a scale of 1 (poor) to 7 (excellent). The four questions concerned the communication, support, coordination and common operational picture of the team. The questions were Likert-scale with well-defined end points. The questions deviate slightly from the norm of Likert-scales where a single statement is used along with marked or unmarked options. Instead a question is asked and the end-points had defined end points that frame the scale, and the participants were asked to circle the appropriate number on the scale, see the example in figure 4-10 below (the example has been translated from Swedish, the original can be found in Appendix A) and all variables measured are listed in table 4-4.

1. Communication **To what extent did you proactively share important information without it being asked for?** **1** **2** **3** **4** **5** **6** **7** **Never** **Always** **7**: Important information was always shared without it being asked for **1**: Important information was never shared without it being asked for

Figure 4-10 Teamwork Example Question

Table 4-4 Teamwork Questions

Variable	Purpose	Scale
Communication	Subject rating of the team communication, low rating indicating little or no sharing of vital information between team members	1 (never sharing vital information) – 7 (always sharing vital information)
Support	Subject rating of the team support, low rating indicating that the team members never or rarely adjusted the workload between them to support each other	1 (No adjustment) – 7 (Always timely adjustment)
Coordination	Subject rating of the team coordination, low rating indicating that the team never or rarely worked together towards the same goal	1 (poor coordination) – 7 (excellent coordination)
Common operational picture	Subject rating of the common operational picture between team members, low rating indicating that team members never or rarely had the same understanding of the current tasks and goals	1 (Never agree on goals) – 7 (always agree on goals)

This questionnaire introduces a 7-point scale unlike the previous two who used an unmarked 0-100 scale. The chart shown in figure 4-11 has organised the average of each condition on the four questions.

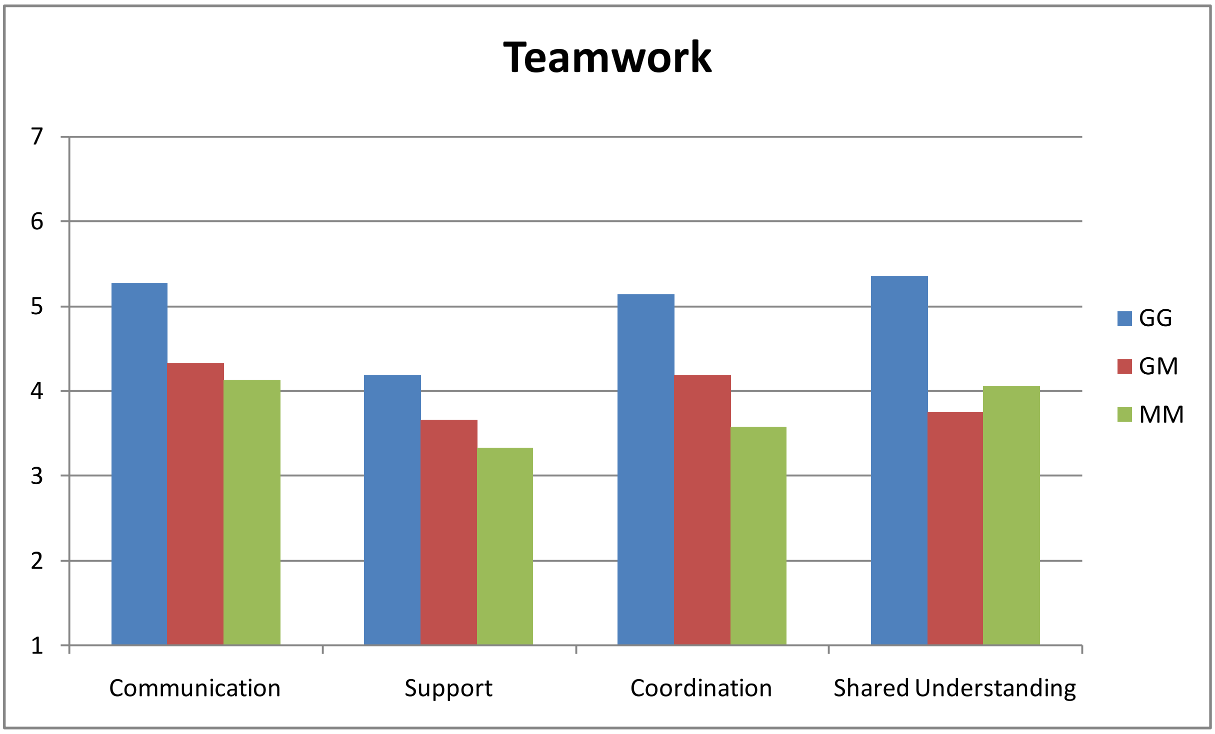


Figure 4-11 Teamwork Result Overview

GUI/GUI deviates from GUI/Map and Map/Map just as in the previous two charts. In this case, the higher bars indicate superior teamwork. The differences between GUI/Map and Map/Map are less clear and in the case of Common operational picture it is even the opposite of what has previously been shown. However, these differences are small, just as in previous charts, and when the teamwork scores are averaged into a single number, there is no difference between GUI/Map and Map/Map (see figure 4-12 below).

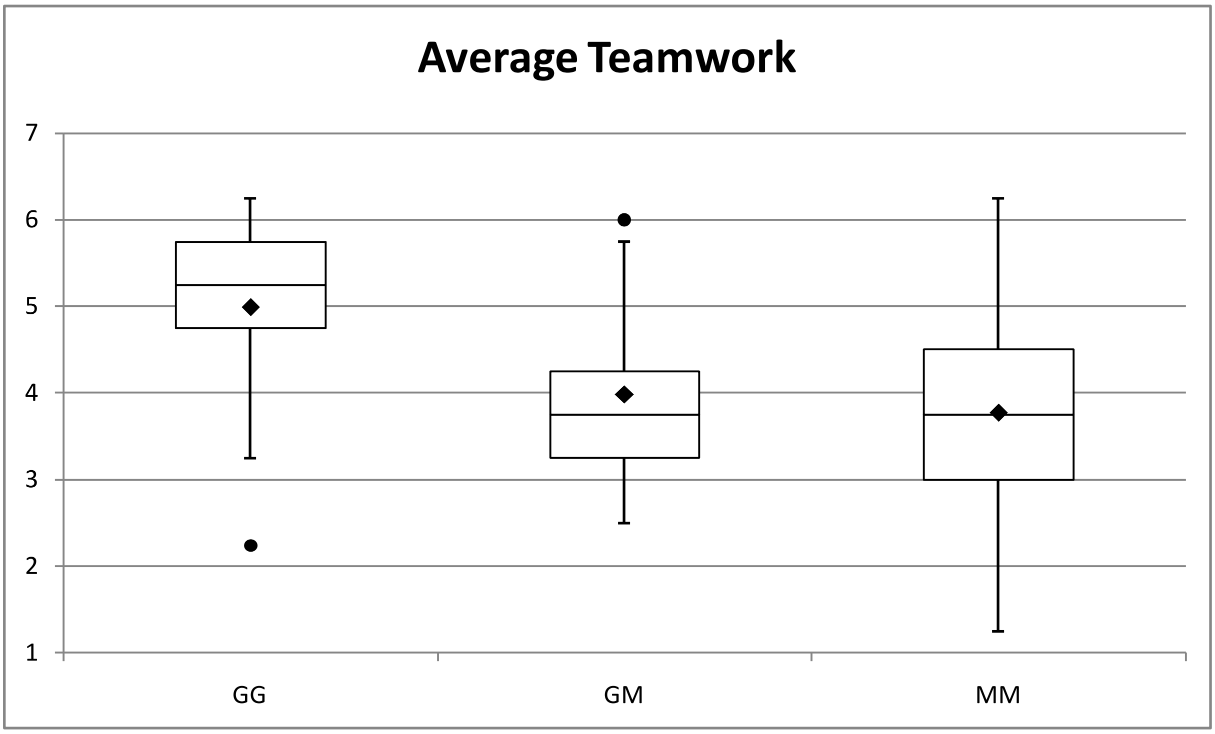


Figure 4-12 Teamwork

A repeated measure ANOVA shows significant differences in the average teamwork score. Mauchly's test of sphericity was significant (p=.021) so Greenhouse-Geisser correction was used, $F(1.663, 58.210) = 24.919$, $p<0.001$. Sidak post-hoc test shows that GUI/GUI (mean 4.993, std error 0.159) is higher than GUI/Map (mean 3.986, std. error 0.152), $p<0.001$, and Map/Map (mean 3.778, std. error 0.183), $p<0.001$. GUI/Map and Map/Map are not significantly different, p=.73.

4.1.6 CARS

The CARS questionnaire uses four scales with two aspects each to investigate Situation Awareness. The four scales are Perception, Comprehension, Projection and Integration which each refers to a different level of SA according to Endsley's (1995) model. The two different aspects are the *content* of each level and the *process* of each level, where the content mainly refers to the current information possessed whereas the process refers to the process of gathering information. An example question is provided below in figure 4-13 and all variables measured are listed in table 4-5.

1. **Perception – Gathering Information**

Content: how reliable and accurate is the information you have now? (E.g. fire/unit positions)

(Best case) 1 2 3 4 (Worst Case)

Figure 4-13 CARS Example Question

Table 4-5 CARS Questions

Variable	Purpose	Scale
Perception Content	Subject rating of the reliability and accuracy of his/her current information	1 (Best possible case) – 4 (Worst possible case)
Perception Process	Subject rating of the difficulty in maintaining his/her current level of information	1 (Best possible case) – 4 (Worst possible case)
Comprehension Content	Subject rating of the reliability and accuracy of his/her current understanding of the situation	1 (Best possible case) – 4 (Worst possible case)
Comprehension Process	Subject rating of the difficulty in maintaining his/her current level of understanding	1 (Best possible case) – 4 (Worst possible case)
Projection Content	Subject rating of the reliability and accuracy of his/her current predictions of future events	1 (Best possible case) – 4 (Worst possible case)
Projection Process	Subject rating of the difficulty in making predictions of future events	1 (Best possible case) – 4 (Worst possible case)
Integration Content	Subject rating of the reliability and accuracy of his/her integration of his/her actions with the aforementioned three aspects	1 (Best possible case) – 4 (Worst possible case)
Integration Process	Subject rating of the difficulty of integrating his/her actions with the aforementioned three aspects	1 (Best possible case) – 4 (Worst possible case)

Here a 4-point scale is used, where 1 indicates the best possible scenario. This means that the lower the score, the better. The data from the experiments were compiled using the average for each condition, see the chart in figure 4-14 below.

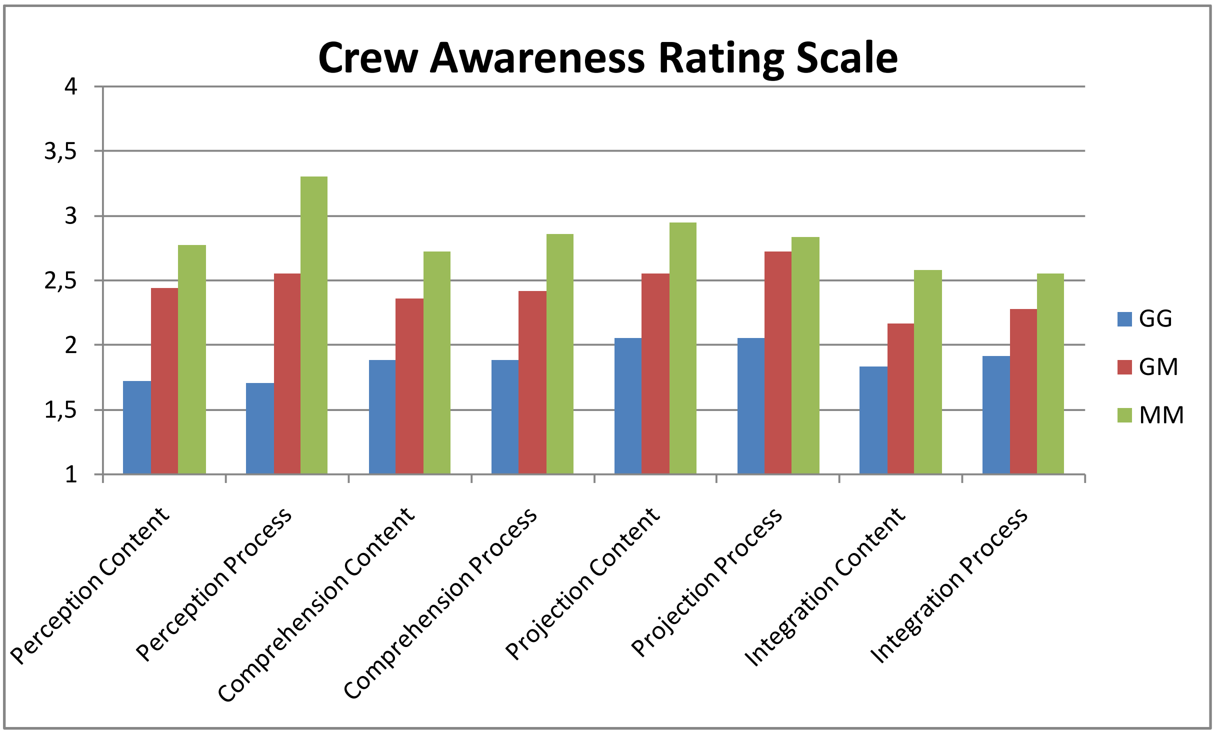


Figure 4-14 CARS Result Overview

It is clear from the chart that GUI/GUI is consistently ranked lower than the other two conditions, with GUI/Map being the second highest and Map/Map the highest. This data is averaged together in the box-and-whiskers plot (figure 4-15 below) to show the differences more clearly.

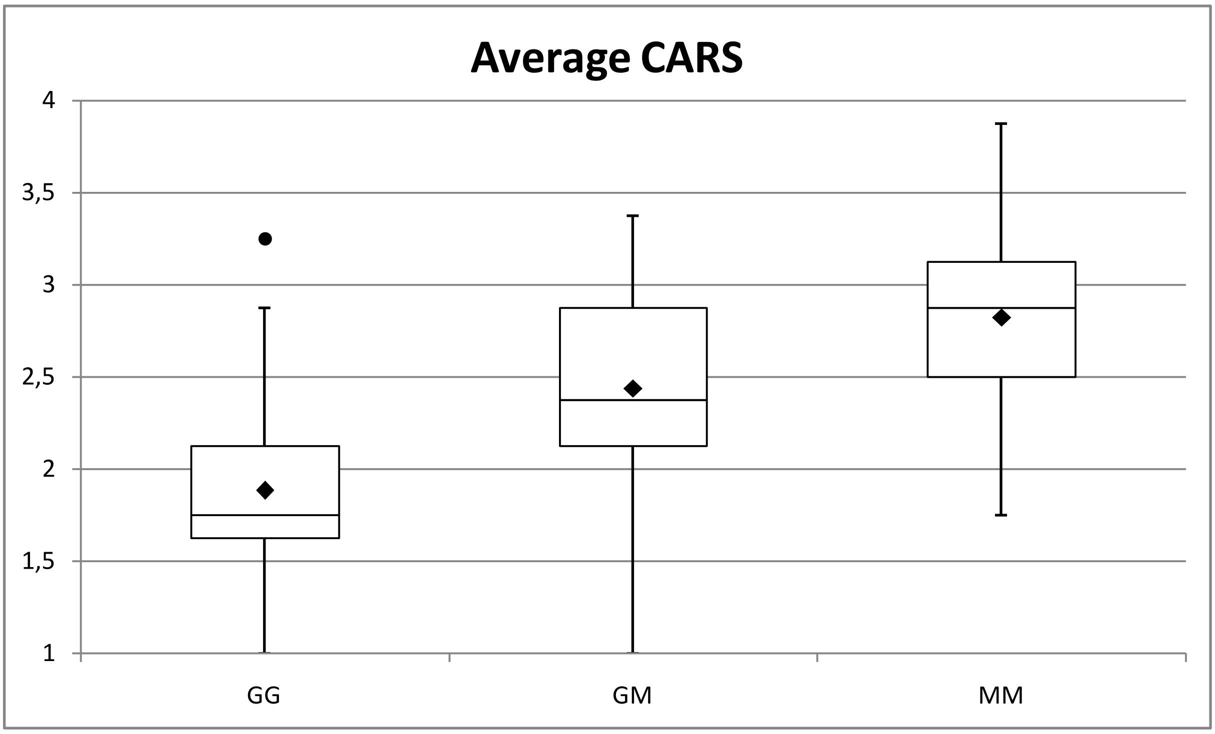


Figure 4-15 CARS

Repeated measure ANOVA shows significant differences in the average CARS rating, $F(2, 70) = 40.535$, $p<0.001$. Sidak post-hoc test shows that GUI/GUI (mean 1.884, std. error 0.078) is ranked lower than GUI/Map (mean 2.438, std. error 0.083), $p<0.001$, and Map/Map (mean 2.823, std. error .082), $p<0.001$. GUI/Map is also significantly lower than Map/Map, $p<.01$.

4.1.7 C3Fire Performance

C3Fire Performance was calculated by adding all the points that was given to a team, 1 point for each burned out or burning terrain square plus an additional 2 points for each house or structure that occupied a burning or burned out square. This score was then divided by the maximum score (which was the amount of points that would have been awarded had no fire been extinguished at any point in the scenario), which gave a value between 0 and 1 where a low number indicated that a large amount of squares that could have burned down was saved. This score was then reversed (by subtracting the score from 1) for ease in data interpretation since it made a high score equal high performance. Thus, the final performance score indicates, in percent, the amount of squares that was saved. The box-and-whiskers plot below (figure 4-17) shows the distribution of performance scores across conditions.

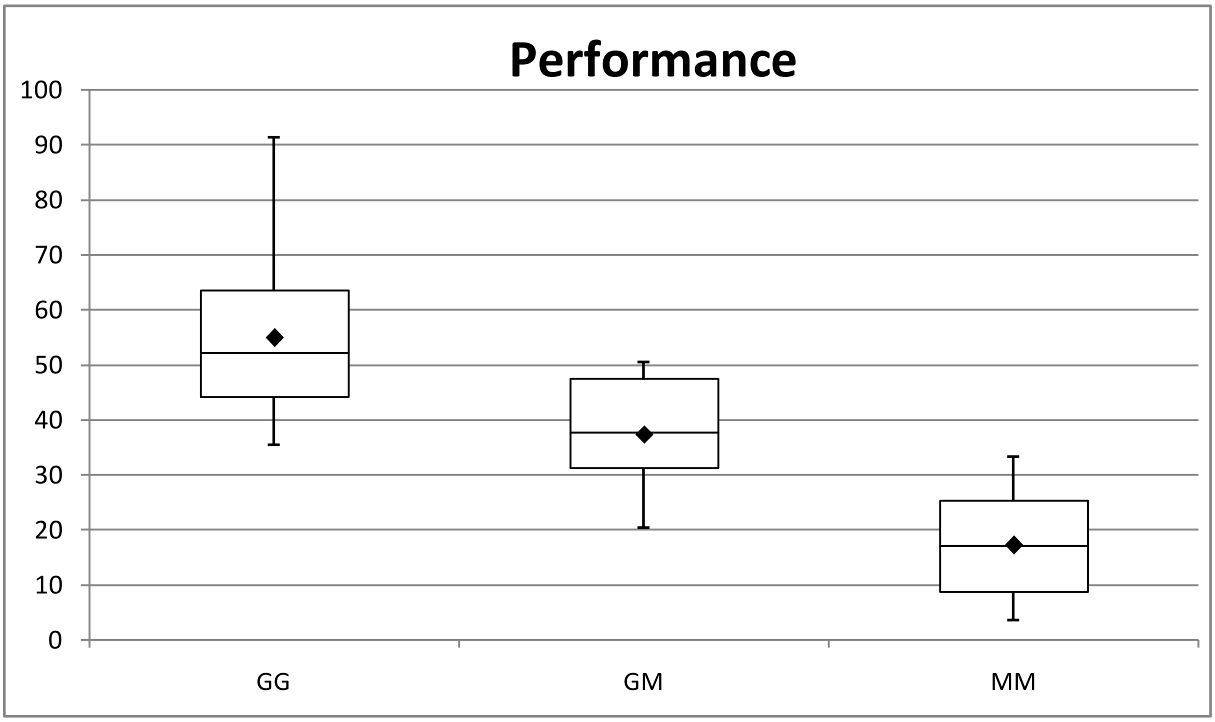


Figure 4-16 Performance Overview

It is worthy of note that the boxes are rather compact with short whiskers, indicating that the values are tightly grouped. Also, no whisker or outlier is at 0 or 100, which shows that no team completely failed to extinguish any fire and no team managed to extinguish all fire. Further, there is a clear difference between conditions and a repeated measure ANOVA shows significant differences. Greenhouse-Geisser correction was used since Mauchly's test of sphericity was significant (p=.002) which yelded $F(1.533, 53.66) = 124.193$, $p<0.001$. Sidak post-hoc test shows that GUI/GUI (mean 54.839, std. error 2.313) is higher than GUI/Map (mean 37.336, std. error 1.580), $p<0.001$, and Map/Map (mean 17.139, std. error 1.577), $p<0.001$. GUI/Map is also higher than Map/Map, $p<0.001$.

4.2 Shared Understanding Results

In this section hypotheses 1 and 2 will be answered by calculating Kendall's W on the SP-data and repeated measures ANOVAs on the discrepancy scores from the different DATMA questionnaires.

4.2.1 Hypothesis 1: Shared Understanding by Shared Priorities

The shared priorities measurement consisted of two sets of five items, each one created by one team member, where each item in the set had two ratings between 1 and 25, one from each team member. Kendall's W was calculated between these sets in order to determine the concordance. The items were organised after their order without taking the rating into account, which gave each set item a number 1 to 5. The ratings were not used as Kendall's W could not be used with them. Next, the average sum ranking was calculated (6) and the maximum sum of squared deviations, which is $(K^2(N^3\text{-}N))/12 = (2^3(5^3\text{-}5)/12 = 40$ where K is the number of rates (subjects) and N the number of items. Kendall's W is then calculated by dividing the individual sum of square deviations for each rank, by adding the squared differences of the average sum rankings (6) with the sums of the ratings of each item, with the maximum sum of squared deviations (40) (Norman & Streiner 2003). The resulting value between 0 and 1 indicates the concordance, or general agreement, between the rates. A value of 1 indicates that the rates agree with each other fully, a value of 0 is complete disagreement and .5 indicates that there is no association between rates. The results of the calculations are summarized in the box-and-whiskers plot (figure 4-17) below.

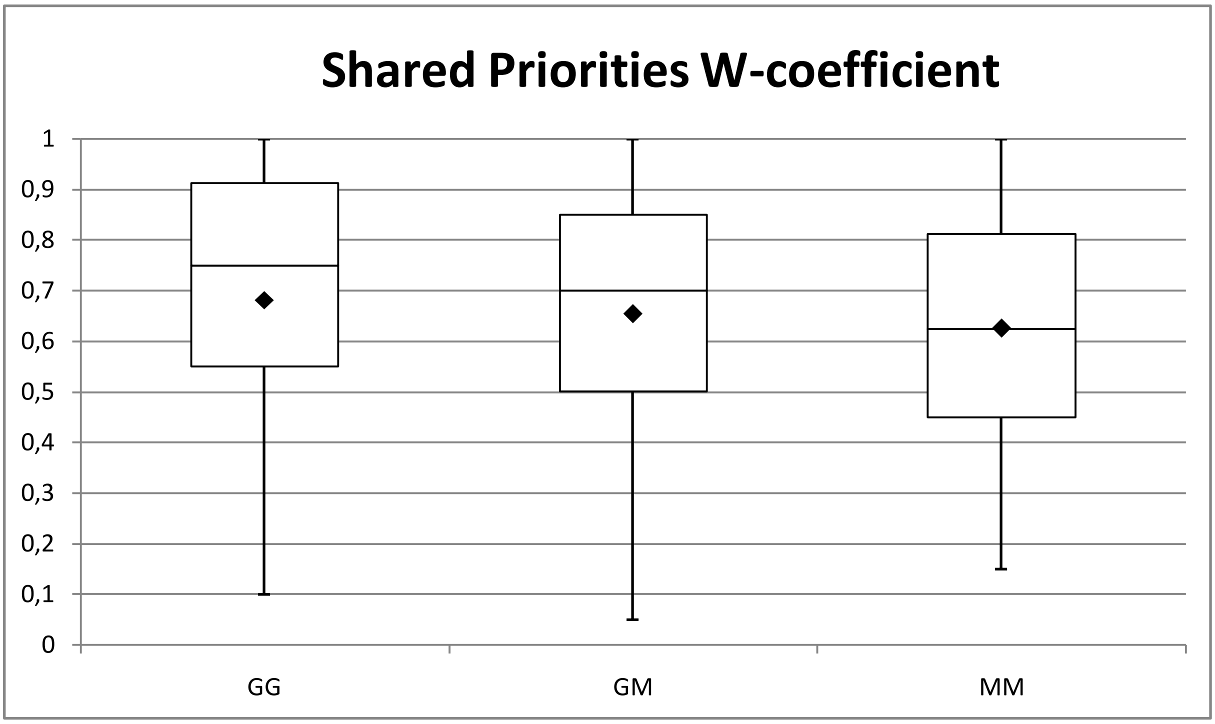


Figure 4-17 Shared Priorities

Repeated measures ANOVA shows no significant differences between any of the conditions for the Shared Priorities W-coefficient. Thus, the null hypothesis of hypothesis 6 cannot be rejected, see table 4-6.

Table 4-6 Hypothesis 1 Answered

Hypothesis 1: Mixed conditions will have an impact on the shared understanding of the team as measured by Shared Priorities
H_0: There is no difference in the Shared Priorities measure between conditions.

4.2.2 Hypothesis 2: Shared Understanding by DATMA

DATMA approaches shared understanding by looking at discrepancies in ratings of workload, teamwork and taskwork. This is similar to the approach of Shared Priorities in saying that discrepancy between team members is a measure on shared understanding. Thus, there will be three different measures of shared understanding by DATMA; discrepancies in workload, teamwork and taskwork.

The workload discrepancies comes from the Team Workload Awareness questionnaire and is calculated by taking the absolute value of the difference between the two team members ratings of each subscale. Repeated measure ANOVAs was calculated on the results. It was shown that there was a significant discrepancy in the Team Mental Workload subscale, $F(2, 70) = 5.383$, $p<0.001$, where GUI/GUI had higher discrepancy (mean 18.611, std. error 1.822) than GUI/Map (mean 10.278, std. error 1.605), $p<0.1$. There was also a significant difference in the Team Time Pressure subscale. Mauchly's test of sphericity was significant, $p=.002$, so Greenhouse-Geisser correction was used, $F(2, 53.87) = 3.377$, $p=0.053$ (a more liberal correction using Huyn-Feldt gives $p=0.051$). Sidak post-hoc test indicates that the difference is between GUI/GUI (mean 22.778, std. error 2.792) and Map/Map (mean 15.555, std. error 1.688), $p=.063$. The remaining subscales (Team Performance, Team Effort and Team Frustration) had no significant differences.

As for the Teamwork discrepancies, only communication had significant differences between conditions. Greenhouse-Geisser correction was used since Mauchly's test of sphericity was significant ($p<.001$) which yielded $F(2, 51.441) = 4.458$, $p<0.05$, and Sidak post-hoc test indicates that GUI/GUI (mean 15.873, std. error 2.261) was lower than Map/Map (mean 27.778, std. error 3.730), $p=.052$.

Last, a repeated measure ANOVA shows significant differences in the Taskwork Congruency scores between conditions. Greenhouse-Geisser correction was used since Mauchly's test of sphericity was significant ($p=.005$) which yelded $F(2, 55.222) = 10.865$, $p<0.01$. Sidak post-hoc test shows that GUI/GUI (mean 1.611, std error 0.170) and GUI/Map (mean 1.556, std. error 0.227) are higher than Map/Map (mean 0.611, std. error 0.140), $p<0.01$. GUI/GUI and GUI/Map are not significantly different. However, the scores indicate overall poor performance on the Taskwork Congruency, which has a max of 8 points.

In conclusion, the different shared understanding measurements shows that there are some differences between some, but not all, conditions. Thus, H_0 and H_a are rejected in favor of H_{a2}, with the caveat that only some differences could be measured. This is summarized in table 4-7.

Table 4-7 Hypothesis 2 Answered

Hypothesis 2: Mixed conditions will affect team mutual awareness as measured by DATMA.
H_{a2}: There is a difference in team mutual awareness as measured by DATMA in some but not all of the three conditions.

4.3 Effects of Mixed Conditions

This section explores hypotheses 3-6 via a Principal Component Analysis (PCA). The PCA was carried out in order to reduce the data set into a fewer numbers of factors.

4.3.1 The PCA

The PCA was conducted on 24 of the dependent variables from the different questionnaires, table 4-18, using an oblique Promax rotation with Kaiser normalization. A few measure variables, such as Fire Awareness, Task Score and Position Score, was included in a first iteration of the PCA but later excluded due to them not correlating with any other variables. The Kaiser-Meyer-Olkin test was used to verify the sampling adequacy, KMO = .85 which is “great” (Field 2009) and above the minimum acceptable threshold of .5 (Field 2009). The individual KMO values were all above the .5 threshold, with the lowest value at .68 and most (21 out of 24) above .80. Bartlett’s test of sphericity, χ^2 (276) = 1808.65, p<0.001, shows that the correlations between items were sufficiently large for the PCA (Field 2009).

A threshold for the Eigenvalues of 1 and a scree-plot, observing the point of inflexion, was used to extract 5 factors. Component loadings below .4 were suppressed when determining the factors. Figure 4-18 below show the loadings, after rotation, above .4 with the highest values for each variable marked.

Pattern Matrix					
	Factor				
	Workload	SA	Performance	Ext. Factors	Teamwork
Individual Effort	.864				
Individual Timepressure	.772				
Individual Mental Workload	.733		.417		
Team Mental Workload	.728				
Team Effort	.724				
Integration Content		.938			
Integration Procedure		.876			
Projection Procedure		.857			
Comprehension Procedure		.731			
Projection Content		.639			
Perception Procedure		.537			
Comprehension Content		.529			
Team Performance			-.948		
Individual Performance			-.901		
Perception Content			.563		
C3Fire-Performance			-.498		
Estimated Teammate Mental Workload				.784	
Team Frustration				.656	
Common operational picture				-.554	
Team Timepressure	.523			.534	
Individual Frustration				0.372	
Support					.877
Communication					.604
Coordination				-.449	.535

Figure 4-18 PCA Factors

Only one variable, Individual Frustration, was below the .4 threshold, but had a highest value close enough to be included in a factor. The first factor included three questions from the individual workload questionnaire (DATMA 1); Individual Effort, Individual Time Pressure and Individual Mental Workload. It also included two questions from the second DATMA questionnaire, concerning team workload; Team Mental Workload and Team Effort. One more variable, Team Time Pressure, had a high factor loading for this factor, .523 compared to the highest .534, and the Individual Mental Workload variable was also high for the third factor (Performance). This first factor was concluded to be a Workload measure, as it draws from questionnaires built to measure (cognitive) workload, both for the team and individual level.

The second factor had 7 out of the 8 questions on the CARS measure (all but Perception Content, the very first question corresponding to the lowest SA-level) and was therefore labelled an SA-measure.

The third factor included the two performance measures from the DATMA questionnaires, Team Performance and Individual Performance, as well as Perception Content from the CARS questionnaire and the C3Fire-Performance measure (number of saved squares). This factor is a clear Performance measure, and the inclusion of Perception Content shows the tight coupling between the ability to perceive vital information and success at reaching the scenario goal.

The next factor included the estimate of the Team Mate Mental Workload, the Team Frustration, Common operational picture (one of the teamwork questions from DATMA), Team Time Pressure and Individual Frustration. Since these all seem to be variables heavily influenced by outside sources (for instance whether the team mate is using GUI or map), it was labelled an “External Factors” factor.

The last factor consists of three out of four teamwork questions from the teamwork questionnaire; Support, Communication and Coordination. All of these variables are meant to measure teamwork, and the factor was thus concluded to be a Teamwork measure.

4.3.2 Factorial Split-Plot ANOVAs on Factors

All factors were aligned so that a low score was “good” and a high “poor”, thus a low score on workload indicated low workload, a low score on performance meant a strong performance while a high score would be a poor performance, a low score on SA indicates what is normally described as good, or “high”, SA. This was done to ease the comparison of the different scores, and to nullify the differences that comes from the questionnaires since the scales were different (i.e. in some cases a high score was a good thing, in some it was not). The only questionnaire that used a high score to indicate something positive was the Teamwork questionnaire (where 7 was good teamwork) and the (inverted) C3Fire-performance measure. Thus, these four questionnaire items were reversed (thus 7 would indicate poor teamwork) and the non-reversed C3Fire-performance measure was used. This created factors where a low score was “good” and a high score was “poor”. An overview chart, figure 4-19, displays all 5 factors below.

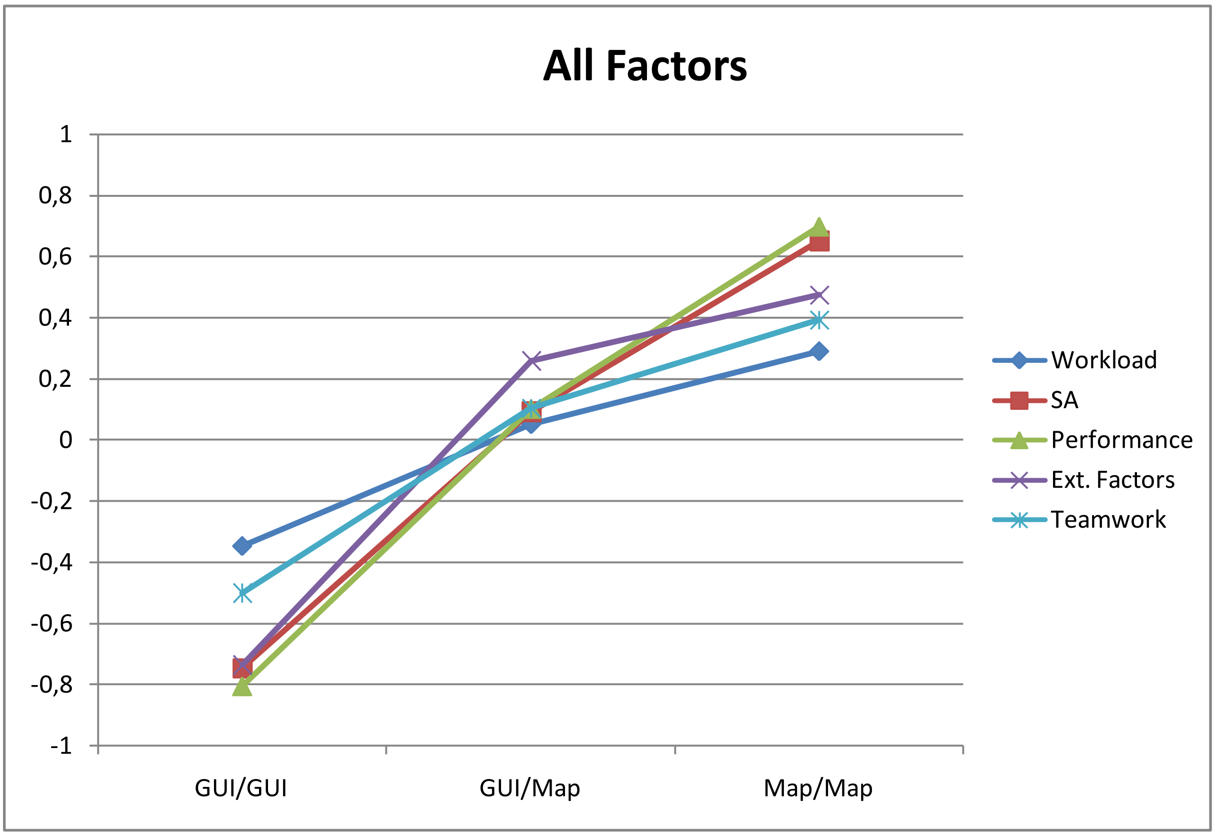


Figure 4-19 PCA Factors Overview

Note how all factors are low in the GUI/GUI condition, rising to the GUI/Map condition and having their highest score in the Map/Map condition. A number of 2 (roles) x 3 (conditions) factorial split-plot ANOVAs were carried out to determine, mainly, the differences in the factors between the conditions, but also to investigate whether there were any differences in the factor scores for the two roles used and any interaction effects between these two variables. The results are presented in the next sections.

4.3.2.1 Hypothesis 3: Workload

The ANOVA on the Workload factor confirms H_{a2} of hypothesis 3 (see table 4-9) while rejecting H_0 and H_a.

Table 4-8 Hypothesis 3 Answered

Hypothesis 3: Mixed conditions will affect mental workload.
H_{a2}: There is a difference in subject and team mental workload in some but not all of the three conditions.

There are differences in Workload between some of the three conditions, $F(2, 68) = 7.83$, $p<0.001$. Sidak post-hoc test shows that GUI/GUI has lower workload than GUI/Map (mean difference -0.398, $p<0.05$) and Map/Map (mean difference -0.637, $p<0.01$). There are no significant differences between GUI/Map and Map/Map in workload, $p=.396$.

There are no significant differences in workload between the roles of the participants, $F(2, 34) = 1.09$, $p=0.304$.

There are significant interaction effects between different combinations of conditions and roles, $F(2, 68) = 5.93$, $p<0.01$. Tukey post-hoc test shows that workload is lower for Commander 1 in GUI/Map than Map/Map (mean difference -0.769, $p<0.05$) and lower in GUI/GUI than Map/Map (mean difference -1.06, $p<0,05$). Workload was also significantly higher for Commander 2 than Commander 1 in GUI/Map, $F(1, 102) = 4.755$, $p<0.05$. Figure 4-20 illustrates this.

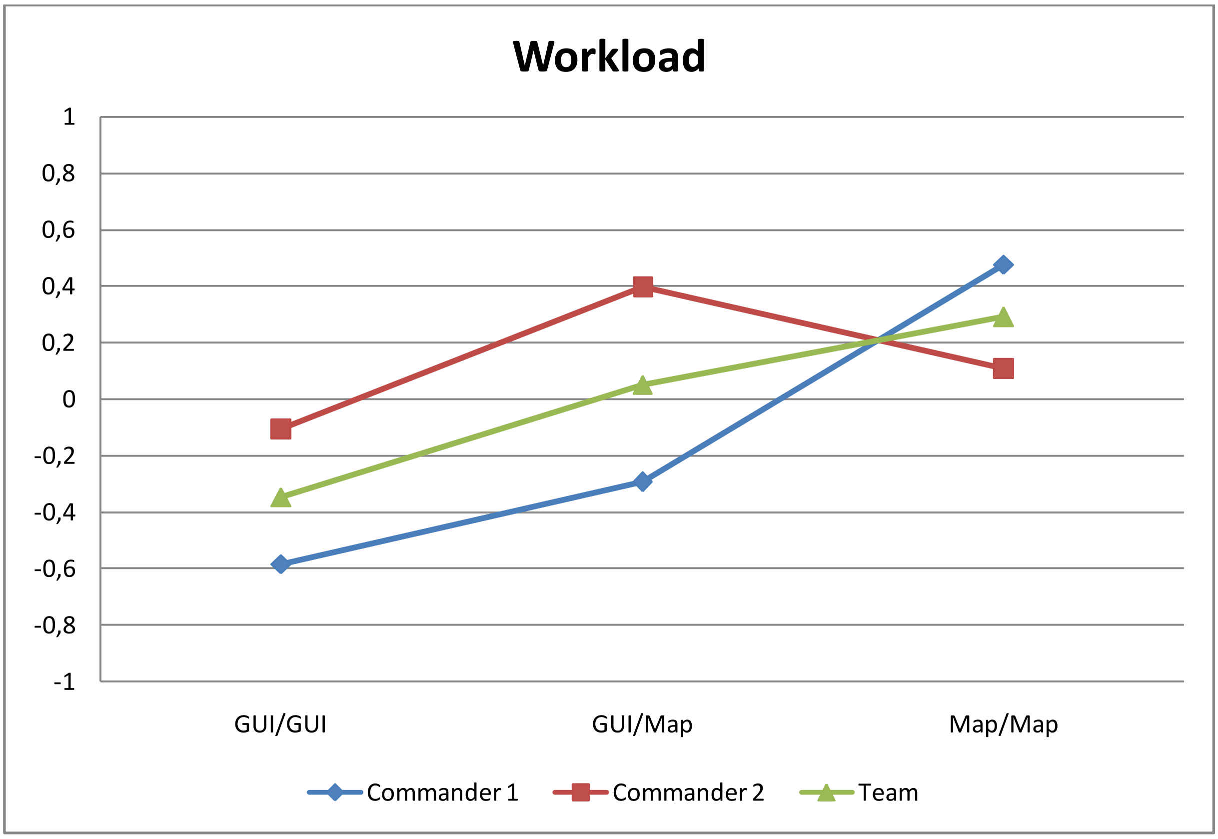


Figure 4-20 Workload Factor

4.3.2.2 Hypothesis 4: SA

The ANOVA on the SA factor confirms H_a for hypothesis 4 and rejects H_0 and H_{a2}.

Table 4-9 Hypothesis 4 Answered

Hypothesis 4: Mixed conditions will affect Situation Awareness (SA).
H_a: There is a difference in SA between all the three conditions.

There are differences in SA between all three conditions, $F(2, 68) = 36.54$, $p<0.001$. Sidak post-hoc test shows that GUI/GUI has better (i.e. lower measurement on) SA than GUI/Map (mean difference -0.838, $p<0.001$) and Map/Map (mean difference -1.396, $p<0.001$) and GUI/Map has better SA than Map/Map (mean difference -0.558, $p<0.01$).

There are no significant differences in SA between the roles of the participants, $F(2, 34) = 0.37$, $p=0.55$.

There are significant interaction effects between different combinations of conditions and roles, $F(2, 68) = 4.13$, $p<0.05$. Tukey post-hoc test shows that SA is better for Commander 1 in GUI/Map than Map/Map (mean difference -1.03, $p<0.05$) and better in GUI/GUI than Map/Map (mean difference -1.644, $p<0,05$) but showed no differences between GUI/GUI and GUI/Map. Tukey post-hoc test also shows that SA is better for Commander 2 in GUI/GUI than GUI/Map (mean difference -1.06, $p<0.05$) and Map/Map (mean difference -1.14, $p<0.05$) but showed no difference between GUI/Map and Map/Map. SA was also significantly better for Commander 1 than Commander 2 in GUI/Map, $F(1, 102) = 4.64$, $p<0.05$, see figure 4-21.

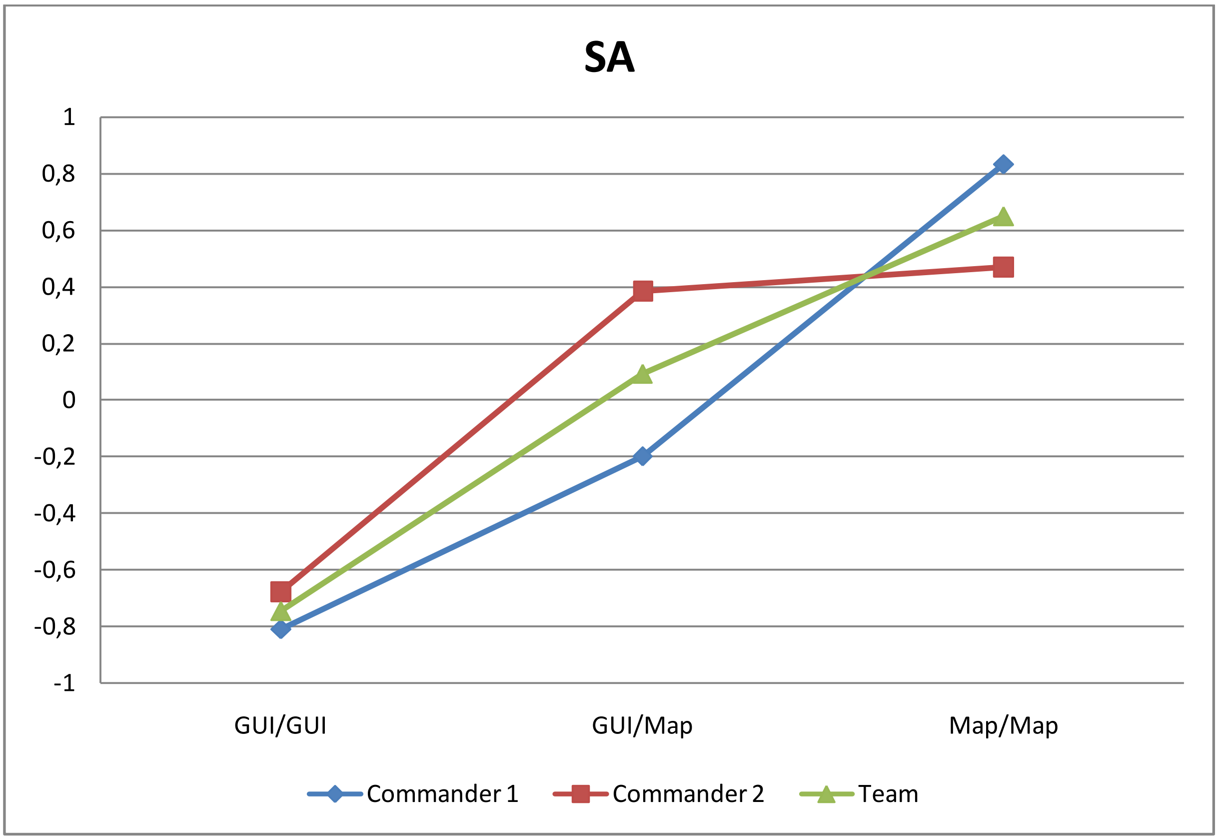


Figure 4-21 SA Factor

4.3.2.3 Hypothesis 5: Teamwork

H_{a2} of hypothesis 5 is confirmed by the ANOVA on the Teamwork factor, while H_0 and H_a are rejected.

Table 4-10 Hypothesis 5 Answered

Hypothesis 5: Mixed conditions will affect teamwork.
H_{a2}: There is a difference in teamwork in some but not all of the three conditions.

There are differences in Teamwork (illustrated in fig. 4-22) between some of the three conditions, $F(2, 68) = 16.11$, $p<0.001$. Sidak post-hoc test shows that GUI/GUI has better (i.e. lower) Teamwork score than GUI/Map (mean difference -0.607, $p<0.01$) and Map/Map (mean difference -0.896, $p<0.01$). There were no significant differences between GUI/Map and Map/Map in Teamwork score.

There are significant differences between the roles of the participants, $F(2, 34) = 10.27$, $p<0.01$. Sidak post-hoc test shows that Commander 2 had better (i.e. lower) Teamwork score than Commander 1 over all conditions (mean difference -0.716, $p<0.01$).

There are no significant interaction effects between different combinations of conditions and roles, $F(2, 68) = 0.56$, $p=0.55$.

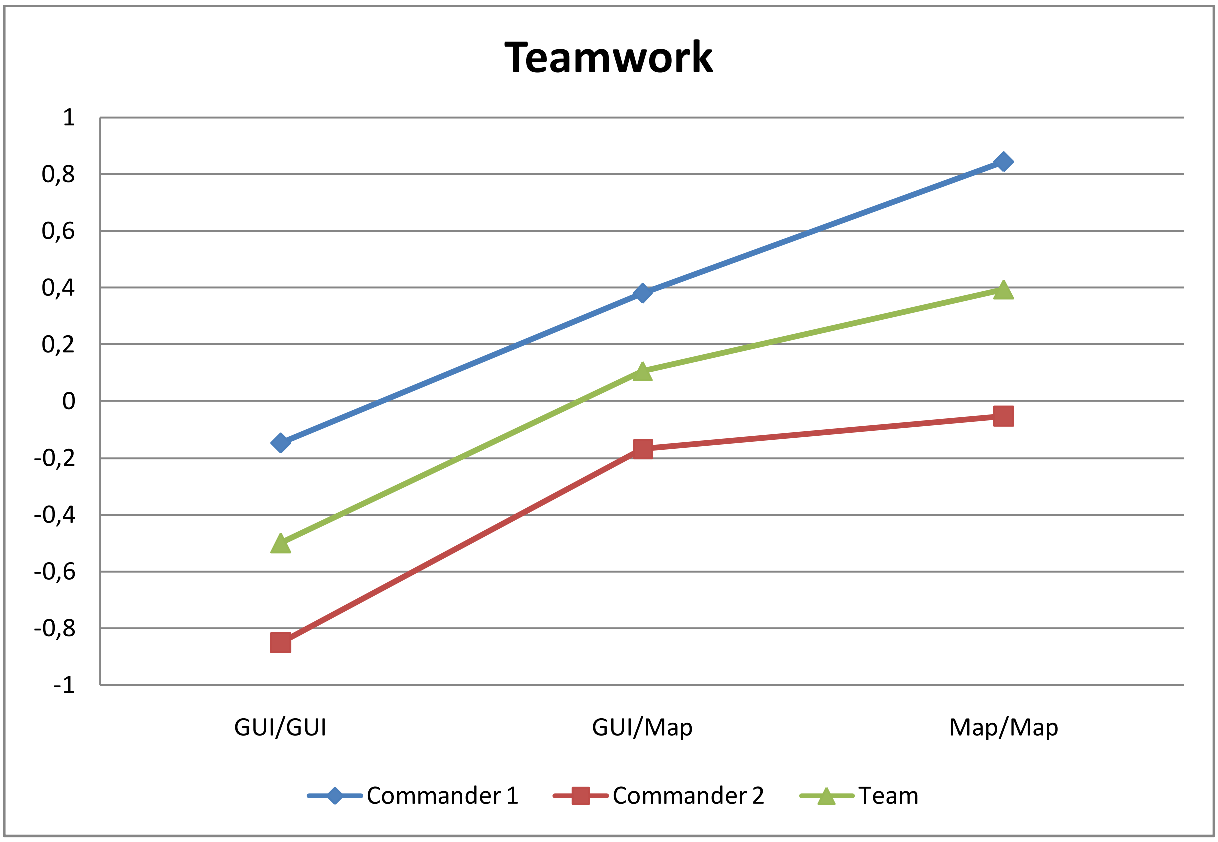


Figure 4-22 Teamwork Factor

4.3.2.4 Hypothesis 6: Performance

The ANOVA on performance reveals that H_a is true for hypothesis 6.

Table 4-11 Hypothesis 6 Answered

Hypothesis 6: Mixed conditions will affect performance in C3Fire.
H_a: There is a difference in performance between all three conditions.

There are differences in Performance between all three conditions, $F(2, 68) = 40.99$, $p<0.001$. Sidak post-hoc test shows that GUI/GUI has better (i.e. lower measurement on) Performance than GUI/Map (mean difference -0.907, $p<0.001$) and Map/Map (mean difference -1.502,

$p<0.001$) and GUI/Map has better Performance than Map/Map (mean difference -0.596, $p<0.01$).

There are no significant differences in Performance between the roles of the participants, $F(2, 34) = 0.006$, $p=0.94$.

There are significant interaction effects between different combinations of conditions and roles, $F(2, 68) = 6.97$, $p<0.01$. Tukey post-hoc test shows that Performance is better for Commander 1 in GUI/GUI than GUI/Map (mean difference -0.81, $p<0.05$) and better in GUI/Map than Map/Map (mean difference -1.18, $p<0.05$) as well as better in GUI/GUI than Map/Map (mean difference -1.98, $p<0.05$). Tukey post-hoc test also shows that Performance is better for Commander 2 in GUI/GUI than GUI/Map (mean difference -1.01, $p<0.05$) and Map/Map (mean difference -1.02, $p<0.05$) but showed no difference in Performance for Commander 2 between GUI/Map and Map/Map. Performance was also significantly better for Commander 2 than Commander 1 in Map/Map, $F(1, 102) = 8.17$, $p<0.05$.

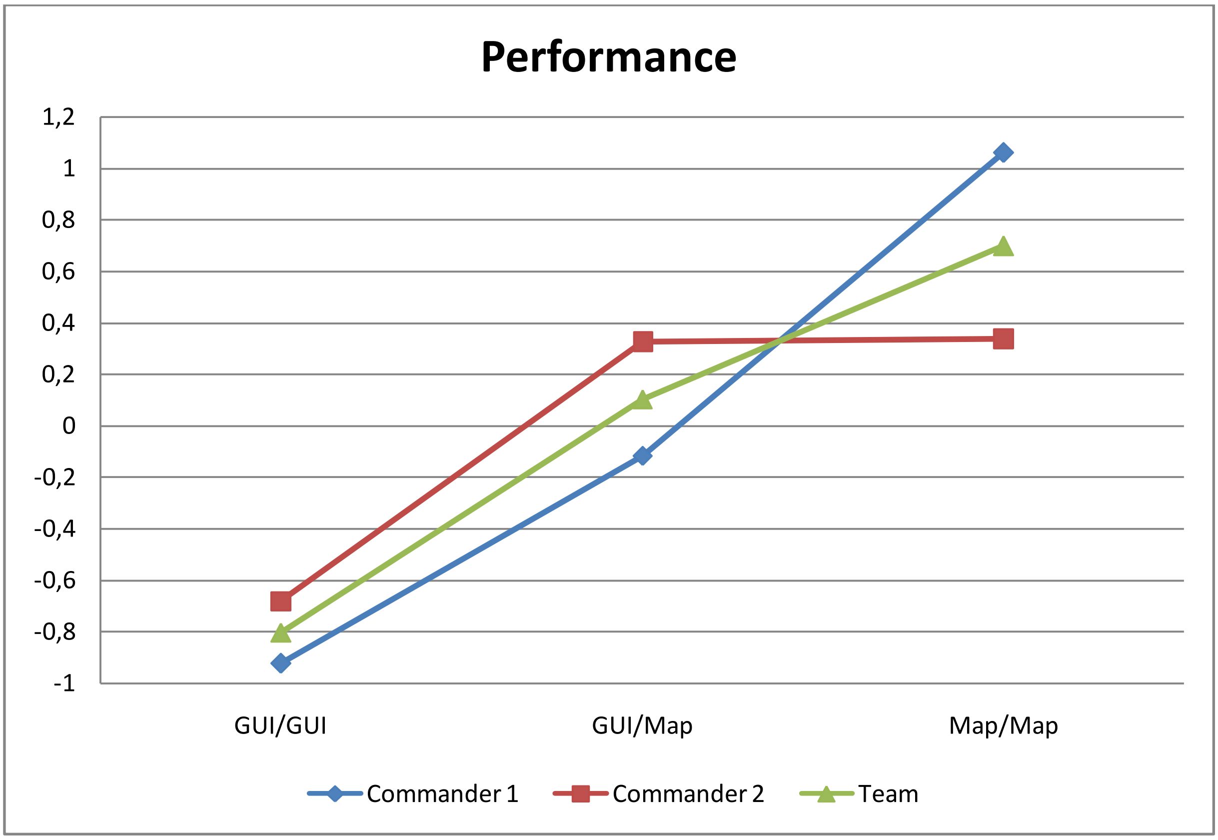


Figure 4-23 Performance Factor

4.3.3 External Factors

There are differences in External Factors between some of the three conditions, $F(2, 68) = 31.98$, $p<0.001$. Sidak post-hoc test shows that GUI/GUI has lower score on the External Factors factor than GUI/Map (mean difference -0.994, $p<0.001$) and Map/Map (mean difference -1.213, $p<0.001$). There were no significant differences between GUI/Map and Map/Map on the External Factors score.

There are no significant differences in External Factors between the roles of the participants, $F(2, 34) = 3.08$, $p=0.12$.

There are no significant interaction effects between different combinations of conditions and roles, $F(2, 68) = 3.08$, $p=0.052$.

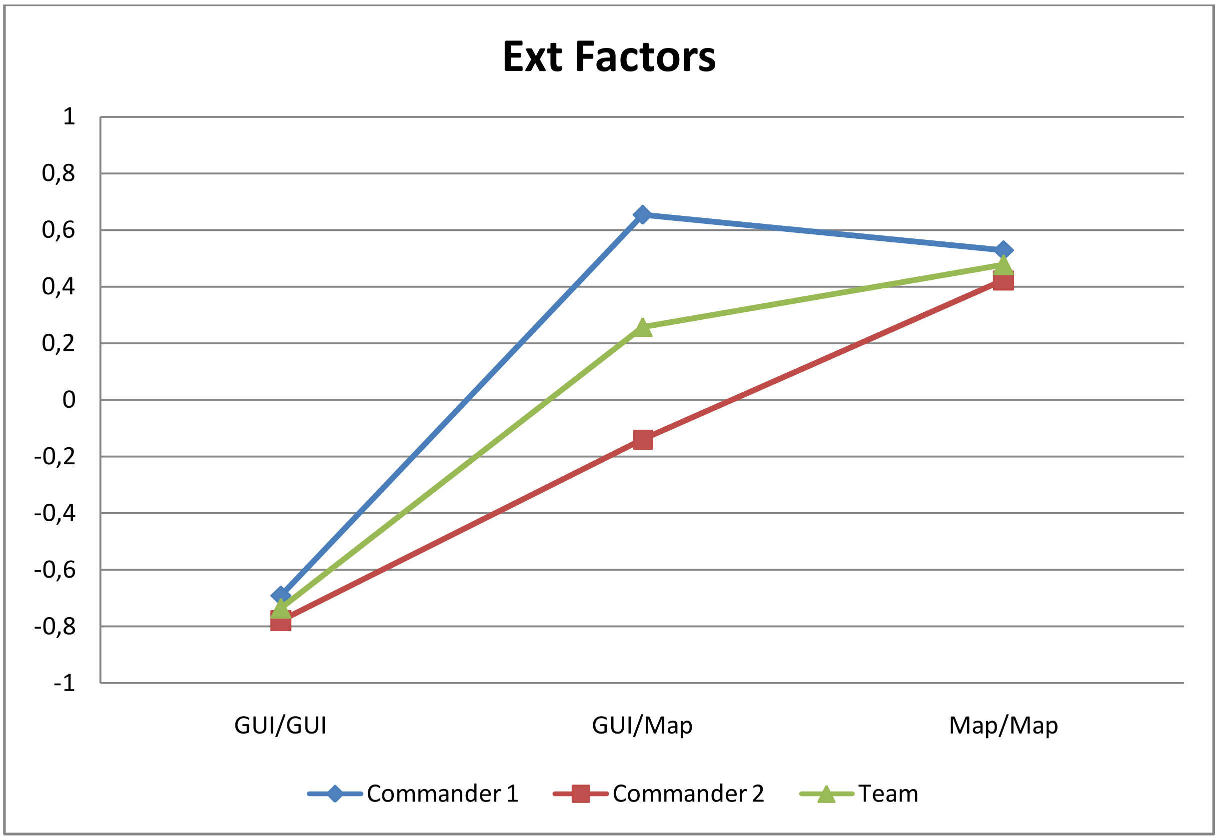


Figure 4-24 External Factors Factor

4.4 Relations between Factors

The fourth research question concerns the relationship between the different concepts used in previous analyses as well as their effects on each other. This was re-interpreted after the factor analysis to mean the relationships between factors. A Structural Equation Model was determined to be a good way of exploring this question.

4.4.1 What is SEM?

Structural Equation Modelling (SEM) is essentially a general, multivariate statistical technique for creating linear models. SEM can be used in exploratory or confirmatory ways, and some applications are casual modelling (testing casual models) or confirmatory factor analysis (an extension to a factor analysis to test specific hypotheses concerning the structure of the factors). A SEM is sometimes referred to as a LISREL model, but the two should not be confused. LISREL, which stands for Analysis of Linear Structural Relationships, was simply one of the first software programs developed to calculate SEMs and this has led to a conflation of the two terms.

In simple terms it can be said that while a PCA or Factor Analysis reduces the data into a smaller, and more manageable, number of factors, a SEM is one way of structuring these factors in a structured order. In this thesis, a SEM is constructed to describe the collected data and indicate how this particular data may be ordered. The SEM should not be seen as something that is necessarily able to be generalised to the population, particularly because the sample (N=108) is very low. However, creating a model has merits on its own in terms of data description and understanding, and was thus deemed a useful approach.

While a normal statistical approach would be to deduce a model from collected data, SEM takes the approach of testing whether a specified model could have generated the collected data. As such, it is normal to start with a hypothesis of relations between the data. If used as a confirmatory factor analysis, this would mean that the first step is to, based on available theory and hypotheses, create a model for how the factors from the factor analysis are related.

The second step would be to perform the SEM and see if this model has a good "fit". Fit is essentially how well the model is supported by the data, and there are several different types of fit indexes. However, the fit of a model should be interpreted with care. First of all, fit does not tell if the model is useful or not, it does not say whether there is a model with a better fit or not, there will (nearly) always be ambiguity with some indexes indicating acceptable values (favouring acceptance of the model) and some unacceptable (favouring rejection), and, finally, that a model cannot be evaluated void of theory and solely using fit indexes as criteria for accepting or rejecting models would be hampering scientific progress (Diamantopolous & Siguaw 2000, citing Byrne 1998, Darden 1983, Bagozzi & Yi 1988). Further, there is no consensus in the scientific field concerning what indexes to use or what values to use as cut-off scores. Therefore a range of recommended indexes will be reported for the analysis done in this thesis in table 4-13.

Another important issue that is highly relevant in this case is that of the size of the sample size, N. A SEM needs a large N in order for certain fit indexes to be useful and also to properly calculate the model. An N smaller than 200 will cause overestimations in some goodness of fit indexes, although some indexes, such as Root Mean Square Error Approximation (RMSEA) and Incremental Fit Index (IFI), are less sensitive to sample size. The N in this case was 108. Chi-Square is an often-used fit index that is very sensitive to sample size. A significant Chi-Square value would be a rejection of the null-hypothesis where the null-hypothesis is that the model "fits the population data perfectly" (Diamantopolous & Siguaw 2000). Thus, a significant Chi-Square would possible be cause for rejection of the model tested.

There are some implications of using a within-group, repeated measures design for this type of statistical analysis. The intra-subject variation, i.e. the variation that is inherent in each individual and transferred between conditions (sometimes called individual differences) can be seen as a violation to the criteria of statistical independency. However, this type of violation is common in applied studies as it may not be practically feasible to take these theoretical violations into account. As previously stated, this analysis does not aim to generalize a model to the population but to explore the available data and test a theory-based model. See Castor (2009) for a discussions on the merits of multivariate analyses as statistical tools contra the disadvantages in terms of risks of false conclusions.

McDonald & Ho (2002) argue strongly that the correlation matrix should be included for up to 30 variables when reporting a SEM. This is in order to provide the reader with a fuller understanding of the variables, and the opportunity to create other plausible explanations. The correlation matrix used for this analysis is therefore provided in appendix B. All significant correlations are marked in green (p<.01) or yellow (p<.05).

4.4.2 The SEM Analysis

A SEM was created with the components from the PCA as latent factors. The model was created in LISREL 8.3 (Jöreskog & Sörbom 1993). Previous studies show clear links between Workload, SA and Performance (Svensson et al. 2006; Nählinder et al. 2004) and this was thought to be the case here as well. The different factors are loaded with variables from questionnaires specifically designed to investigate workload, SA, and so on, and should thus exhibit the same structure as has been shown previously. The factors were also in and of themselves very clear and structured which would enable a SEM despite the small N. During the iterations of creating the model, two variables were deemed as having high modification index values and were moved to more appropriate factors. These were Common operational picture, which was moved from External Factors to Teamwork, and Team Timepressure, which was moved from External Factors to General Workload. The created model showed a strong connection between External Factors and General Workload. The three remaining variables in External Factors were thus moved to General Workload, which, based on the available theory, seemed more plausible. This modification of factors did not negatively impact the fit indexes compared to keeping the factors as they were created by the PCA.

Table 4-12 SEM Fit Indexes

Term	Value	General rule for acceptable fit	Notes
N	108		N>200 is preferred
Degrees of Freedom	235		With a large number of variables, this number is large. Calculated as: Df = ½k(k+1)-t, where k = number of observed variables and t = number of parameters
Chi-Square	334.15		
Chi-Square *p*-value	<.001	>.05	
Goodness of Fit Index (GFI)	0.79	>.95 (Schreiber et al. 2006)	This used to be a common value to use, but it is no longer recommended. With large df compared to N, GFI will be biased downwards. Calculated as: GFI = 1-(ChiSq for default model/ChiSq for null model)
Adjusted Goodness of Fit Index (AGFI)	0.74	>.95(Schreiber et al. 2006)	Similar to GFI but adjusts for df. May underestimate fit for small sample sizes. Calculated as: AGFI = 1-[(1-GFI)*(p*(p+1)/2*df)] where p = number of parameters.
Comparative Fit Index (CFI)	0.93	> .95 (Hu & Bentler 1999)	Similar to NFI but penalizes for sample size. Calculated as: CFI = (1-max(ChiSq-df,0))/(max(ChiSq-df,(ChiSqn-dfn),0)), where ChiSq and ChiSqn is the Chi-Square value for the model and null model, and df and dfn is the degrees of freedom for the model and null model.
Normed Fit Index (NFI)	0.82	>.95 (Schreiber et al. 2006)	Ranges 1 to 0 where 1 indicates perfect fit. Calculated as: NFI = (ChiSq for null model – ChiSq for default model)/ChiSq for null model
Non-Normed Fit Index (NNFI)	0.91	> .95 (Hu & Bentler 1999)	Relatively independent of sample size. Calculated as: NNFI = (ChiSqn /dfn-ChiSq/df)/ (ChiSqn/dfn-1)
Incremental Fit Index (IFI)	0.93	>.90	Relatively independent of sample size. Calculated as: IFI = (ChiSq for null model – ChiSq for default model)/(ChiSq for null model – df for default model)
Root Mean Square Error of Approximation (RMSEA)	0.063	< .06 (Hu & Bentler 1999)	RMSEA does not require comparison to null model as many others do, which makes it a popular fit index.
Standardized Root Mean Square Residual (SRMR)	0.089	< .08 (Hu & Bentler 1999)	A small value indicates good fit, cut-off values of .10, .09, .08 or .05 have been suggested. A large sample size normally gives lower values.

These different fit indexes give a slightly mixed message, but that is understandable given the low N. GFI, AGFI and NFI indicate a poor fit, while CFI, NNFI and IFI indicate a rather good fit. SRMSR and RMSEA are also a low and indicate a good fit. A CFI of .93 can be said to indicate that this model is able to describe 93% of the available data . All in all this indicates that the model should be seen as a valid description of the available data, which was the purpose of the analysis.

The created model can be seen below in figure 4-25. It was based on the Workload-SA-Performance model that has been established in earlier studies (Nählinder et al. 2004) but with the addition of Teamwork as a mediating factor between the General Workload and SA. The model will be analysed further in section 5.2.4 "Research Question 4" in the Discussion chapter.

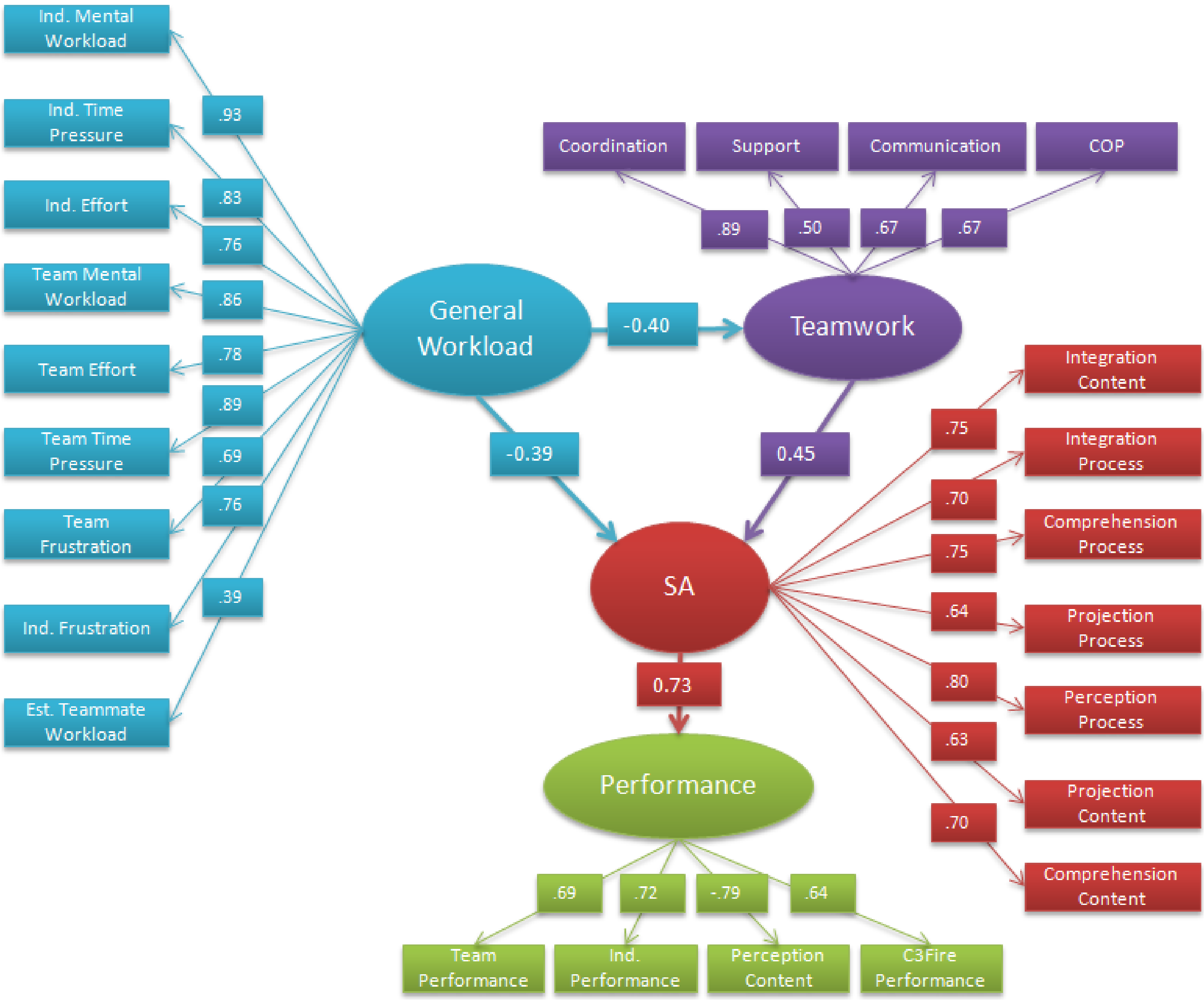


Figure 4-25 Structural Equation Model

5 Analysis/Discussion

This chapter will analyze and discuss the results presented in the previous chapter and tie these to the purpose of the thesis and the theories discussed in chapter 2. The first section, 5.1 Method Discussion, will discuss the methods used in investigating the purposes as well as look critically at the statistical methods used, while 5.2 Result Analysis will focus on interpreting the results. Last, 5.3 Conclusions and Future Works will summarize the thesis and point at potential future studies.

5.1 Method Discussion

Problems and other issues relating to the experimental setup, the statistical tests used and the results overall will be discussed in this section.

5.1.1 Task, Fire and Position Awareness

One identified problem was that the Task/Fire/Position Awareness questionnaire performed poorly. While a significant difference exists between the GUI and the Map conditions for the Position and Fire Awareness scores, the results are overall very low. The highest mean score, GUI/GUI, for Position Awareness was 7.667 out of 24, for task score the average for all conditions was between 2 and 3 out of 8 and in Fire Awareness the highest mean, again GUI/GUI, was 35.81 out of 100.

It is unclear if these low scores come from the questionnaire being misinterpreted, if the task in itself was too difficult, or a combination of these two. Some potential reasons for these low scores have been identified. First, concerning the Fire Awareness, many participants failed to mark down more than a few squares. This could be because they did not properly understand the instructions to mark down all current and previous fires (i.e. including those squares that had previously been on fire but now were extinguished or had burned out) and only marked down the squares that were on fire at the moment. It could also be the case that the participants had not yet had time to formulate an accurate mental model of the behavior of the fire in the microworld. In order to get high scores, the participants had to infer where the fire should be at the moment, i.e. how much it should have spread in directions where they had not explored. This task might have been too complex or demanding.

As for the position score, the participants had problems discerning between the units of the same kind. A common remark to the experiment leader was that they knew *a* fire or water truck to be at a specific place, but not which one out of the four or two possible units it was. While the scoring took this into account and gave partial points for marking the wrong ID on a position occupied by another unit as long as they were of the same type, it still might have contributed to the low mean.

The task score did not discriminate between conditions and was also very low overall. One potential reason is that the question was not adequately adapted to the C3Fire microworld. Participants often marked, for example, a fire truck as extinguishing fire when it was in fact inactive. This seem to indicate that the participants did not have a moment-to-moment awareness of what each specific unit was doing, but rather a mental model on a grander time scale. The fire trucks might just have finished extinguishing a fire and was inactive at the time of the questionnaire, but in the mind of the participant, the fire truck was still involved in fighting fire. Whenever the participant would have noticed the fire truck standing inactive, it would have been sent to fight fire again. One way of phrasing this is to say that the participant did not consider each square a separate fire, which is a valid point of view when fighting the fire. However, it means that until all squares are extinguished or a fire truck is clearly without the ability to fight it (for example if it is lacking in fuel or water), the fire truck are involved in fighting fire even though they at the moment are inactive. In short, the fighting fire task is seen as something spanning the entire scenario, or perhaps from the point that the fire is discovered to the point that it is extinguished, and a short moment of inactivity does not change the task of the unit. The questionnaire did not take this into account.

Another problem with this questionnaire is that the map and the GUI should perhaps not have been removed as the agent and the artefact should be seen as one unit. Removing the artefact isolated the user and the questionnaire was answered based on what the participant had internalized. However, meta-SA or meta-cognition (i.e. knowing what do to do gain an understanding or knowing where to look to see certain information) is important as well. It is plausible that the participants would have performed better if they were allowed to keep the map or GUI when filling in the questionnaire because they had a meta-SA, they could have known where to look to correctly answer a question. This would also have been more in line with the theoretical approach of viewing the participant and the artefact together as a system. Thus, those results should be treated with caution not only because their poor performance but because the method was not used properly based on the theoretical assumptions made in this thesis.

5.1.2 The PCA

A PCA is a good tool, but interpretation is required to understand the underlying factors. In this study, all factors but one (External Factors) has good theoretical underpinnings. The workload factor contained only variables measured by NASA TLX, a scale used for 20 years to measure workload. The teamwork and SA factor only had variables from questionnaires specifically designed to measure teamwork and SA. The performance factor contained variables from four sources, but three of them were in fact specifically designed to measure performance. That these performance measures were based on two different concepts; a subjective rating of either individual or team performance, and a more objective approach using the number of saved squares only serves to strengthen the factor. It is in fact one indication that the participants were well aware of how well they performed in the microworld. The last variable in the performance factor was, as mentioned in the result section, a variable from the SA-questionnaire CARS that aimed to measure the very lowest form of SA; if the information that the participant had at that time was reliable. If the structure shown in the SEM model, similar to that found in Nählinder et al. (2004), is indeed an accurate description of the world, it would make sense to connect this lowest form of SA with performance.

The one factor that is hardest to interpret would be the External Factors, containing variables from three different questionnaires; individual workload, team workload and teamwork. It was also the only factor that was changed during the creation of the SEM model. It was shown that the Team Time Pressure variable from the team workload questionnaire was a better fit in the General Workload factor, and that the Common operational picture variable from the teamwork questionnaire was a better fit in Teamwork. This left External Factors with three variables; individual and team frustration, and estimated teammate workload, that could be combined with the General Workload factor to create a more theoretically sound model.

5.1.3 The effect of roles

The roles differed in their purpose (extinguish fires vs. providing water/fuel to enable fire fighting) even though the system goal was shared. While it "forced" teamwork as the goal could not be met without cooperation, it also brought with it the risk that the requirements on the different roles would produce a different view on workload, teamwork and so on. It was shown that Commander 2, the role of providing water and fuel, consistently ranked the teamwork as lower (that is, better) than Commander 1, the fire fighter. One possible explanation to this is that Commander 1 was more active in communication needs, directions and even orders to Commander 2. Commander 1 had units in closer proximity to the fire (since a unit had to be on the square that was burning in order to extinguish it) and also knew which units needed water/fuel (via the info-window below the chat window). This meant that Commander 1 was more likely to "take command" of the situation and requests water or fuel to specific positions. Commander 2 on the other hand, could not contribute with as much information in return unless his/her units happened to spot a fire that Commander 1 was unaware of. This discrepancy might account for why Commander 2 thought the teamwork worked better; Commander 1 simply sent a lot of information without receiving much back. The main indication for Commander 1 that the teamwork was working would be that the fire trucks always had fuel and water. Commander 2 on the other hand received a lot of information and did not directly see if Commander 1 was performing well or not (since Commander 2 was not as in as close proximity to the fire). Since the teamwork questionnaire specifically asked, among other things, for the participants to rate the information dissemination within the team, Commander 1 and 2 may have had different ideas as how well this was working.

5.1.4 Paper Map Training Effects

A potential training effect could be discerned in some data showing for Commander 2, the participant who received two paper map tasks. The effect was not great enough for a significant difference between the two roles to appear on all factors, but a lack of significant difference between the GUI/Map and Map/Map conditions for Commander 2 but not for Commander 1 in the Performance, SA and Workload factors seem to indicate that Commander 2 was not as affected by Map/Map as Commander 1. Of course, this could be due to inherent differences in the two roles, and a plausible explanation is that in the GUI/Map condition, Commander 1 with a GUI worked faster than Commander 2 who had to struggle to keep up and thus experienced a greater workload and lower SA. In the Map/Map condition both participants were on equal footing and this could be a relief for Commander 2. But another plausible explanation is a training effect. None of the participants did get to train on using a paper map. This was partly due to an erroneous belief that university students should be familiar with the basic functions of a map (for instance that north was “up” on the map) which was not always the case. But most of all it was due to time constraints. Each experiment ran for three hours and adding a training sessions or second map-scenario would increase the time to the point where recruiting participants would be difficult as well as increasing the time needed for data collection beyond what was acceptable within the given time constraints of the thesis work. The lack of training on paper maps might have affected the result in the following ways:

1. Abnormal increase in workload in comparison to trained map users
2. Training effects for role Commander 2 who received two paper map tasks

If replicated, participants should be trained in map use and both roles should receive the paper map in separate conditions. This would allow more valid comparisons to be made. This thesis was not designed to investigate the effect of training on paper map tasks, and the broken down data that seem to indicate a difference between the roles and more importantly between the results of Commander 2 depending on whether Map/Map or GUI/Map was performed first have too low N (9) to draw any valid conclusions. While this could be an interesting future study, it is not directly relevant to the results of this thesis as the balancing of conditions should be enough to minimize the potential training effects of the paper map for Commander 2.

5.2 Result Analysis

In this section, the results will be discussed in a broader context relating to the purposes and research questions of this thesis.

This thesis had two purposes. One purpose was to validate a recently developed shared understanding measurement called Shared Priorities by examining how well it measures differences in conditions and how it relates to the other measures used in this study. Research questions 1 and 2 and hypotheses 1 and 2 are relevant for this purpose.

The second and main purpose of this thesis was to examine how mixed conditions affect a C^2-system, particularly in terms of shared understanding, situation awareness, performance and workload. Research questions 3 and 4 as well as hypotheses 3 through 6 are relevant for this purpose. These hypotheses were answered by comparing factors from a PCA between the different experimental conditions. It was found that the situation awareness and performance of the participants gradually deteriorated between conditions, so that the best case was GUI/GUI, then GUI/Map and last Map/Map as the worst case. For workload and teamwork, there was no difference between the mixed condition (GUI/Map) and the Map/Map condition while GUI/GUI had lower workload and better teamwork.

5.2.1 Research Question 1

How will mixed conditions affect the shared understanding in the system?

Unfortunately, no clear results were found to explain this question in full. The SP measure did not discriminate between conditions, and the second method used to measure shared understanding, a part of DATMA, could only measure some differences between conditions. The only discrepancies detected in the shared understanding between conditions was in team mental workload, where GUI/GUI had higher discrepancy than the other two conditions, in team time pressure where GUI/GUI was higher than Map/Map, in Communication, where GUI/GUI was lower than Map/Map, and finally Taskwork Congruency, where GUI/GUI and GUI/Map had a higher congruency score than Map/Map. These results are fairly weak as only two out of nine variables has significant discrepancies and the congruency scores where overall very low (the highest mean score being GUI/GUI with 1.611 out of 8). The problems of this part of DATMA are also discussed further in 5.1.1 Task, Fire and Position Awareness. Last, these results give a mixed message where the disagreement on the mental workload and communication of the team appears greatest in the GUI/GUI condition, but the agreement on what tasks were performed was best in GUI/GUI, although this was not significantly different from GUI/Map. Thus, while H_0 was technically rejected for hypothesis 2, the room for interpretation is limited.

5.2.2 Research Question 2

How well will Shared Prioties be able to measure shared understanding?

The SP measure did not discriminate between any of the conditions. Based on previous studies done using this measurement and some observations made in this study some possible explanations for this was formulated:

- The participants were not experts and thus had no conception of appropriate priorities
- The participants only experienced one role and could therefore not fully understand the situation of the other
- The scenarios might have been too short for a shared understanding to materialize
- The difference in roles where one was communicating less than the other might have led to a discrepancy in the understanding of the situation
- Individual differences might have existed prior to the experiment, e.g. in cultural or socio-economical background, etc.

Unlike other studies (for instance Höglund et al. 2009), the goal was clear and understood by the participants so this could be ruled out as an influencing factor in this study. In a study performed on professional soldiers in the Swedish armed forces, some differences were found between the conditions (Berggren & Johansson 2010). While not statistically significant, they seem to indicate that the SP-measurement might not be appropriate for teams of novices that do not know each other, as this study was set up.

Previous research showed that there are clear differences in priorities and behavior in C3Fire depending on the cultural background (for example Lindgren & Smith 2006). Thus, this might also be an influencing factor in this study. While all participants were Swedish citizens and all students at Linköping University, they still might have had very different backgrounds and this was not checked. For a continued future development of SP as a valid measurement, these possible influencing factors listed above must be checked in controlled experiments and a more solid theoretical connection to existing work should be done as well.

Another possible approach would be to take a more qualitative stance and look at the priorities themselves as a form of communication. Their content and use could be analyzed from such a theoretical stance which might provide insights in how the method could best be put to use and eventually construed as a quantitative measurement.

5.2.3 Research Question 3

> *How will mixed conditions affect the mental workload, SA, teamwork, or performance of the system?*

First of all, this result indicates that the GUI was superior to the paper map in terms of these measured aspects. This was not unexpected. SA for instance, should logically be sensitive to the switch from a dynamic map, where the information is readily available and up-to-date, to a paper map, where the information has to be entered and updated manually. The interesting result lies in the mixing of the two mediums. All aspects suffered when the paper map was introduced to one team member, but only SA and performance further deteriorated when the second paper map was introduced. Another way of putting it is that going from having two paper maps to one paper map and one computer interface will only benefit the SA and performance of the system but not necessarily the workload and teamwork.

Returning to the example given in the introduction of the battle tank Leopard 122 with its electronic system and the Combat Vehicle 90 with its lack of the same, the following implications could be the case for the commander based on the results of this experiment. A rear staff of commanding officers who are attempting to lead these two vehicle types on a joint mission from a rear HQ would have a workload comparable to that that they would have if Leopard 122 lacked an electronic system. The same goes for teamwork, there would be no benefit from the added capabilities of the electronic system. However, the commanders' understanding of the situation, SA, would improve and the overall performance of the system would as well.

It is conceivable that the effects of these mixed conditions are not clearly manifested in the daily operations of the armed forces thanks to training effects and greater cohesiveness of the units, two variables not controlled in this experiment. By training the soldiers and commanders in using a paper map, the workload may be reduced greatly. Joint exercises with the two different vehicle types would also increase the teamwork, as well as give other potential benefits. However, in novel situations these effects may resurface. It could be the case with units from different regiments, or during international joint missions where the systems may differ greatly and be completely new to the commanders. The results from this thesis indicate that the commanders would then only benefit in some ways, e.g. higher SA and performance, but not in others, e.g. workload and teamwork when forces with higher capabilities are added to forces that has previously lacked these (particularly, as in this case, added sensor capabilities). Conversely, if units with a system poor in these regards (i.e.

comparable to units using only a paper map) was to join a unit with a high-performing system (i.e. comparable to the GUI in this experiment), the rear staff would experience an increase in workload and a decrease in SA, teamwork and performance[4].

5.2.4 Research Question 4

> *How are the concepts used to investigate the effects of mixed conditions (SA workload, performance, shared understanding) related, and how do they affect each other?*

The data was further explored by looking at the relations between factors via a SEM. Based on the available data a model was created that followed previous designs where workload affects SA, which in turn affects performance. This model has strong theoretical ties. A “good” performance is essentially a measure on how successful a system is in accomplishing its goals by manipulating its environment. In order to implement the “correct” actions (i.e. one that leads to high performance), the “correct” decision (a decision that leads to the “correct” action) must be made. This decision is grounded in the system’s understanding of the current situation in relation to its goal, or SA for short, as has been argued both from a CSE (“the Contextual Control Model (COCOM) described earlier specifically states that it is the context that, together with the competence of the system, determines the actions that lead to a change in the environment”), C^2 (“The sensemaking function [of DOODA] aims to produce an understanding of the situation, in the form of “what should be done” in the current situation”) and SA (“A person with poor SA will...make poor decisions and thus have a poor performance”) perspective. In Endlsey’s three-step model (Endsley 1995), workload is a factor affecting SA, and it has previously been shown that these two concepts are indeed linked (Nählinder et al. 2004), so also in this case.

What is different about this model is the mediating factor teamwork that is affected by workload and in turn affects SA. With the experimental setup in mind this is not a major deviation, but rather a logical progression that comes with moving from a single-user perspective to a more team or system based view, which will be developed further below.

---

[4] The overall performance of the system, in terms of achieving the intended effects, would potentially increase with the added units, but performance here refers to an arbitrary performance-per-unit measure. That is, the added units would not contribute as much to the system overall as they would if they all used the same system.

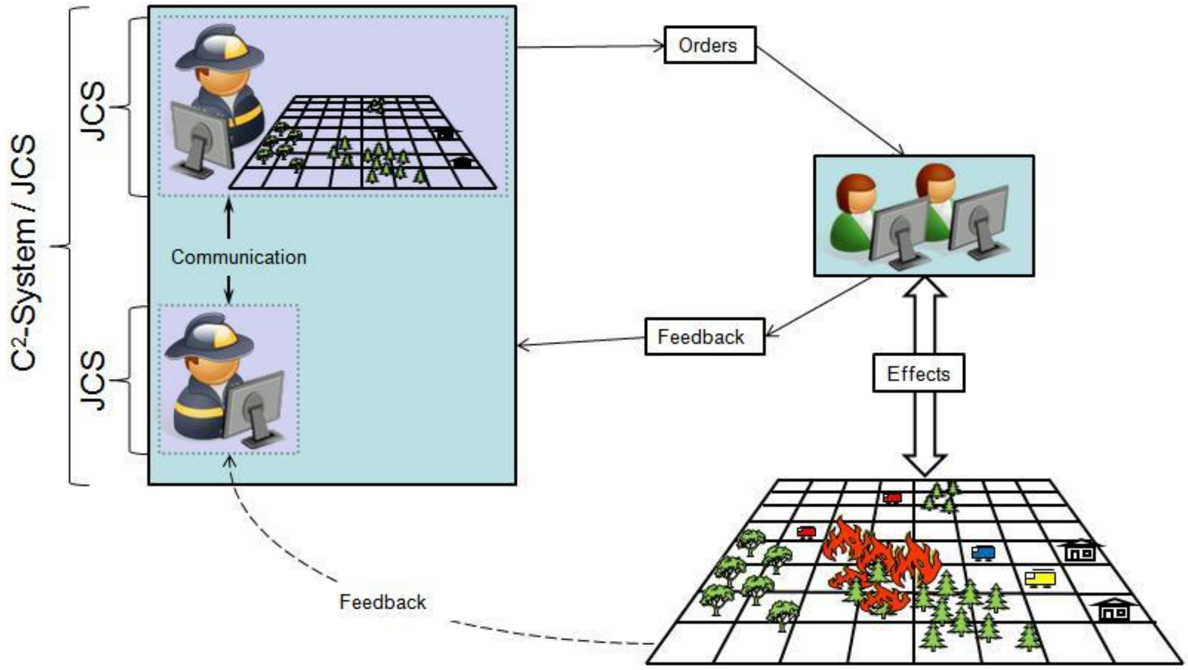


Figure 5-1 Experiment setup as DOODA with multiple JCSs

The fourth question can also be approached in a more theoretical sense by returning to the DOODA-loop as a model of C^2-systems and the CSE perspective, illustrated above in figure 5-1 as it was manifested in this experiment. The concern is the workload, SA, performance, etc. of the C^2-system, which can be described as a JCS consisting of, in this case, two JCSs with different capabilities. The situation awareness of the C^2-system exists as a system-wide state and process (as was argued in more detail in section 2.7 Summary), as does the workload (much like an airplane cockpit can be seen as having a memory, see Hutchins 1995). The teamwork could be seen as something within the C^2-system as well as between the C^2-system and the system that performs the actions ("military activity" in terms of DOODA), and performance could also be affected by frictions or outside disturbances. However, these things were controlled for thanks to the experimental setup with a controlled microworld, so that the teamwork and performance of the C^2-system can be analyzed alone.

Using Endsley's three-step model it could be said that there were "task factors", for instance workload, that negatively affected the SA of the system, leading to poorer decisions and poorer performance. Using the Extended Control Model (ECOM, see section 2.2.3 COCOM & ECOM) instead, another explanation is possible. ECOM uses parallel processes on different levels to explain system performance. The targeting level, which is concerned with situation assessment, creating an understanding of the current situation and setting goals and targets for the lower level processes, can be said to have been disturbed by the gradual degradation of the capabilities of the two JCSs. The disturbance would have manifested itself as a lack of understanding of the current situation, or SA, which indeed happened. This disturbance of the targeting process can be partially explained by the workload and teamwork (also affected by the workload) that the change in capabilities brought with it, as shown by the SEM analysis. A disturbance in the targeting process would have implications for system performance as the formulation of goals, and thus the criteria for selecting appropriate actions, would be hindered. It is not unreasonable to assume a disturbance in the monitoring level independent from the one in targeting. The monitoring level is also concerned with setting goals based on the feedback from the actions. Thus the change in feedback capabilities that came with the different conditions could directly influence this level as well as the targeting level. In comparison, the goals in Endsley's model are "individual factors" that are not directly affected by workload. Thus the CSE perspective, here exemplified by ECOM, may be more appropriate, thanks to its functional, and holistic, approach, to explain the system performance (performance in a broader sense, including e.g. workload and SA) and the effects of mixed conditions.

5.3 Conclusions & Future Works

In summary, the Shared Priorities measure did not discriminate between conditions and could not be cross-validated with any other measure used. Overall, the shared understanding measurements were problematic, and it is unclear if this stems from the nature of the task or the measurement methods themselves. As mentioned previously, more controlled studies are needed to further develop the SP method. Training effects, group cohesiveness and many other factors may influence the measurement in ways not yet understood or explored. The SP measure also needs to be explored in a more theoretical way to understand how it may relate to other concepts in the domains in which it is intended to be used. This would aid future attempts to cross-validate the method.

The overall performance of the system, using GUI/GUI as baseline, was negatively affected by the GUI/Map (mixed) condition, but not to the degree of the Map/Map condition. The SA was similarly affected, while the workload and teamwork did not differ significantly between GUI/Map and Map/Map. It was shown based on arguments based on available theory and supported by the Structural Equation Model that the performance of a system was affected by the SA of said system, which in turn was influenced by the workload and teamwork.

By showing this, this study validates further research about how mixed conditions affect C^2-systems. The results may be generalized to existing military systems in use today, which would indicate that problems, or frictions, could arise when units using different systems perform joint operations. To validate this, the research needs to step outside of the controlled microworld environment and look at these factors, which influence each other, in the wild.

6 Bibliography

Aminoff, H., Johansson, B., & Trnka, J. (2007). Understanding Coordination in Emergency Responce. *Proc. of the 26th European Annual Conference on Human Decision-Making and Manual Control (EAM '07).* Lyngby, DK.

Andriole, S. J., & Halpin, S. M. (1986). Information Technology for Command and Control. *IEEE Transactions on Systems, Man, and Cybernetics , SMC-16* (6), 762-765.

Artman, H., & Garbis, C. (1998). Situation Awareness as Distributed Cognition. *Proceedings of Ninth Conference of Cognitive Ergonomics (ECCE-9),* Limerick, Ireland.

Artman, H., & Granlund, R. (1998). Team Situation Awareness using Graphical or Textual Databases in Dynamic Decision Making. *Ninth European Conference on Cognitive Ergonomics (ECCE-9),* Limerick, Ireland.

Bedney, G., & Meister, D. (1999). Theory of Activity and Situation Awareness. *International Journal of Cognitive Ergonomics , 3* (1), 63-72.

Berggren, P., & Johansson, B. (2010). Developing an instrument for measuring shared understanding. *Seventh International ISCRAM Conference.* Seattle, WA, May 2010.

Billings, C. (1996). Situation Awareness Measurement and Analysis: A Commentary. In D. Garland, & M. R. Endsley, *Experimental Analysis and Measurement of Situation Awareness* (p. 3). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Boyd, J. (1987). *A Discourse on Winning and Losing.* Maxwell Air Force Base, AL (Briefing slides): Air University Library Document No. M-U 43947.

Brehmer, B. (1987). System Design and the Psychology of Complex Systems. In J. Rasmussen, & P. Zunde, *Empirical Foundations of Information and Software Science.* Plenum Publishing Corporations.

Brehmer, B. (2008). Hur man åstadkommer en snabbare OODA-loop: överste Boyds syn på ledning. *Kungl Krigsvetenskapsakademiens Handlingar och Tidskrift* (4), 42-68.

Brehmer, B. (2005a). Micro-worlds and the circular relation between people and their environment. *Theoretical Issues in Ergonomics Science , 6* (1), 73-93.

Brehmer, B. (2005b). The Dynamic OODA Loop: Amalgamating Boyd's OODA Loop and the Cybernetic Approach to Command and Control. *10th International Command and Control Research and Technology Symposium.* McLean, VA, June 17-21.

Brehmer, B. (2006). One Loop To Rule Them All. *11th International Command and Control Research and Technology Symposium.* Cambridge, UK, September 2006.

Brehmer, B. (2007). Understanding the Functions of C^2 Is the Key to Progress. *The International C^2 Journal , 1* (1), 211-232.

Brehmer, B., & Dörner, D. (1993). Experiments with Computer-simulated Microworlds: Escaping both the Narrow Straits of the Laboratory and the Deep Blue Sea of the Field Study. *Computers in Human Behavior* (9), 171-184.

Brehmer, B., & Svenmark, P. (1994). Distributed decision making dynamic environment: time scales and architectures of decision making. In J. Calverni, M. Bar-Hillel, F. Baron, & M. Jungermann, *Contributions to Decision Making* (pp. 155-174). Amsterdam: Elsevier.

Breton, R., Tremblay, S., & Banbury, S. (2007). *Measurement of individual and team situation awareness: A critical evaluation of the available metrics and tools and their applicability to command and control environments*. Defence R&D Canada.

Castor, M. (2009). *The use of structural equation modeling to describe the effect of operator functional state on air-to-air engagement outcomes*. Linköping: Linköping University Institute of Technology (Dissertation No. 1251), Linköping: Unitryck

Command and Control Research Program (2002). *NATO COBP: Code of Best Practice for C^2 Assessment*, CCRP.

Dekker, S., & Hollnagel, E. (2004). Human Factors and Folk Models. *Cognition, Technology & Work* 6 (2), 79-86.

Dekker, S., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human-automation coordination. *Cognition, Technology, and Work, 4* (3), 240-244.

Diamantopoulos, A., & Siguaw, J. (2000). *Introducing LISREL*. London: Sage.

Dominguez, C. (1994). Can SA be defined? In M. Vidulich, C. Dominguez, E. Vogel, & G. McMillan, *Situation Awareness: Papers and Annotated Bibliography*. Wright-Patterson Airforce Base, OH: Air Force Command: Report AL/CF-TR-1994-0085.

Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). *Paper presented at the National Aerospace and Electronic Conference (NAECON).* Dayton, OH, May 1988.

Endsley, M. R. (1993). A Survey of Situation Awareness Requirements in Air-to-Air Combat Fighters. *The International Journal of Aviation Psychology*, *3* (2), 157-168.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors , 37* (1), 32-64.

Endsley, M. R. (2000). Direct Measurement of Situation Awareness: Validity and Use of SAGAT. In M. R. Endsley, & D. J. Garland, *Situation Awareness and Measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (2004). Situation Awareness: Progress and Directions. In S. Banbury, & S. Tremblay, *A Cognitive Approach to Situation Awareness: Theory, Measurement and Application* (pp. 317-341). Aldershot: Ashgate Publishing.

Endsley, M., & Jones, W. (1997). *Situation Awareness Information Dominance & Information Warfare*. Belmont, MA: Armstrong Laboratories (Technical report).

Farley, T. C., Hansman, R. J., Endsley, M. R., Amonlidvirman, K., & Vigeant-Langlois, L. (1998). The Effect of Shared Information on Pilot/Controller Situation Awareness and Re-route Negotiation. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*. Santa Monica, CA, October 1998.

Field, A. (2009). *Discovering Statistics Using SPSS - Third Edition*. Dubai: SAGE Publications LTD.

Flach, J. M. (1995). Situation Awareness: Proceed with Caution. *Human Factors , 37* (1), 149-157.

Granlund, R. (1997). *C3Fire - a microworld supporting emergency management*. Linköping, Sweden: Linköping University (Licentiate thesis no. 598), Linköping: Unitryck.

Granlund, R. (2003). Monitoring experiences from command and control research with the C3Fire microworld. *Cognition, Technology and Work , 3* (5), 183-190.

Hansberger, J. T., Schreiber, C., & Spain, R. D. (2008). C^2 Network Analysis: Insights into Coordination & Understanding. *13th International Command and Control Research and Technology Symposium*. Bellevue, WA, June 2008.

Hansman, J. (2004). *Workload and Situation Awareness*. Cambridge, MA: MIT Press.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later. *Human Factors and Ergonomics Society 50th Annual Meeting Proceedings*. San Fransisco, CA, October 2006.

Hart, S. G., & Staveland, L. E. (1988). Development of a Multi-dimensional Workload Rating Scale: Results of Empirical and Theoretical Research. In P. A. Hancock, & N. Meshkati, *Human Mental Workload*, 139-183, Amsterdam: Elsevier.

Hartman, B. O., & Secrist, G. E. (1991). Situaional Awareness is more than exceptional vision. *Space, and Environmental Medicine* (62), 1084-1089.

Hendy, K. C. (1995). Situation Awareness and Workload: Birds of a Feather. *AGARD AMP Symposium on Situation Awareness: Limitations and Enhancements in the Aviation Environment*. Brussels, April 1995.

Hollnagel, E., & Woods, D. D. (1983). Cognitive Systems Engineering. New Wine in New Bottles. *International Journal of Man-Machine Studies , 18*, 583-600.

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

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal , 6* (1), 1-55.

Hutchins, E. (1995). *Cognition in The Wild*. London, England: MIT Press.

Höglund, F., Berggren, P., & Nählinder, S. (2009). *Using Shared Priorities to Measure Shared Situation Awareness*. Linköping: Swedish Defence Research Agency (Nr. FOI-R-2791-SE, ISSN 1650-1942).

Jeannot, E., Kelly, C., & Thompson, D. (2003). *The Development of Situation Awareness Measures in ATM Systems*. (No. HRS/HSP-005-REP-01), European Air Traffic Management Programme.

Johansson, B., Persson, M., & Granlund, R. (2003). C3Fire in command and control research. *Cognition, Technology and Work , 3* (5), 191-196.

Jöreskog, K., & Sörbom, D. (1993). *LISREL: Structural equation modeling with the SIMPLIS command language*. Hillsdale: Lawrence Erlbaum Associates, Inc.

Kendall, M., Smith, B., & Babington. (1939). The problem of m ranking. *Ann. Math. Statist. , 3* (10), 275-287.

Klein, G. A. (1989). Recognition-primed Decisions. In W. B. Rouse, *Advances in man-machine systems research* (pp. 47-92). Greenwich, CT: JAI.

Klein, G. A. (1998). *Sources of Power - How People Make Decisions*. Cambridge, MT: MIT Press.

Klein, G. A., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems , 3* (18), 81-85.

Lawson, J. S. (1981). Command and Control as a Process. *IEEE Control Systems Magazine , 1* (1), 5-16.

Lindgren, I., & Smith, K. (2006). Using Microworlds to Understand Cultural Influences on Distributed collaborative Decision Making in C^2 Settings. *Proceedings from the 11th International Command and Control Research and Technology Symposium.* Cambridge, UK, 26-28 Sep.

Lövborg, L., & Brehmer, B. (1991). *NEWFIRE - a flexible system for running simulated fire-fighting experiments.* Roskilde: Risö National Laboratory (Report nr. Risö-M-2953).

MacMillan, J., Paley, M. J., Entin, E. B., & Entin, E. E. (2005). Questionnaires for Distributed Assessment of Team Mutual Awareness. In N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendricks, *Handbook of Human Factors Methods.* London: Taylor and Francis.

Matthews, M. D., & Beal, S. A. (2002). *Assessing Situation Awareness in Field Training Exercises.* (No. 1795), US Army Research Institute for the Behavioural and Social Siences.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and Practice in Reporting Structural Equation Analysis. *Psychological Methods , 7* (1), 64-82.

McGuinness, B. (1999). Situation Awareness and the Crew Awareness Rating Scale (CARS). *Proceedings of the 1999 Avionics Conference.* Heathrow, London, 17th-18th November.

McGuinness, B., & Ebbage, L. (2002). *Assessing Human Factors in Command and Control: Workload and Situational Awareness Metrics.* BAE Systems Advanced Technology Centre, Bristol, UK.

McGuinness, B., & Foy, L. (2000). A subjective measure of SA: the Crew Awareness Rating Scale (CARS). Presented at the *Human Performance, Situational Awareness and Automation Conference*

Neisser, U. (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology.* San Fransisco: Freeman.

Norman, G. R., & Streiner, D. L. (2003). *PDQ Statistics (Third Ed.).* London: BC Decker Inc.

Nählinder, S., Berggren, P., & Svensson, E. (2004). Reoccuring LISREL patterns describing mental workload, situation awareness and performance. *Proceedings of the Human Factors and Ergonomics society 48th Annual Meeting.* New Orleans, LA.

Parasuraman, R., Shreidan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making , 2* (2), 140-160.

Pigeau, R., & McCann, C. (2002). Re-Conceptualizing Command and Control. *Canadian Military Journal , 3* (1), 53-62.

Riley, J. M., Endsley, M. R., Bolstad, C. A., & Cuevas, H. M. (2006). Collaborative planning and situation awareness in Army command and control. *Ergonomics , 49* (12-13), 1139-1153.

Salas, E., Prince, C., Baker, P., & Shrestha, L. (1995). Situation awareness in team performance. *Human Factors 37* (1), 123-136.

Salmon, P. M., Stanton, N. A., Walker, G. H., & Jenkins, D. P. (2009). *Distributed Situation Awareness.* Ashgate Publishing Limited.

Sarter, N. B., & Woods, D. D. (1991). Situation Awareness: A Critical but Ill-defined Phenomenon. *International Journal of Aviation Psychology, 1* (1), 45-57.

Schreiber, J. b., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research, 99* (6), 323-338.

Smith, K., & Hancock, P. (1995a). Situation Awareness is Adaptive, Externally Directed Consciousness. *Human Factors* 37 (1), 137-148.

Smith, K., & Hancock, P. (1995b). The risk space representation of commercial airspace. *Proceedings of the 8th International Symposium on Aviation Psychology*. Columbus, OH, April 1995.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). *Human Factors Methods - A Practical Guide for Engineering and Design.* Ashgate Publishing Limited.

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M. S., Linsell, M., Dymott, R., & Green, D. (2006). Distributed Situation Awareness in Dynamic Systems: Theoretical Development and Application of an Ergonomics Methodology. *Ergonomics, 49* (12-13), 1288-1311.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, Shared Mental Models, and Coordinated Performance: An Empirical Link is Established. *Human Factors , 41* (1), 61-71.

Svensson, E., Rencrantz, C., Lindoff, J., Berggren, P., & Norlander, A. (2006). Dynamic Measures for Performance Assessment in Complex Environments. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Fransisco, CA, October 2006.

Wickens, C. D. (2008). Situation Awareness: Review of Mica Endsley's 1995 Articles on Situation Awareness Theory and Measurement. *Human Factors , 50* (3), 397-403.

Wiener, E. L. (1993). Life in the Second Decade of the Glass Cockpit. *Proceedings of the Seventh International Symposium on Aviation Psychlogy*, Columbus, OH, April 1993.

Woltjer, R., & Hollnagel, E. (2007). The Alaskan Airlines Flight 261 accident: a systemic analysis of functional resonance. *Proceedings of the 14th International Symposium on Aviation Psychology*. Dayton, OH, April 2007.

Woltjer, R., Prytz, E., & Smith, K. (2009). Functional Modeling of Agile Command and Control. *Proceedings of the 14th International Command and Control Research and Technology Symposium.* Washington D.C., USA, June 2009.

Appendix A – Questionnaires

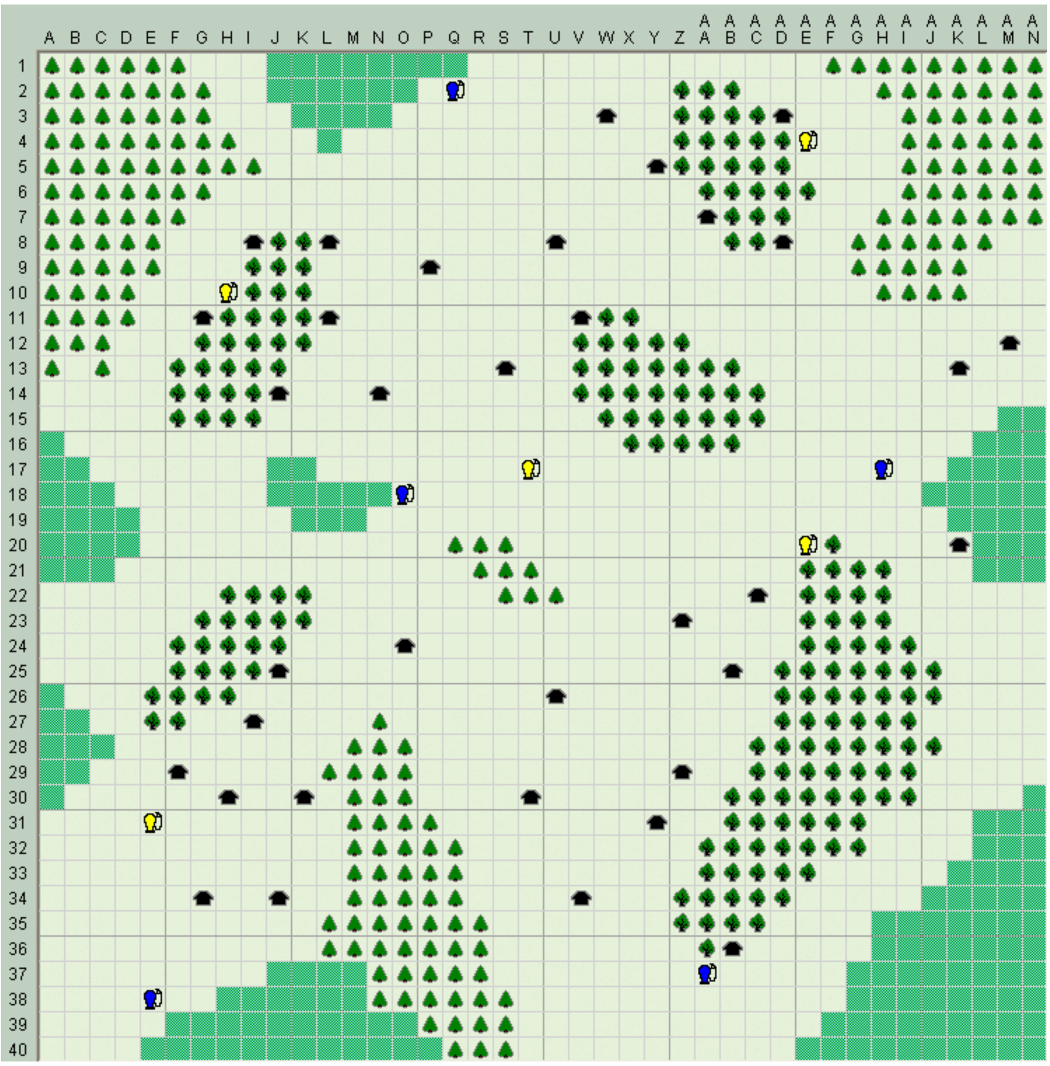
Task Mutual Awareness

Kategori	Uppgift
1	Bekämpar eld
2	Förflyttar sig för att bekämpa eld
3	Förflyttar sig med annat syfte
4	Inaktiv på grund av brist på bränsle eller vatten
5	Inaktiv av annan anledning
6	Fyller på eget bränsle/vatten från tankbil eller pump
7	Fyller på annat fordons bränsle/vatten
8	Vet ej
9	Övrigt – vänligen specificera

Enhet	Kategori (Uppgift)
F1	
F2	
F3	
F4	
W5	
W6	
G7	
G8	

Varje kategori beskriver en viktig uppgift eller grupp av uppgifter som kan utföras av enheterna i C3Fire. Använd dessa kategorier i formuläret nedan för att identifiera ***vad enheterna gör just nu*** i scenariot. Använd kartan nedan för att fylla i ***var enheterna står just nu***. (Markera F1 som 1, F2 som 2, G7 som 7 etc)

Fyll även i ***var det brinner just nu***. Använd **X** för att markera eld i de rutor som du tror brinner/har brunnit.



Team Workload Awareness – Part 1

Markera med ett X på skalan var du anser att ***din personliga*** arbetsbelastning låg under scenariot.

Mental Arbetsbelastning – Hur mentalt ansträngande var scenariot ***för just dig personligen***?

Väldigt lågt Väldigt högt

Tidspress – Hur mycket tidspress kände ***just du personligen*** under scenariot?

Väldigt låg Väldigt hög

Prestation – Hur väl lyckades ***just du personligen*** prestera under detta scenario?

Väldigt lågt Väldigt högt

Ansträngning – Hur hårt behövde ***just du personligen*** jobba för att upprätthålla denna nivå av prestation under detta scenario?

Väldigt lite Väldigt hårt

Frustration – Hur osäker, irriterad, stressad och frustrerad var ***just du personligen*** under detta scenario?

Väldigt lite Väldigt mycket

Markera med ett X på skalan var du personligen uppskattar ***din arbetskamrats*** totala arbetsbelastning.

Total Arbetsbelastning – Vad är din uppskattning av den totala arbetsbelastningen ***för din arbetskamrat*?**

Väldigt låg Väldigt hög

Team Workload Awareness – Part 2

Markera med ett X på skalan var du anser att arbetsbelastningen för ***laget som helhet*** låg under scenariot.

Mental Arbetsbelastning – Hur mentalt ansträngande var scenariot ***för er som lag***?

Väldigt lågt | Väldigt högt

Tidspress – Hur mycket tidspress kände ***ni som lag*** under scenariot?

Väldigt låg | Väldigt hög

Prestation – Hur väl lyckades ***ni som lag*** prestera under detta scenario?

Väldigt lågt | Väldigt högt

Ansträngning – Hur hårt behövde ***ni som lag*** jobba för att upprätthålla denna nivå av prestation under detta scenario?

Väldigt lite | Väldigt hårt

Frustration – Hur osäkra, irriterade, stressade och frustrerade var ***ni som lag*** under detta scenario?

Väldigt lite | Väldigt mycket

Teamwork Assessment

Ringa in det alternativ 1 – 7 som stämmer bäst.

Kommunikation
I vilken utsträckning gav ni varandra relevant information på ett förebyggande vis, utan att den behövde efterfrågas?

1 2 3 4 5 6 7

Aldrig **Alltid**

7: Viktig information vidarebefordrades alltid utan efterfrågan

1: Viktig information vidarebefordrades aldrig utan efterfrågan

Stöd
Justerade ni ansvarsfördelningen i laget för att förhindra att endera lagmedlem blev överbelastad?

1 2 3 4 5 6 7

Ingen justering **Alltid snabb justering**

7: Lagkamraten var alltid medveten om min arbetsbelastning och reagerade snabbt för att justera arbetsbördan för att jämna ut arbetsbelastningen

1: Lagkamraten var generellt *inte* medveten om min arbetsbelastning; liten eller ingen justering av arbetsbördan gjordes för att jämna ut arbetsbelastningen innan det påverkade vår uppgift negativt

Koordination
I vilken utsträckning var ni koordinerade som lag?

1 2 3 4 5 6 7

Dålig koordination **Bra koordination**

7: Bra koordination är när lagmedlemmarna skickar kritisk information till berörda parter, lagmedlemmarna är medvetna om de andras arbetsuppgifter och laget arbetar synkroniserat mot ett gemensamt mål

1: Dålig koordination är när lagmedlemmar är ineffektiva i sina egna arbetsuppgifter vilket leder till att andra medlemmar misslyckas med sina uppgifter; när medlemmar är oförutsägbara eller underlåter sig att vidarebefordra kritisk information vilket leder till sämre prestation som lag.

Lägesförståelse
I vilken utsträckning var din och din lagkamrats lägesförståelse lika?

1 2 3 4 5 6 7

Aldrig överens **Alltid överens**

7: Alla i laget var helt överens gällande mål, uppgifter och begrepp som ingick i scenariot

1: Laget var sällan överens gällande mål, uppgifter och begrepp som ingick i scenariot

Crew Awareness Rating Scale

Ringa in det alternativ 1 - 4 som stämmer bäst.

Perception - Insamlande av information
Innehåll: hur pålitlig och korrekt är informationen ni har just nu? (Ex. eldens/enheters position)

(Bästa möjliga fall) 1 2 3 4 (Sämsta möjliga fall)

Processen: Hur lätt är informationsnivån att underhålla? (Ex. hur lätt är det att lokalisera enheter på kartan?)

(Bästa möjliga fall) 1 2 3 4 (Sämsta möjliga fall)

Förståelse - Förståelse av information i kontexten
Innehåll: Hur pålitlig och korrekt är er förståelse av läget? (Ex. förståelse för hur enheternas positioner påverkar er förmåga att bekämpa elden)

(Bästa möjliga fall) 1 2 3 4 (Sämsta möjliga fall)

Processen: Hur lätt är det att underhålla denna nivå av förståelse? (Ex. hur lätt är det att förstå den information som ni samlar in?)

(Bästa möjliga fall) 1 2 3 4 (Sämsta möjliga fall)

Förutseende - Förutseende av framtida händelseutvecklingar
Innehåll: Hur pålitliga och korrekta är förutsägelserna? (Ex. till vilken grad kan du förutse hur elden kommer sprida sig?)

(Bästa möjliga fall) 1 2 3 4 (Sämsta möjliga fall)

Processen: Hur lätt är det att förutse framtida händelseutvecklingar? (Ex. hur lätt är det att utifrån er förståelse förutse vad som kommer att hända?)

(Bästa möjliga fall) 1 2 3 4 (Sämsta möjliga fall)

Integration - Sammanfogande av perception, förståelse och förutseende med ens egna handlingar
Innehåll: Hur pålitligt och korrekt är sammanfogandet av ovanstående tre aspekter och dina egna handlingar? (Ex. till vilken grad baserar du dina beslut på informationen, förståelsen och förutsägelserna?)

(Bästa möjliga fall) 1 2 3 4 (Sämsta möjliga fall)

Processen: Hur lätt är det att sammanfoga ovanstående tre aspekter med dina egna handlingar? (Ex. hur lätt är det att utifrån de tre aspekterna att se vad som måste göras?)

(Bästa möjliga fall) 1 2 3 4 (Sämsta möjliga fall

Bakgrundsenkät

Sysselsättning:____________________ **Ålder**: _______ **Kön**:_________

(Om student, även utbildning/program)

Erfarenhet av C3Fire; antal gånger du använt/spelat C3Fire

0 1 2 3 4 5+

Har du tidigare generell datorvana?

1 2 3 4 5
Ingen vana Mycket stor vana

Har du tidigare datorspelsvana?

1 2 3 4 5
Ingen vana Mycket stor vana

Känner du den andra försöksdeltagaren? Om ja, sedan hur länge?

Ja, sedan _______ månader Nej

Har du gjort militärtjänst? (ex. lumpen, hemvärnet, FBU)

Ja, befattning:_________________, i ______ månader Nej

Har du tidigare erfarenhet av ledningssystem? (ex. stridsledning eller andra större insatser)

1 2 3 4 5
Ingen erfarenhet Mycket stor erfarenhet

Har du tidigare erfarenhet av brandsläckningsinsatser? (ex. hel/deltidsbrandman)

1 2 3 4 5
Ingen erfarenhet Mycket stor erfarenhet

Har du erfarenhet av att läsa papperskartor? (ex. orienterare, utbildning, fritidsaktivitet)

1 2 3 4 5
Ingen erfarenhet Mycket stor erfarenhet

Är du van att använda elektroniska kartor? (ex. GPS, Google maps, karta i mobilen)

1 2 3 4 5
Ingen vana Mycket stor vana

Appendix B – Correlation Matrix

The key provided (table below) is for interpreting the correlation matrix on the next page. Significant results in the correlation matrix are marked in green (*p*<.01) or yellow (*p*<.05).

Label in the correlation matrix	Variable
A	C3Fire Performance
B	Ind. Mental Workload
C	Ind. Time Pressure
D	Ind. Effort
E	Ind. Frustration
F	Team Mental Workload
G	Team Time Pressure
H	Team Effort
I	Team Frustration
J	Perception Content
K	Perception Process
L	Comprehension Content
M	Comprehension Process
N	Projection Content
O	Projection Process
P	Integration Content
Q	Integration Process
R	Communication
S	Support
T	Coordination
U	Common operational picture
V	Ind. Performance
X	Team Performance
Y	Est. Teammate Workload

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	X	Y
A	1,00																							
B	0,36	1,00																						
C	0,28	0,80	1,00																					
D	0,13	0,68	0,66	1,00																				
E	0,40	0,56	0,58	0,47	1,00																			
F	0,37	0,79	0,66	0,61	0,53	1,00																		
G	0,41	0,62	0,66	0,48	0,60	0,79	1,00																	
H	0,21	0,47	0,55	0,63	0,43	0,68	0,68	1,00																
I	0,44	0,46	0,46	0,38	0,84	0,56	0,64	0,47	1,00															
J	0,53	0,52	0,42	0,36	0,49	0,45	0,36	0,34	0,40	1,00														
K	0,52	0,51	0,43	0,29	0,42	0,50	0,43	0,31	0,36	0,58	1,00													
L	0,44	0,44	0,24	0,35	0,41	0,38	0,27	0,25	0,41	0,52	0,59	1,00												
M	0,42	0,39	0,36	0,33	0,50	0,36	0,39	0,28	0,39	0,49	0,68	0,55	1,00											
N	0,38	0,53	0,48	0,37	0,51	0,38	0,46	0,29	0,38	0,41	0,45	0,46	0,42	1,00										
O	0,35	0,38	0,31	0,24	0,45	0,29	0,33	0,21	0,36	0,37	0,43	0,47	0,44	0,68	1,00									
P	0,41	0,39	0,40	0,29	0,46	0,32	0,36	0,31	0,39	0,51	0,60	0,53	0,59	0,49	0,54	1,00								
Q	0,32	0,38	0,34	0,32	0,50	0,35	0,33	0,29	0,36	0,42	0,54	0,47	0,56	0,51	0,55	0,61	1,00							
R	0,30	0,19	0,06	0,02	0,24	0,12	0,08	-0,05	0,25	0,34	0,30	0,37	0,34	0,22	0,30	0,23	0,19	1,00						
S	0,20	0,25	0,18	0,17	0,14	0,19	0,10	0,18	0,07	0,32	0,35	0,25	0,40	0,10	0,20	0,17	0,28	0,43	1,00					
T	0,38	0,38	0,21	0,20	0,36	0,38	0,40	0,26	0,39	0,36	0,42	0,44	0,41	0,36	0,38	0,37	0,32	0,61	0,43	1,00				
U	0,46	0,18	0,15	0,05	0,34	0,25	0,32	0,22	0,37	0,38	0,38	0,37	0,41	0,24	0,37	0,32	0,29	0,39	0,27	0,63	1,00			
V	0,41	0,55	0,45	0,31	0,56	0,38	0,41	0,25	0,44	0,59	0,49	0,45	0,35	0,46	0,36	0,37	0,47	0,34	0,23	0,45	0,35	1,00		
X	0,57	0,50	0,34	0,21	0,44	0,41	0,39	0,12	0,35	0,56	0,49	0,49	0,40	0,46	0,39	0,41	0,43	0,37	0,20	0,44	0,36	0,74	1,00	
Y	0,20	0,29	0,21	0,27	0,27	0,55	0,45	0,40	0,38	0,04	0,14	0,13	0,22	0,13	0,11	-0,01	0,02	0,07	0,09	0,27	0,17	0,13	0,14	1,00