ePartners

for dynamic task allocation and coordination

Tjerk de Greef

*e*PARTNERS FOR DYNAMIC TASK ALLOCATION & COORDINATION

*e*PARTNERS FOR DYNAMIC TASK ALLOCATION & COORDINATION

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op 3 juli 2012 om 12.30 uur
door Tjerk Everardus DE GREEF
doctorandus in de Informatica en Ingenieur in de Elektrotechniek
geboren te Rheden.

Dit proefschrift is goedgekeurd door de promotoren:
Prof. dr. M.A. Neerincx
Copromotor: Dr. ir. W.P Brinkman

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. M.A. Neerincx,	Technische Universiteit Delft, promotor
Dr. ir. W.P. Brinkman,	Technische Universiteit Delft, copromotor
Prof. dr. J.M.C. Schraagen,	Technische Universiteit Twente
Prof. dr. C.M. Jonker,	Technische Universiteit Delft
Prof. dr. ir. Wieringa,	Technische Universiteit Delft
Prof. dr. J-J.C. Meyer,	Universiteit Utrecht
Dr. J.M. Bradshaw,	Florida Institute of Human Machine Cognition

ISBN 978-90-8891-431-7

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

1.1 ePartners serving Humans

"*For a long time it puzzled me how something so expensive, so leading edge, could be so useless, and then it occurred to me that a computer is a stupid machine with the ability to do incredibly smart things, while computer programmers are smart people with the ability to do incredibly stupid things. They are, in short, a perfect match.*"

Bill Bryson (1999, p. 352)

In this quote, Bill Bryson ironically states that computers evidently can do clever things but that their potential is not fully exploited. Fitts analyzed as early as 1951 that humans and computers have different capabilities by composing a list of general task abilities summarizing where "*Men-Are-Better-At*" and where "*Machines-Are-Better-At*". The so-called Fitts' list (1951) helped designers to allocate functions or tasks either to a human or a machine. Moreover, Bryson's quote illustrates that a superficial "perfect match" does not necessarily provide the best outcome for the human end-user. According to the joint cognitive systems paradigm (Hollnagel & Woods, 2005), the match concerns the collaboration between the human and the machine by focusing on the joint activity of these actors and their joint performance (Figure 1). Based on this paradigm, current research regard machines as members of a human-machine team (cf. Salas, Cooke, & Rosen, 2008, p. 544), for example calling the machine actors electronic partners (ePartners; Neerincx & Grant, 2010) or agents. The present dissertation focuses on the design and evaluation of such ePartners that support dynamic task allocation and coordination in, possibly distributed, teams.

1.2 High-risk Professional Domains

Particularly in high-risk professional domains, such as defense and crisis management, managing the task allocation and coordination during teamwork processes is complex and critical. The envisioned ePartner supports task allocation and coordination during teamwork in such a way that the team can cope with the dynamics of the work environment. The support necessities increase due to the following trends. First, there is a trend to increase efficiency and safety by using technological advanced systems to compensate for crew reduction initiatives or increased situational complexity. The U.S. Navy, for example, targets to reduce the manning of navy destroyers by 60 to 70% in order to lower operational cost mandated by increased technological capabilities (Laurent et al., 2003). Drones serve as another example of reduced cost of ownership and an increased safety facilitated by sophisticated technology ("Flight of the drones," 2011). Using highly technological systems with less people requires a seamless integration of the human and the machine ensuring a proper fit of human capabilities, omitting misuse of the systems, and preventing errors and accidents. Second, there is a trend to economic globalization, which leads to various types of collaboration over organizational, geographical, and temporal boundaries. As an example, the rescue endeavor at the 2010 Haiti earthquake (striking Port au Prince heavily) involved rescue organizations of many different nations. Such ad-hoc deployments in chaotic circumstances require proper coordination of activities in order to work effectively within safety boundaries. The failure to support coordination between rescue teams of different nations at the Pakistani earthquake in 2005 (USAR.nl, 2005) resulted in searching similar buildings, leading to a less effective deployment of resources. Another example where improper coordination lead to inefficient deployment of natural resources concerns the Mont Blanc tunnel fire incident (Sergiu & Luchian, 1999). Rescue teams at both sides of the tunnel were unaware of decisions and assumptions leading to a failure to scale up the crisis organization. As a consequence, the fire went on for 52 hours taking 41 lives.

High-risk professional domains distinguish other domains because of the disastrous and irreversible consequences when incorrect decisions are made. In 1988, for example, the USS Vincennes naval combat frigate mistakenly shot down a commercial Iranian airline because it was misidentified as an Iranian F-14 combat fighter (Klein, 2001). The misidentification led to the launch of a missile hitting the airline killing all on-board. The consequences were evidently disastrous and irreversible.

Much is demanded of the professionals that work in high-risk professional domains. The training is extensive and takes typically multiple years. It takes a naval warfare officer, for example, multiple years of training prior to being authorized to make weapon-launch decisions. In addition, mentality, physical health, and life style are important as around-the-clock operation is normal and horrifying scenes and difficult decisions challenge the human mind.

The naval command and control and the Urban Search and Rescue (USAR) domains are selected because of the intrinsic complexity of task allocation and coordination during distributed teamwork. Both domains provide a natural platform to study and prototype ePartners from the perspective of dynamic task allocation and coordination during teamwork.

The naval command and control domain provides a natural domain to study dynamic task allocation because increasingly complex and dynamic environments lead to excessive workload variation. Human operators work in a sophisticated technical environment with large amounts of complex information to process (Grootjen, Neerincx, & Weert, 2006). In the coastal areas, for example, complexity increases due to asymmetric threats and restrictive legislative rules of engagements. An asymmetric threat is characterized as a civilian entity having a hostile intention, requiring an increased cognitive effort to properly distinguish it from a non-threat.

Multiple rescue teams working at different locations towards the rescue of entombed victims after a natural disaster characterizes the USAR domain. A USAR rescue team, for example, works at a remote working site while the staff operates at base camp. Much effort is required to generate awareness on activities, intention, progress, fitness, and morale of remote actors, at times leading to a coordination breakdown. Coordination breakdowns lower the effectiveness of the mission. The nature of distributed activities provides a natural domain to study the effect of ePartners that support coordination of teamwork while working distributed.

Designing ePartners in high-risk professional domains requires dealing with cognitive reasoning strategies that might differ from classical decision-making theories. Naturalistic decision-making promotes that a number of cognitive functions emerge in natural settings that are not easily replicated in laboratory settings. Classical information processing and decision-making theories fail to consider factors that are inherent to the real world (Klein, 2001) leading to the collapse of some classical cognitive theories when confronted with, for example, time pressure, vague goals, or high stakes. When designing ePartners, it is important to realize that specific cognitive functions might (fail to) emerge.

1.3 Human Machine Collaboration

Figure 1.1 displays the relation between terms that are essential to human machine collaboration. Central is the concept of joint activity that is being defined as an activity "*that is carried out by an ensemble of people acting in coordination with each other*" (Clark, 1996, p. 4). Throughout this thesis, a focus is put on an ensemble of actors that need to collaborate in order to handle complex tasks that are beyond the capacity of a single individual actor. Similar to humans who are goal driven and require activities to achieve those goals, ensembles have a joint goal and require joint activities in order to accomplish these goals. Joint activity leads to joint goal accomplishment and is measurable as joint performance.

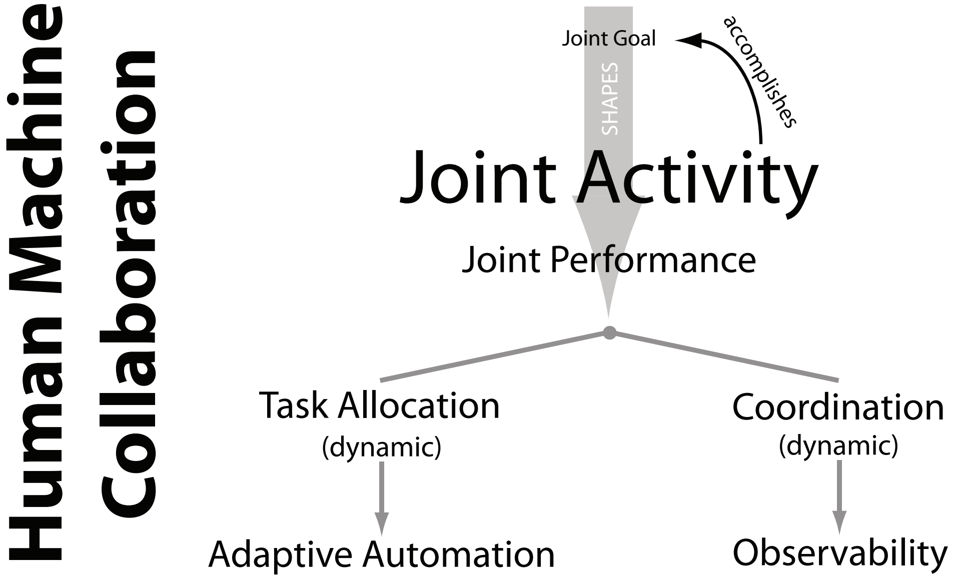


Figure 1.1 – A human and machine collaborate on joint activities leading to a joint performance. The joint activities accomplish a joint goal. Task allocation and coordination are two processes essential in human machine collaboration.

Two processes that are key to collaboration are 1) task allocation and 2) coordination (Figure 1.1). Task allocation refers to a process that assigns specific actors to specific tasks appropriate to the current situation. Said differently, task allocation refers to the process that decides that, for example, actor A is responsible for task 1 while actors B and C are jointly responsible for task 2. Coordination refers to a process in which dependencies between activities are managed. Coordination refers thus to a process that determines that, for example, actor B can only start a specific task after actor A has finished a task. Both the task allocation and coordination are important aspects of human machine collaboration and determine joint performance.

Both processes are substantially affected by the dynamics of the work environments causing additional effort to maintain the level of joint performance. These dynamic task demands may cause overload for the human actors. Overloaded actors have too much work at hand leading to a negative effect on joint performance. In such situations, a process of reallocating tasks often improves the joint performance to the original level (this is coined dynamic task allocation). Adaptive automation is a special case of dynamic task allocation where work is dynamically divided between the human and the machine based on a machine decision to reallocate work. The dynamics of the work environment in which these processes take place increase the complexity of the coordination processes that are needed to integrate and complete tasks within established temporal constraints. The complexity of the coordination process becomes even more complicated when teams are separated in geography or time. This separation leads to additional cognitive costs because it impedes the observability of the team members' activities that might require some coordination with corresponding effort (e.g. phone calls, progress reports). The common denominator in these distributed settings is the failure to directly observe actions or responses and sense states of remote actors. Heath & Luff (1992) highlight the value of observing activities that benefit the coordination of joint activities leading to superior performance. These observations help to anticipate information processing needs and create an awareness of the weak spots in the team. Observability is proposed as a way to make performance, behavior, intention, task progression, and conditional information of the remote actors visible using human interaction technology. Said differently, observability allows actors to detect remote co-workers in the shared environment leading to comprehend what they are doing and how this impacts the joint tasks.

1.4 Human ePartner relationship

A human ePartner relationship follows the joint cognitive systems (Hollnagel & Woods, 2005) paradigm shift from *automation extending* human capabilities to *automation partnering with* the human. The computer should be regarded as an electronic partner where the human and the ePartner collaborate in a symbiotic relation to achieve the best performance while operating within safety boundaries. An ePartner is a computerized entity that partners with a human (development of a relationship) and shares tasks, activities, and experiences. Similar to human partners, explicit agreements are made and mutual reciprocity exists between the partners. This latter means that you need each other to achieve goals and engage in tasks and activities. ePartners are proposed in various domains such as space missions (Neerincx & Grant, 2010) or self health care services (Blanson Henkemans, 2009).

The central objective is to design and evaluate the effects of an ePartner that collaborates with a human by supporting 1) dynamic task allocation via adaptive automation and 2) coordination via the provision of observability displays.

1.4.1 Adaptive Automation

Various environmental conditions require the human to divide his or her attention between different environmental items resulting in a varying workload, at times leading to a cognitive over- or under-load. These over- or under-load conditions hamper effective working significantly due to over-stimulating or under-stimulating our cognitive system (Wickens & Hollands, 2000). The capacity of automated systems in which the division of labor between human and automation is flexible and responsive to task or human demands is called adaptive.

Adaptive automation refers to an approach that dynamically divides work between the human and machine based on a machine decision to reallocate work (Hancock, Chignell, & Lowenthal, 1985; Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992; Rouse, 1988; Scerbo, 1996). In contrast to adaptive automation, adaptable automation (Opperman, 1994; Scerbo, 2001) refers to a mechanism where a human decides on the reallocation of tasks whereas the decision lies within the machine in the adaptive automation paradigm. The concept traces back to 1988, when Rouse introduced adaptive aiding as a way to have the machine "…*intervene and assume authority*…" (1994, p. 30) but he rejected any "…*conditions under which it is appropriate for computers to unilaterally hand tasks to humans*." (1994, p. 30). However, empirical data (e.g. Parasuraman, Mouloua, & Molloy, 1996) show beneficial effects when automation hands tasks back to the human.

Simply increasing the level of automation has its own problems. Although high levels of automation might help the human in periods of high workload, research has indicated that offering high levels of automation is not necessarily the best solution during periods of low workload for two reasons. First, high levels of automation makes the human a passive monitor (i.e., the human checks the machine on erroneous behavior) in essence pushing the human out of the loop (Endsley & Kiris, 1995). Taking both under-load and overload into account, it is essential to keep the human within a bandwidth of workload. Secondly, skill degradation is seen as another issue of highly automated machines (Billings, 1997; Kaber, Onal, & Endsley, 1999). Generally, adaptive automation is seen as a solution to tackle high workload situations while avoiding the aforementioned risks (Clamann, Wright, & Kaber, 2002; Hilburn et al., 1997; Kaber et al., 2006; Kaber & Riley, 1999; Parasuraman et al., 1996; Wilson & Russell, 2007). Adaptive automation is thus regarded as a trade-off between two interlocking ideas. At times of high workload, the adaptive mechanism should transfer work from the human to the machine. This shift of work does take the human out-of-the-loop but allows the human to cope with increased task demands.

On the other hand, the automation should be reset to lower levels at times of low workload to let the human process all information. The human gets in the loop leading to improved situation awareness. Taking the human in the loop at times of low workload has the additional advantage that skills remain trained.

Literature suggests that this dynamic behavior represents the best match between task demands on one side and the available cognitive resources of a human on the other hand (Parasuraman, Mouloua, & Molloy, 1996; Wickens & Holland, 2000). A number of studies have shown that adaptive automation can regulate workload, improve performance, and enhance situation awareness (Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006; Hilburn, Jorna, Byrne, & Parasuraman, 1997; Kaber & Endsley, 2004; Kaber, Perry, Segall, McClernon, & Prinzel III, 2006; Moray, Inagaki, & Itoh, 2000; Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2003).

1.4.2 Observability

The differentiated functions and roles enable a team to collaborate on problems beyond the limits of the individual. However, collaboration requires addition effort to coordinate with those who are working towards the same goal (Cooke, Salas, Cannon-Bowers, & Stout, 2000). Complicated as it is, coordination becomes more and more difficult when actors are separated in time or space. Having various teams operational at various locations diminishes inter-predictability because limited cues are available that feed what others are doing (Thompson & Coovert, 2006). Inter-predictability relates to a capability to plan actions based on accurate predictions what 'others' will do (Klein, Woods, Bradshaw, Hoffman, & Feltovich, 2004). Skilled teams become mutually predictable through shared knowledge and coordination devices that have developed through experience in working together (cf. Heath & Luff, 1992).

Observability is proposed as a display solution that fills part of that gap. Observability allows actors to detect remote co-workers in the shared environment using human computer interaction technology and comprehend what they are doing and how this impacts joint activity. Observability is defined as "*the perception what 'others' in your environment are doing and how they are operating allowing to determine the impact on joint activities. This requires information about the performance, behavior, intention, task progression, and mental and physical condition of the remote actors*". The advantages of observability displays relate to increased awareness of your team members facilitating the coordination of joint activities and increased resilience to unexpected events. Observability displays show information that isn't primary to the (shared) operational goal but secondary allowing actors in a distributed team to be operational in a variety of changed conditions and unexpected events.

1.5 Research Objective

Bringing adaptive automation to real world settings requires building upon previous laboratory studies (e.g. Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006; Hilburn, Jorna, Byrne, & Parasuraman, 1997; Kaber & Endsley, 2004; Moray, Inagaki, & Itoh, 2000; Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2003). Current adaptive automation task models insufficiently motivate how to divide work within a single task. The models merely describe that complete tasks are allocated and the models lack flexibility towards humans task division models (cf. Miller & Parasuraman, 2007). Preferably, ePartners should be able to handle a division of work alike how humans divide work within a single task. This has the advantage that within a task, responsibility can be delegated gradually. Moreover, current approaches fail to clarify how the end-user determines what levels of automation the ePartner is authorized to reach, which is important because the end-user is the domain expert and can weigh decisions best. In addition, the models in place to trigger adaptive automation (e.g. Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006; Clamann, Wright, & Kaber, 2002; Inagaki, 2000a; Moray, Inagaki, & Itoh, 2000; Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2000; Wilson & Russell, 2007) need to be operationalized to real world settings.

Supporting the coordination of joint activities requires addressing questions that are hardly addressed in literature. Support technologies require not merely inserting technology but also study the impact on teamwork. There exists limited validated knowledge on the effects of coordination support displays on teamwork (cf. Carroll, Rosson, Convertino, & Ganoe, 2006; Dabbish & Kraut, 2008; Rocker, 2009). Questions remain on the effect of observability displays on performance, coordination, and related factors such as backing-up behavior, workload, and communication. Another question that needs attention is the question whether awareness on deviations to predefined plans leads to timely responses and adequate actions (cf. Feltovich, Bradshaw, Clancey, Johnsn, & Bunch, 2008).

Consequently, the research objective of this dissertation reads:

Design an ePartner in high-risk professional domains that varies its authority on tasks in response to workload dynamics and supports the coordination of joint activities in distributed settings, all leading to improved joint performance.

In the first part, a focus is put on a framework that is capable of reassigning work to the ePartner at a fine-grained level. A framework needs to be designed from the philosophy that both the human and the ePartner observe elements in the environment and create a mental representation of the environment. Moreover, the ePartner should to be able to adjust the work division using predetermined working agreements to lower the workload of the human or, in case of under-load, transfer more work to the human. Furthermore, the framework needs to be able to make a division between critical and less critical elements in terms of severity or responsibility or those that are more repetitive and monotonous in comparison to cognitive demanding elements. Therefore, the first key objective reads (chapter 2, see Figure 1.2):

(1) To develop a framework capable of dividing the work between the ePartner and the human according to predetermined working agreements fitting the whole chain of information processing in a high-risk professional domain.

One of the challenging factors in the development of successful adaptive automation concerns the question of *when* changes in the level of automation must be effectuated. The ePartner therefore needs to observe the human to determine over-load and underload situations. Previous papers (Rouse, 1988, 1994; Scerbo, 2001) discuss the idea of 'the workload being too high or too low' as a reason to instigate a reallocation of work between the human and the ePartner. At the same time, Gopher & Donchin (1986) acknowledge that it remains difficult to give workload a concrete form. The present dissertation aims to keep the human in a bandwidth of workload requiring a study on available indicators. Consequently, our second key objective is (chapter 3, see Figure 1.2):

(2) To identify adaptive automation triggering models that assess the momentary capacity of the human and the task demands upon the human.

In addition to designing ePartners that are capable of delegating work in a fine-grained way, we are also interested in the effects of such systems on human performance in a high-risk professional domain. Given that performance and workload are interrelated constructs, our third key objective is (chapter 4, see Figure 1.2):

(3) To determine the effect of the adaptive object-oriented task model on the performance and workload of navy professionals.

While the first three key objectives relate to the ePartner's role to initiate a new division of work (i.e. adaptive automation), the following objectives relate to the ePartner's role to support coordinating joint activities when being separated in time or space. The present dissertation proposes observability as a way in which human computer interaction technology overcomes temporal or spatial boundaries requiring a discussion on the problems related to working distributed. It is also important to discuss the hypothetical benefits and costs of using an observability display. Therefore, our fourth key objective reads (chapter 5, see Figure 1.2):

(4) To define which elements are important to present on an observability display that increase awareness of remote actors.

Task complexity and team experience are two factors that potentially influence the use of observability displays. Complexity is known to inflict conflicts in goals and tasks that negatively impact the coordination process. On the other hand, team experience leads to developed knowledge structures that facilitate the coordination process on its turn reducing the need to fall back on other coordination tools such as observability displays. Consequently, the fifth key objective is (chapter 6, see Figure 1.2):

(5) To understand whether coordination and the frequency of use of the observability display changes when a team gains experience and when the task gets less or more complicated.

The aim of an observability display is to improve the coordination of joint activities. Coordination can manifest in many different forms. The most common effect of an improved coordination process surfaces as increase in performance. However, there are a number of alternative manifestations of improved coordination. Backing-up is a manifestation of the coordination processes. McIntyre and Salas (1995) emphasize the importance of backing-up behavior as a component of teamwork. Therefore, the sixth key objective is (chapter 7, see Figure 1.2):

(6) To determine the backing-up behavior effects of using an observability display to coordinate joint activities when being separated geographically.

The previous three key objectives provide a theoretical and empirical understanding about the effects of observability displays on team performance and team-related factors. The main challenge is to apply the acquired knowledge to an observability display that is situated in the Urban Search & Rescue domain. Urban Search and Rescue is a highly dynamic environment leading to deviations to plan. However, the observability display should allow actors to be at the right place, at the right time. Therefore, the seventh key research question is (chapter 8, see Figure 1.2):

(7) To test whether observability displays lead to improved performance and coordination in situations that not always follow the predefined plan.

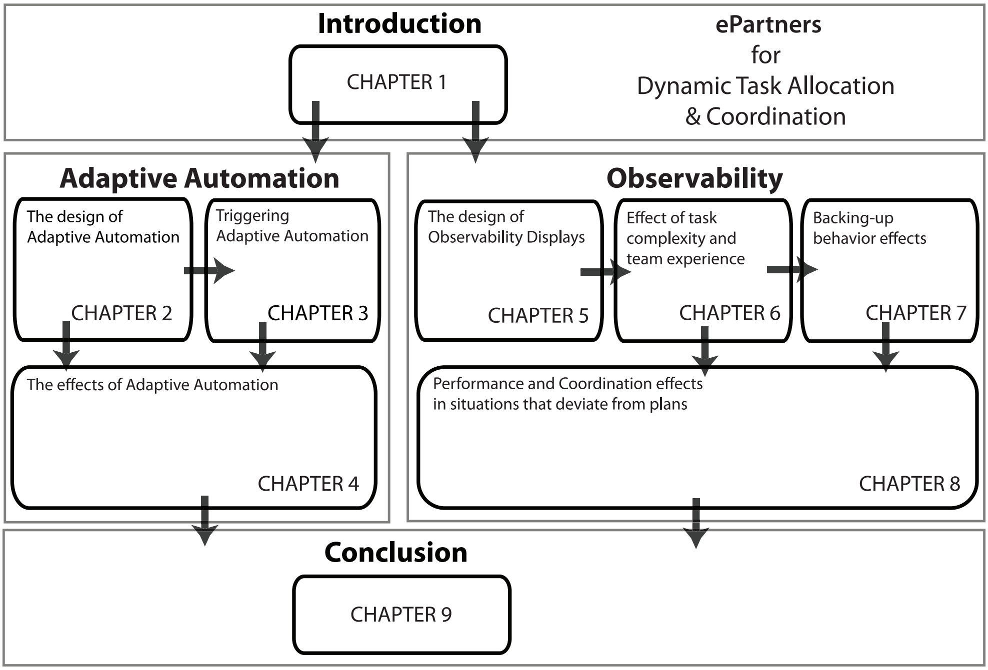


Figure 1.2 – Dissertation outline

2

AN OBJECT-ORIENTED APPROACH TO APPLY ADAPTIVE AUTOMATION IN THE WILD

ABSTRACT — There is a continuing trend of letting fewer people deal with larger amounts of information in more complex situations using highly automated systems. In such circumstances there is a risk that people are overwhelmed by information during intense periods or do not build sufficient situational awareness during periods that require little attention. Adaptive automation provides a solution. A number of studies show encouraging results in increasing the efficiency of human-machine systems by making the automation adaptive in response to human workload. However, these studies are mainly conducted in laboratory settings. An alternative work division model is presented that focuses on the objects central in the domain and allows implementing adaptive automation in real world settings. A fine-grained adaptation framework is proposed that is based on easy comprehension and acceptance by the end user. The machine is regarded like a virtual team member in that it continuously builds its own view of the situation independent from the human. In addition, working agreements between human and machine provide lower and upper bounds of automation that are in advance determined by the end user to avoid undesirable authority taking by the machine. The framework is applicable across a wide range of complex systems because it takes the objects that are central in the domain as a starting point. It gives researchers a framework that they can use to get adaptive automation up and running relatively quickly and easily.

This chapter is predominantly based on*:

Arciszewski, H.F.R, de Greef, T.E., and van Delft, J.H. (2009). Adaptive Automation in a Naval Combat Management System, *IEEE Transactions on Man, Systems, and Cybernetics: Part A: Systems and Humans*, 39(6), 1188-1199.

**The first and second author equally contributed to this study and agreed to alternate first and second authorship for this journal publication and the publication referred to in chapter 4.*

2.1 Introduction

In many domains (e.g. air traffic control, military command and control, crisis management) humans are assisted by computer systems during their assessment of the situation and their subsequent decision making. A continuous technology push has led to innovative but at the same time complex systems. Technological development has enabled humans to work more efficiently and/or effectively using such systems. In such information-rich and dynamic environments, however, a competition for the users' attention is going on between numerous different information items, at times leading to a cognitive overload. This overload originates in the limitations of human attention and constitutes a well-known bottleneck in human information processing. Research has indicated repeatedly that aiding the crew by as much automation as technologically feasible does not necessarily lead to a better performance (Parasuraman & Riley, 1997; Woods, 1996). Prolonged periods of low activity (i.e., underload) lead to performance degradations because the operator gets out of the information processing loop as he or she becomes a passive monitor (Endsley & Kiris, 1995). Taking both underload and overload into account, it is important the keep the human within a bandwidth of workload for optimum performance. In order to reach an optimal human-machine collaboration, research is required to attain the right balance between technologically feasible levels of automation on the one hand, and human requirements and responsibilities on the other hand.

Various studies have been conducted that provide indications on the level of control that can be allocated towards a human or a system (for an overview see de Greef, van Dongen, Grootjen, & Lindenberg, 2007). Since 1951 various suggestions have been proposed starting with Fitts's list (Fitts, 1951), continuing with the taxonomies of Sheridan and Verplank (Sheridan & Verplank, 1978) and Endsley (Endsley, 1987), and finishing with the model of Parasuraman et al. (Parasuraman et al., 2000). In this last model, information processing is divided into four stages (information acquisition, information analysis, decision & action selection, and action implementation) and each stage can be automated at a different level. In the military world this four-stage information processing loop is usually referred to as the OODA (Observe, Orient, Decide, Act) loop first introduced by Boyd (Coram, 2002). Parasuraman et al. (Parasuraman et al., 2000) propose to choose a type and level of automation based on primary (e.g., human performance consequences) and secondary (e.g., automation reliability and costs of action) criteria. They argue for the application of higher levels of automation when applied to the sensory and action levels (information acquisition and action implementation) compared to the cognitive levels (information analysis and decision and action selection) (Clamann et al., 2002; Endsley & Kaber, 1999; Parasuraman et al., 2000). High levels of automation in the information analysis phase can severely impact the situational awareness of the humans and make it difficult for them to monitor proper system behavior and to correct system errors when these occur (Parasuraman, Mouloua, & Molloy, 1996). Likewise, high levels of automation in the decision making phase make it difficult to ensure that proper decisions are indeed being made.

Even if the system designer gets the amount of automation right, (highly) varying circumstances will still produce a (highly) varying workload. Hence, a flexible division of work between the human and the machine is seen as a solution to those varying workload levels. The approach to a dynamic division of work between the human and automation is called adaptive automation (Hancock, Chignell, & Lowenthal, 1985; Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992; Rouse, 1988; Scerbo, 1996). Although Rouse proposed to have the automation "*…intervene and assume authority…*" (Rouse, 1994, p. 30) he rejected any "*…conditions under which it is appropriate for computers to unilaterally hand tasks to humans.*" (Rouse, 1994, p 30). However, empirical data (e.g. Parasuraman et al., 1996) show beneficial performance effects when automation hands tasks (back) to the human, thereby overcoming a number of pitfalls related to highly automated systems. Consequently, this dissertation disagrees with Rouse's limitation and positions adaptive automation as a trade-off between two interlocking ideas. At times of high workload when human information processing limitations emerge, the adaptive mechanism should transfer work from the human to the automation. And, adaptive automation should, at times of low workload reset the automation to lower levels of automation to let the human operator observe all information, thereby increasing situation awareness (Endsley & Kiris, 1995), letting him overcome over-reliance issues, and reducing loss of skill (Billings, 1997).

Adaptive automation thus refers to a mechanism that aids the human operator in real-time by managing his or her workload, the latter fluctuating due to varying environmental conditions. Literature suggests that this dynamic behavior represents the best match between task demands on one side and the available cognitive resources of a human on the other hand (Parasuraman et al., 1996; Wickens & Hollands, 2000). A number of studies have shown that adaptive automation can regulate workload, improve performance, and enhance situation awareness (Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006; Hilburn, Jorna, Byrne, & Parasuraman, 1997; Kaber & Endsley, 2004; Kaber, Perry, Segall, McClernon, & Prinzel, 2006;

Moray, Inagaki, & Itoh, 2000; Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2003). These results highlight some of the potential advantages of adaptive automation.

Whether coined adaptive automation (Scerbo, 1996), dynamic task allocation, dynamic function allocation, or adaptive aiding (W. Rouse, 1988), they all reflect the real-time dynamic reallocation of work in order to optimize performance. In contrast with adaptive automation, *adaptable automation* (Opperman, 1994; Scerbo, 2001) refers to a mechanism where a human makes the reallocation decision whereas the decision lies within the automation in the adaptive automation paradigm.

This chapter puts forward the deployment of adaptive automation to a complex domain, more specifically to naval command and control (C2). The work has been part of a larger research program that investigates adaptive teams and adaptive automation for the Royal Netherlands Navy (RNLN). The RNLN is preparing for a future in which a large variety in missions will have to be undertaken and executed in new and demanding environments with smaller crews. The last years have seen a marked shift of operational deployment from open-ocean ('blue water') to littoral waters ('brown water') in the vicinity of hostile territory where missions are largely in support of land operations. An extended range of threats characterizes littoral operations. Besides the danger from traditional platforms (military ships and aircraft), the operational area is covered by land-based weapons (guns, missile launchers) and there is an increased chance of asymmetric attacks by small surface vessels and civilian aircraft. Situation assessment is made more difficult due to the presence of numerous neutral and civilian entities, smaller detection ranges (and thus reaction times) and stricter rules of engagement. In addition, the amount and complexity of available information continually increases because of, among other things, better sensors and communication and information technology. At the same time, crews are being scaled down due to increasing maintenance and personnel costs.

In order to keep in line with these developments, the RNLN needs to have flexible teams that can adapt to dynamic operational situations. The advantages of adaptive teams are a better chance of fulfilling mission goals and a more efficient deployment of personnel. Adaptive automation in turn is intended to aid the crew in this continuing adjustment to the changing environment.

The central idea of the proposed approach is to have adaptive automation help the human operator focus on the high priority (difficult) cases at times of high workload while ignoring less important or critical work by letting the automation take care of this, and at times of low workload to let the human deal with all processing. This requires to design adaptive automation from a joint activity perspective

(Hollnagel & Woods, 2005, pp. 67-68; Klein, Woods, Bradshaw, Hoffman, & Feltovich, 2004) in that the automation should be regarded as a virtual partner like a human actor. In this thesis the automation is considered a (junior) team member capable of delegating work too and therefore following a philosophy that both the human and the automation observe the elements in the environment and create a mental representation of this environment (Figure 2.1). The automation, however, additionally observes the human operator to determine an overload or an underload situation. Whenever such a situation occurs, the automation can adjust the current work division using predetermined working agreements and reassign work in order to lower the workload of the human, or, in case of underload, reset itself and transfer more work to the latter.

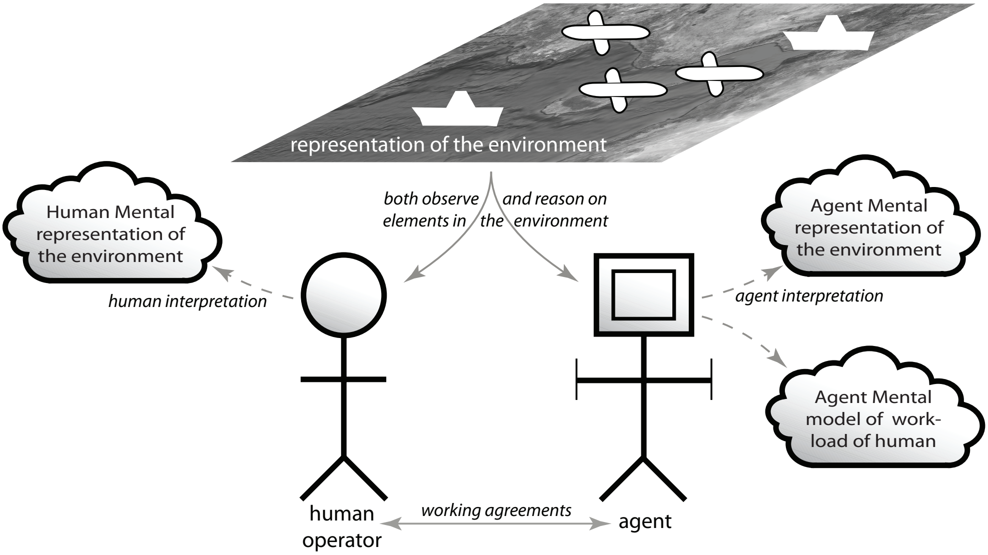


Figure 2.1 – Both the human and the automated machine observe elements (objects) in the world and create a mental representation of the elements. The machine also observes human reaction to these elements and enabling the machine to spot overload and underload situations allowing it to reassign work (i.e., taking over some tracks identification tasks) using predetermined working agreements.

This chapter discusses the implementation of adaptive automation for the C2 identification task requiring an elaboration on *how* adaptive automation is implemented using an object-oriented task model. The question *when* adaptation should take place (i.e., which conditions should trigger the automation to adapt) is discussed in the next chapter of this dissertation (chapter three). Chapter four of this dissertation discusses the effects of such an adaptive automation system by comparing a number of dependent measures (e.g. performance, workload, accuracy, and timeliness of decisions) when working with adaptive automation to working without such a mechanism using both classical attack scenarios and smuggling scenarios.

This chapter continues with an introduction to the domain in section 2.2, sections 2.3 to 2.7 describe the task allocation model using an object-oriented framework. Section 2.3 explains the rationale behind the object-oriented framework and section 2.4 discusses the importance of separating the view of the human and the view of the machine. Section 2.5 considers the usage of five levels of automation and section 2.6 and 2.7 combine the levels of automation with the object-oriented approach leading to a fine-grained distribution of work. How such task allocation can be made adaptive is the subject of section 2.8. Section 2.9 shows how the approach to adaptive automation can be interpreted as a set of working agreements in the human-machine team. The question of when to shift autonomy is taken up in section 2.10 prior to summarizing and drawing conclusions about applying the object-oriented approach ‘in the wild’.

2.2 The Domain: Naval Command and Control

Because adaptive automation is applied to the naval C2 domain, a brief introduction to this domain is provided. Among other things a Combat Management System (CMS) supports the team in the command centre of a naval vessel with its tactical work. This means that operators continuously execute all the stages of information processing in the naval tactical domain and in addition they have to build a situational picture of the surroundings of the ship (including comprehension of the situation and an extrapolation into the future) and the potential undertaking of offensive and defensive actions. As already mentioned in the introduction, this is in the military known as the OODA loop. The loop is similar to the information processing model of Endsley (1987) and Parasuraman et al. (2000). The loop can be further subdivided into distinct tasks like correlation, classification, identification, threat assessment, and engagement. Correlation is the process whereby different sensor readings are combined and integrated over time to generate a track. The term track denotes the representation of an external platform within the CMS, including its attributes and properties. Classification is the process of determining the track’s type of platform (e.g., an F16 fighter aircraft or an Arleigh Burke class destroyer) while the identification process attempts to determine its identity or allegiance in terms of it being friendly, neutral, or hostile to the ship. The threat assessment task assesses the danger a track represents to the own ship or other platforms. At this stage, the information becomes more abstract as singular tracks are bunched together in larger aggregates like military formations and tactical patterns that need to be interpreted as a whole. The engagement task includes the decision to apply various levels of force to neutralize a threat and the execution of the decision. Track attributes like height and speed need to be monitored continuously because these variables are input for more abstract functions like adherence to an air lane or formation in the identification process. Therefore monitoring is also part of the duties of a command team.

All the tasks described above are currently handled in large part by the crew. Therefore they must be replicated in algorithmic form in order to be able to automate the process and be made adaptive. An adaptive CMS could provide naval crews with an answer to the looming risk of operator overload due to increasing information processing requirements and manning reduction initiatives. It should be clear from the outset, however, that there is a definite need to have different levels of automation for these tasks. Apart from considerations of whether it is technically possible to fully automate all of them or sensible in terms of human factors considerations, the question of responsibility immediately emerges when thinking about automating the engagement process. There are less worries with respect to the automation of track correlation or classification. Section 2.9 discusses working agreements and deepens the question of responsibility and how these can be regulated appropriately. But prior to that the object-oriented framework is discussed.

2.3 An Object-Oriented Task Model to Implement Adaptive Automation

The adaptive automation paradigm shifts control dynamically between the human and automation and one major question is, how and to what amount shift control? Literature catalogs a number of models that describe the level of automation that can be allocated towards a human or automation, starting with Fitts' list (1951), continuing with the models of Sheridan & Verplank (1978), Endsley (1987), Endsley & Kaber (1999), and ending with the four-stage model of Parasuraman, Sheridan & Wickens (2000). The latter model is based on human information-processing theories and accordingly recognizes four sequential stages. The model states that all of these stages should be automated at a different level of automation based on primary (e.g., human performance) and secondary (e.g., automation reliability) criteria and is the most sophisticated model to reason about levels of automation.

Although the model describes the level of automation on an abstract level this does not lead to a straightforward *implementation* to share work between a human and automation in 'real world' operational circumstances. This requires, to be more precise, to overcome three shortcomings. First, the four-stage model of Parasuraman et al. (Parasuraman et al., 2000) requires a more fine-grained division of work. The four-stage model states that each (C2) information processing step should be automated at a different level. In the C2 domain, for example, the model enables the identification task to be automated at a high level of automation autonomy while the weapon engagement level can be automated at a low level of automation autonomy. However, this rather abstract way of dividing work does not explain how to divide the work for a single task. One approach is to recursively apply the four-stage model leading to a subdivision of work based on an information processing perspective. However, this is not how humans tend to delegate or divide work when things get busy. A more human way to divide the work follows statements like "you take the contacts in the north" in the military domain or "you take all other approaching aircraft while I focus on the two non-separated aircraft" in the air traffic control domain. In other words, the work is divided by assigning the *objects* (tracks, contacts) the task is dealing with to the operator and not by further dividing the processes involved in the task. Specific characteristics of task stimuli (e.g. tracks, contacts) may dictate different information processing functions or needs. This opens the way to a framework that allows for a more granular division of a task that links closely with how humans delegate work while taking the information processing perspective into account. Second, the position is taken that the automation should aim to take over those parts of the work that are less critical in terms of severity or responsibility or that are more repetitive and monotonous. Again, this can be decided on a task-by-task basis, but again objects that are deemed more problematic can be singled out for special treatment as well (compare the 'two non-separated aircraft' described above). Third, the end-user rather than the system designer should have the final say as to what levels of automation the machine is authorized to reach as the end-user is the domain expert and can weigh the problems and priorities best. Human control is especially important in domains where incorrect decisions lead to outcomes with a high toll (e.g. military command and control). This requires that the automation levels be phrased in terms understandable to the users.

The idea behind the object-oriented approach is as follows. During the information processing loop in domains like military command and control or air traffic control, the operators and the system are dealing with large numbers of real-life objects (military platforms, aircraft). The representation of these objects in the control systems are usually called *tracks*, so in these domains tracks are the domain-related instantiation of objects. In other domains, airplane or network-package objects play a similar role. For each of these real-life objects, a number of tasks are executed (for example classification, identification, possible engagement in the military domain; assignment to a holding pattern or an approach lane in the air traffic domain). The total work involved in each of these tasks thus can be thought of as a repetition of the same task for each object in turn (each track needs to be identified, each aircraft needs to be guided to the right runway). Then, while sharing a task between man and machine, it is much easier to *divide the objects* the task is dealing with between operator and system than let them somehow share the processing the task involves. This way of dividing work between man and machine matches the way crews divide work as well. Now, objects can simply be divided between man and machine in a numerical way (human takes everything when things are slow, fifty-fifty when the workload becomes heavier), but a more subtle approach is easily accommodated. In this approach, the system takes the easy (standard) objects and the human is left with the more difficult ones. In the air traffic domain, the easy objects would be aircraft that are widely separated from other aircraft and the more difficult cases would be aircraft in close proximity and aircraft landing and taking off. In emergencies, most non-problematic aircraft could be handled by the system based on standard rules, allowing the operator to concentrate on the aircraft in trouble. In the military domain, easy objects are e.g. neutral platforms that follow a predefined air or sea way and identify themselves as commercial transports. This object-oriented approach seems suitable for all domains where tasks involve multiple, similar objects.

To be able to divide the objects in a more sophisticated way than by numbers, the proposed object-oriented task model recognizes five levels of automation running from no support via manual, advice, consent, and veto to system. Each of these automation levels is clearly distinguishable by the level of autonomy of the automation and the signaling of warnings. These levels are discussed in depth in section 2.5 subsequent to a discussion about the separation of user and systems worlds in section 2.4.

2.4 System View and User View

Instead of letting a task be performed either by the human, the machine, or a combination of both, a mechanism is put forward for whereby both parties do their job concurrently. In this way each party arrives at their own interpretation of the situation, building their respective world views at the same time (Figure 2.2). It is thus assumed that the system is capable of deriving its own interpretation of the tactical situation and making its own decisions. Using the interaction between these two separate world views is the base for the adaptive mechanism. One important aspect of this concept is the fact that the system always calculates its view and makes its decisions, independent of whether the user is dealing with the same problem or not.

The first step toward an implementation is the provision for 'storage space' where the two parties can deposit the information pertaining to their individual view of the world. Thus two separate data spaces are provided (which are called 'system world view' and 'user world view', respectively) where the results of their respective computational and cognitive efforts can be written to. Having these two distinct world views is akin to two people coming to different conclusions based on the same set of (non-conclusive) data. The system may be viewed as a team member with a limited but scrupulously objective view of the world.

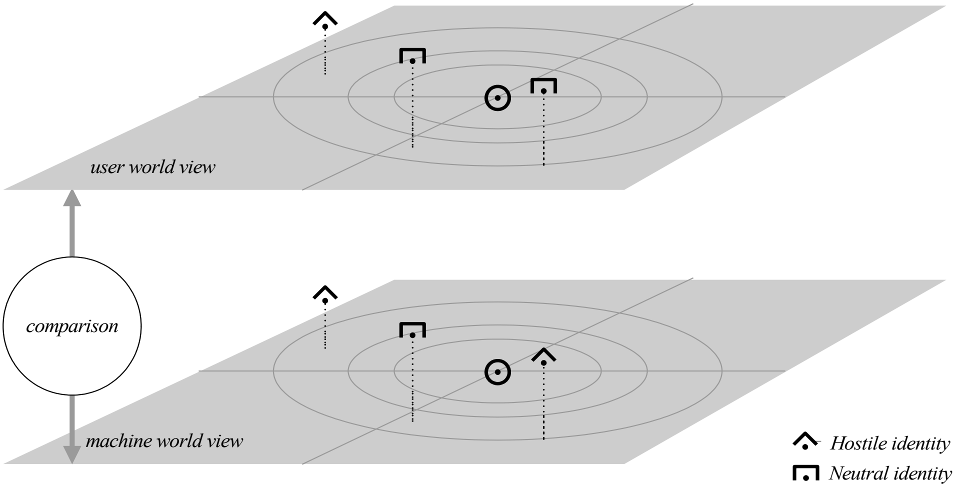


Figure 2.2 – Both the machine and the user world have an interpretation of the world based on the system's computational power and the user's intelligence that is stored in their respective world views. In this case, each view has three tracks and the center represents the navy ship and the circles represent different ranges. The machine world view recognizes two hostile and one neutral track while the user world view recognized two neutral tracks and one hostile. The user and machine world view can be compared for differences in this case producing one different identity.

As such, the user space serves two purposes. First it is a method to create and maintain common ground between the system and the user. When the system and user are engaged in joint activity (i.e., the system becomes a virtual team member), they should maintain a common ground in order to limit coordination breakdowns

(Klein et al., 2004). Second, when multiple actors are engaged in the process, the user world view can serve as a 'blackboard' where a graphical picture of the tactical situation unfolds as different actors add their information. In other words, it can be used to maintain common ground between the various players as well. In this capacity the user world view reflects the usage in current CMS, where only user-controlled attributes (e.g., class, identity) are stored. The data in the user world view are leading in decision making further on in the C2 loop. In other words, a decision to engage a track will depend among other things on its identity in the user world view.

Another advantage of these parallel world views is the increased transparency of the automation, both to the designer and the user. In addition, the separation of the domain algorithms and the processing required for adaptive behavior is another advantage. Finally, the dichotomy between user and system views makes it relatively easy to retroactively add the framework to existing systems. The fact that a 'double bookkeeping' is required is a disadvantage but the advantages outweigh this extra system load.

2.5 Automation Levels

Starting from the system and user world view, four distinct Levels of Automation (LoA) follow logically when a) autonomy is being shifted between the human and the machine and b) whether signaling is executed actively by the machine or not. A fifth level of automation is added in case the machine is not available. Table 2.1 displays the five levels of automation. The *advice* LoA represents the level where the system world view is available for advice. In other words, if users are granted access to the system world view, users always have a secondary opinion available. In the next LoA (*consent*), the machine compares both views and it alerts the human to any discrepancies. Of course, system space is still available for inspection. The signaling only occurs when the discrepancy occurs or when additional evidence is found for a deviating system view. This signaling functionality represents the consent LoA. Advice thus denotes a passive level, where the user must actively pull the opinion of the system (from system world view) and the consent level describes an active level, where the system pushes its view actively to the user.

The machine is granted more authority at the higher levels of automation. The machine has the highest authority at the *system* LoA where the machine takes over the responsibility. One level lower, the *veto* level grants the machine has the same responsibility compared to the *system* LoA but also alerts the human about actions taken allowing the human to intervene. In architectural terms, the machine gets write access to user space in veto and system level and is consequently able to inject its own decisions into the human information processing loop. The only difference between these two levels is that in veto the system warns the user of its decision thereby allowing the user to reject and 'veto' the decision of the system whereas in system mode the system performs the same action in silence. Currently four separate levels of automation are distinguished: advice, consent, veto, and system. A fifth, *manual*, was added to denote the fact that no automation is present for whatever reason (technological, ergonomic, political, financial, etc.). Table 2.1 summarizes the automation levels.

The levels of automation are based on the distinctions between the user and system world view in terms of authority and signaling (Figure 2.3). Signaling happens when the system wants to inform the human of the fact that there is either a discrepancy between the data in the system world view and user world view or a decision has been made by the system that the user may wish to revoke (veto). Authority is defined in terms of whether the system has write access to user world view (Figure 2.3). The idea is that whoever (or whatever) enters data in the user world view has the final say with respect to the task associated with the data. For example, if the machine is allowed to change a track identity in user space, it has the authority for the identification process for that track. And if the machine has authority to alter a track identity, it is allowed to write its identity to the user space. It is in this sense that the data in user world view are 'leading' whereas the user may not actually have that view.

With system and user world view and automation levels, there are enough building blocks to construct an adaptive system by allowing the machine access to the user view in a controlled and gradual manner. This would involve switching between different levels of automation. But before doing so, an object-oriented work allocation approach is proposed. The object-oriented approach allows a granular division of work between human and machine resembling to how humans in a team divide work.

Table 2.1 – Summary of levels of automation used in the object-oriented task model

Level of Automation	Description	Signaling
Manual	No automation is available or allowed to assist the user	No
Advice	The human keeps all responsibility. The automation is available for advice but the automation takes no initiative ("pull").	No
Consent	The human keeps all responsibility but the machine alerts the human to changes in the situation ("push").	Yes
Veto	The human delegates the responsibility to the automation *unless overruled* (vetoed).	Yes
System	The human delegates all responsibility to the automation and there is no interaction between the automation and the human.	No

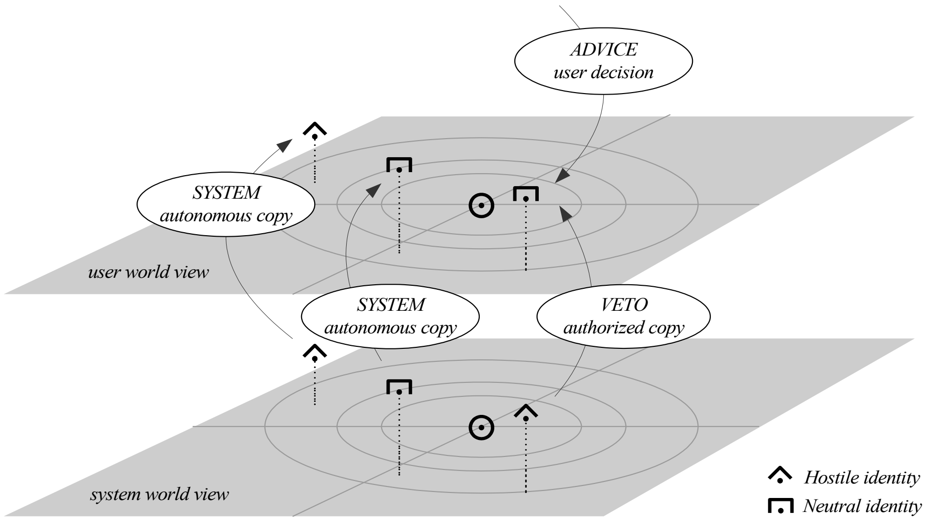


Figure 2.3 - Authority is defined as equivalent to the machine's write access to user space. Either the machine copies its view to user space (autonomous copy in SYSTEM mode) or the human determines what is written in user space (VETO, CONSENT, ADVICE).

2.6 Object-Oriented Work Allocation

Similar to the paradigm shift from functional programming to object-oriented programming, the focus on objects rather than tasks seems fruitful (cf. Bolderheij, 2007). In object-oriented design and programming, objects are abstractions of real-world things or entities that share characteristics and conform to similar sets of rules and policies (Rumbaugh, Blaha, Premerlani, Eddy, & Lorensen, 1991; Shlaer & Mellor, 1992). A number of domains yield objects that have a very tangible and physical nature while other domains concern themselves with more abstract items. In the operation of an airport, objects like airplanes, runways, and air lanes quickly come to the fore whereas communication systems will be concerned with objects such as data frames and acknowledgements. In this terminology, all objects have attributes that lay down the characteristics of the real-world entities that they represent, such as height, temperature, registration number, or location. Furthermore, objects (at least the more interesting ones) can be considered to have a state describing in overall terms the condition of the object. Although the state generally can be derived from the attributes of the object, it usually makes sense to add a state-like description to an object. For an air traffic controller, for example, an aircraft is either within or outside his or her air space, waiting for a landing slot, landing or taking off, taxiing, or being parked.

Once the focus has been put on objects, tasks return into the picture as the processes related to the objects. For example, tasks that can be associated with aircrafts in the air traffic control domain are assignments to an air lane and continuous collision monitoring. One of the advantages of the use of objects is that it is much easier to pin down what exactly it is that a task is trying to do, namely to create new objects, to assign values to attributes, to establish (or remove) associations between objects, and so on. A task like 'situation assessment' is hard to define, but the sharply outlined purpose of the identification task is to fill in the identity attribute of a track object. The focus on objects does not mean that tasks disappear; it is only that the emphasis is on objects first. In air traffic control, most tasks involve collecting, processing, and inspecting data about aircraft tracks and updating or generating other information related to the same objects. In the naval domain, the prime objects of interest to the crew are the platforms present in the tactically relevant surroundings of the own ship. Because these platforms are represented by tracks in the CMS, the latter form the actual objects of attention. Section 2.2 discusses some of the tasks that are associated with tracks in the naval command and control domain. They include classification, identification, threat assessment, and engagement, supplemented by behavioral monitoring (often not explicitly mentioned, but executed nevertheless). Besides tracks, which represent very tangible objects in the ship's surroundings, more abstract objects like formations and task groups, mission objectives, and tactical patterns also play a role in military command and control. It is at these levels of abstraction that warfare officers switch from perception and interpretation of the current situation to predictions of the future situation. In the proposed work on adaptive automation, the focus is put on tracks as objects for which the information processing can be automated to a reasonable extent.

2.7 Assigning a Level of Automation to Objects

The object-orientation allows for a granular work division between the human and the machine. While the previous section discusses the object orientation in relation to tasks, this section discusses how work is divided between the human and the machine based on objects that are associated with tasks or processes.

If a task, like identification, is considered, this task by definition is performed for each object (track) in turn. The human workload can be tuned for this task by setting the LoA of the individual objects the task pertains to lower or higher, in effect parceling out the objects between the human and the machine. The easiest way to achieve this is by means of the object attributes such as velocity or altitude. Or in other words: let the attribute values of the objects determine what LoA is assigned to each object. The object attributes define sets of objects where each set has a different LoA (Figure 2.4). In order to clarify this idea, track sets for tasks are defined that are to be shared between human and machine. Figure 2.5 and Table 2.2 show that it is fairly easy to define which tracks will be handled by the user and which by the system. Figure 2.5 provides a track set where low-speed air tracks are to be handled fully by the automation, while the high-speed air tracks (the suspicious ones) remain the responsibility of the human. A similar setup but with different boundaries is applied to surface tracks. Table 2.2 shows that domain terminology can be used to assign different LoA to objects as object attributes themselves are part of the domain. In the case of the identification task, the determining attribute is the machine assigned identity; in the case of engagement the determining attributes are identity and class (which must be hostile and missile, respectively). In both these cases the relevant attributes are ordinal types but continuous attributes like range or height could be used as well (Figure 2.5). An additional advantage is the fact that the domain-related attributes allow the user to keep control over those objects that are deemed most important (e.g., the suspect and hostile tracks).

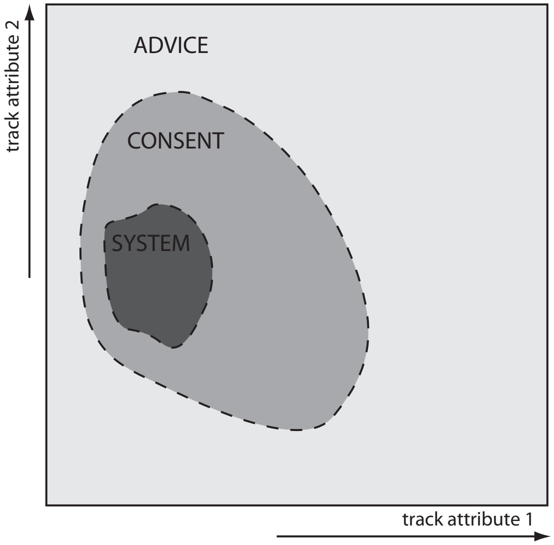


Figure 2.4 - Track sets related to a certain task defined using two track attributes (for example, identity and range or identity and class). Only SYSTEM, CONSENT, and ADVICE are shown. The VETO set is empty and the ADVICE set effectively is what remains from the full track set after the SYSTEM and CONSENT sets have been subtracted.

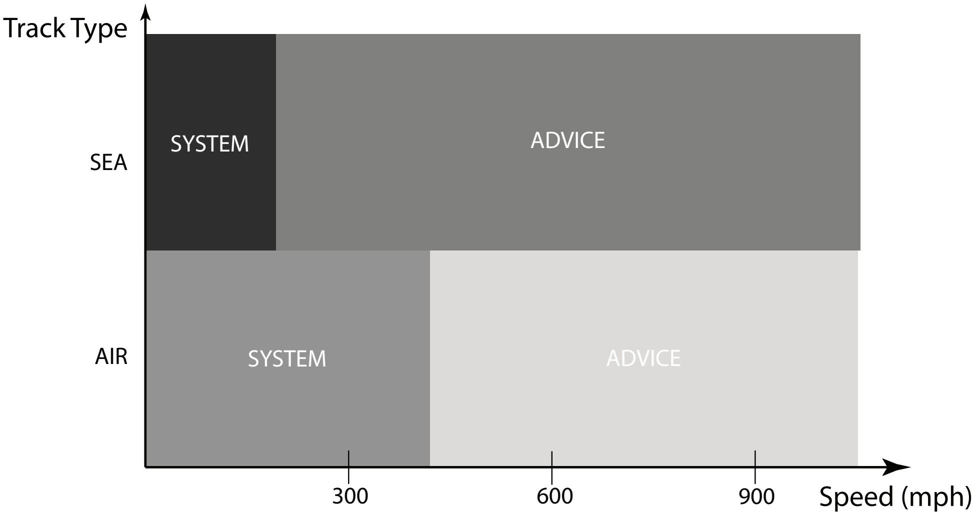


Figure 2.5 - A definition of which objects are in which level of automation based on one or more attributes. Air tracks, for example, traveling with a speed larger than 400 mph are in advice mode making them the responsibility of the human while tracks with a lower speed are the responsibility of the machine (system mode). Sea tracks, on the other hand, use in this working agreement a lower speed to assign tracks in the advice mode or system mode.

Table 2.2 - Division of work between user and system in terms of track attributes for different tasks.

Task	Set Definition	Track Automation Level
Classification	Let the machine automatically classify all tracks	all tracks : SYSTEM
Identification	Let the machine identify all friendly* and neutral* tracks	friendly and neutral tracks*: SYSTEM hostile tracks* : CONSENT **according to the machine's view*
Engagement	Let the machine handle all (hostile) missiles automatically by scheduling weapons against such tracks	hostile missile tracks* : SYSTEM other tracks : CONSENT

It should be clear that during an actual operation, some tracks should be brought to the attention of the human (should be signaled). These include tracks in consent mode where the machine has detected a discrepancy between the system view and the user view, and tracks in veto mode where the machine has already taken steps to correct the situation, but has the duty to inform the human of this fact.

2.8 Adaptive Automation

All that remains in order to be able to shift the workload between the human and the machine, is to be able to adjust the boundaries of the sets (Figure 2.6). When the boundaries of the machine-controlled sets are stretched outward, thus including more tracks, the workload of the human will be reduced as fewer tracks will be his or her immediate responsibility. Even if the responsibility for some of the tracks remains the human's, signaling important changes with regard to these tracks will relieve their workload as less monitoring is required. When the boundaries of the track sets are contracted the workload on the human will increase because more tracks become their responsibility. The boundaries are changed by changing the attribute values that define the sets. For example, by reducing the range at which tracks are identified manually, the set of tracks handled by the machine (either by consent, by veto or system) is increased and the workload of the user is reduced in proportion.

Adaptive automation is achieved by adjustment of the automation levels of the objects, which is accomplished by adjusting the boundary values of the relevant object attributes for the sets concerned. In this way, objects are effectively transferred from the human to the machine or vice versa. Table 2.3 shows two different configurations: one for a low operator workload and another for a high operator workload. The human can set these configurations offline using working agreements.

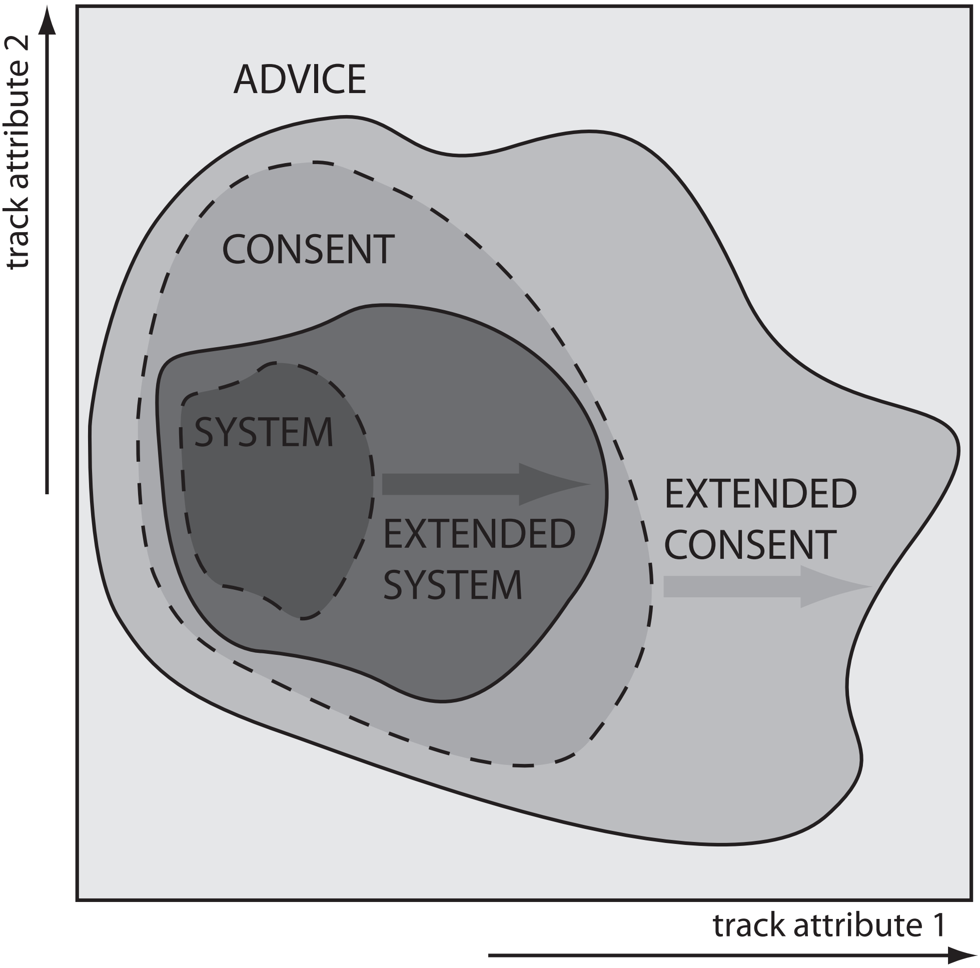


Figure 2.6 - Adaptable track sets related to a certain task defined using two track attributes. By adjusting the set boundaries adaptive behavior is implemented where the machine takes on more or less work thereby adjusting the workload of the human.

Ensuring that more objects (e.g. tracks) are the responsibility of the automation when a different configuration is selected shifts work from the human to the automation (and vice versa, when objects become the responsibility of the human). These working agreements minimize confusion in the operator as to what the machine will take over. Figure 2.7 shows the effect of the adaptive automation mechanism. The left side shows a situation with ten tracks, excluding the centrally positioned navy ship of the operator (the white half circle). Eight of these tracks are gray and require attention as these are tracks that have not been identified yet and two tracks (black) have been identified as hostile and require attention. The right screenshot shows the same display after the automation has been triggered. The adaptive automation system checked its working agreements and was allowed to identify six of the eight tracks as neutral given the identity according to the machine. The remaining two tracks have a high speed and are to be identified by the human operator. In this way, work to be done by the human is significantly lower while allowing the human to focus its attention to these critical objects.

Table 2.3 - Division of work between user and system for different tasks and for different workloads

Task	Operator Workload	Set Definition	Automation Level
CLASSIFICATION	LOW	Limit the system to giving advice with respect to classification	all tracks : ADVICE
	HIGH	Let the system automatically classify all tracks with the possibility for the user to adjust or delimit the system classification	all tracks : SYSTEM
IDENTIFICATION	LOW	Let the user identify all tracks; limit the system to warning the user when something seems amiss with a track	all tracks : CONSENT
	HIGH	Let the system identify all friendly* and neutral* tracks	friendly and neutral tracks* : SYSTEM hostile tracks *: CONSENT **according to the machine's view*
ENGAGEMENT	LOW	Limit the system to giving advice with respect to engagement	all tracks : CONSENT
	HIGH	Let the system handle all hostile missiles automatically by scheduling weapons against such tracks	hostile missile tracks : SYSTEM other tracks : CONSENT

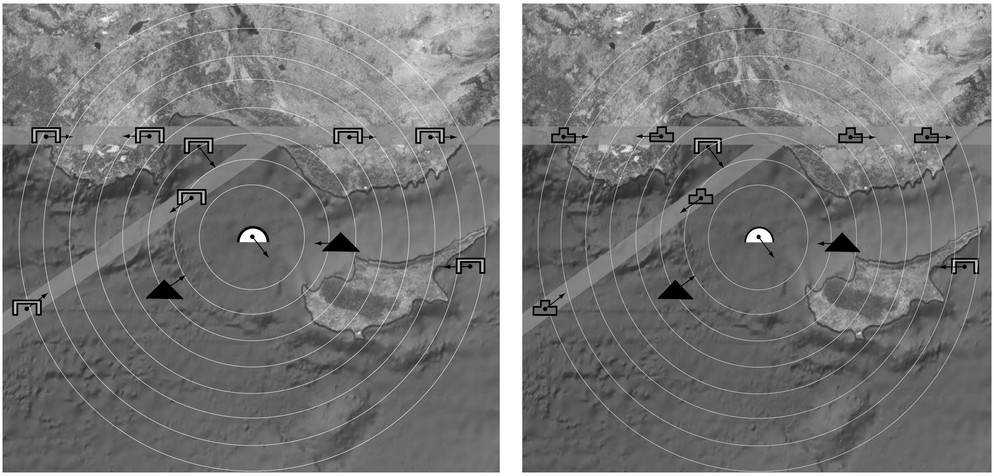


Figure 2.7 - The left shows a situation prior to triggering adaptive automation, and the right side shows the situation after the automation took over some of the work (following the description of Figure 2.5; this means that they are dealt with by the system but a different agreement could set a different automation mode (e.g. consent)). The number of tracks requiring identification (colored gray) has been lowered because the configuration allowed the automation to identify six of the eight tracks that previously required human attention.

2.9 Working Agreements

The moment that high workload conditions are encountered, the adaptive system will increase its authority and assume responsibility for some of the work that was previously the responsibility of the human. At such a time, however, the adaptive system should not negotiate this shift with the human. Were the machine to start a discourse with the human, an extra task would be initiated that would further decrease the latter's performance. It is for this reason that working agreements are introduced.

The general idea behind the concept of working agreements is that as the human and the machine work together to achieve common goals requires both to act as team members. In the field of human teamwork, several people have elaborated the concept of working agreements (Cannon-Bowers & Salas, 1998; Rasker & Willeboordse, 2001). Working agreements aid team members in providing each other with information in time, jointly solving problems, and assisting each other during periods of heavy workloads.

In the proposed framework the machine is regarded as a virtual team member and the different settings of the set boundaries as working agreements. Table 2.3 shows working agreements for three tasks during low and high workload situations, respectively. Using predefined tasks and object attributes, the human instructs the machine on the extent to which authority may be increased prior to a mission.

Furthermore, it is proposed that the human evaluates these working agreements during debriefing sessions following work shifts. Based on the human's most recent experience, the human can improve his or her dealings with the machine by adjusting incomplete or cumbersome agreements.

The role of the working agreements is larger than merely regulating the workload between human and machine. They can reflect the opinions of the system designers and users on the proficiency of the software and be used to establish trust in the workings of the machine by starting at fairly low levels of automation. They will also mirror thoughts on how much authority is ultimately delegated to the machine under the political and strategic constraints of the mission. Peace-keeping operations will likely involve lower maximum levels of automation than a full-out war will allow. The considerations that are taken into account when drafting these working agreements are the same as the secondary evaluation criteria brought to the fore by Parasuraman et al. (2000): reliability and the costs of actions. In the military domain there is a continuous balancing act between the need to further automate processes that require fast reaction times or complicated algorithms (e.g., missile defense) and the need to keep the responsibility for the consequences of these processes firmly in human hands. The working agreements advocated in this chapter allow this to be realized for each distinct task. The primary evaluation criteria described by Parasuraman et al. (2000) (mental workload, situational awareness, and avoidance of complacency) should be met -to some degree at least- by the workings of the adaptive system itself.

2.10 Triggering Adaptation

One of the challenging factors in the successful development of adaptive automation is the question of when changes in level of automation must be effectuated. 'Workload' generally is the key concept to invoke such a change of authority. Most researchers, however, have come to the conclusion that "workload is a multidimensional, multifaceted concept that is difficult to define. It is generally agreed that attempts to measure workload relying on a single representative measure are unlikely to be of use (Gopher & Donchin, 1986). The definition of workload as "…*an intervening variable similar to attention that modulates or indexes the tuning between the demands of the environment and the capacity of the operator*" (Kantowitz, 1988) seems to capture the two main aspects of workload, i.e., the capacity of humans and the task demands made on them. The workload increases when the capacity decreases or the task demands increase. Both the capacity and task demands are not fixed entities and both are affected by many factors. Skill and training are two factors that increase capacity, for example, whereas capacity decreases when humans become fatigued or have to work under extreme working conditions for a prolonged period.

If measuring workload directly is not a feasible way to trigger the adaptive automation mechanism, other ways must be found and chapter three of this dissertation discusses both the theoretical background and some empirical findings.

2.11 Conclusions

The proposed framework for adaptive automation aims at maintaining an operator's workload at a manageable level. This is achieved by adjusting the LoA for the tasks the human must perform, or more specifically, by shuttling the objects these tasks are concerned with between sets of different LoA. Which objects are shuttled to which sets depends on the attributes of the objects, allowing fine-grained control over the transferred work that can be expressed in terms of the domain itself (for example: all hostile tracks or only tracks with high velocities at lower altitudes). Depending on the LoA of an object, a task is either under control of the human or the machine, while the latter may additionally signal discrepancies or gaps in the human's view. The boundaries of the LoA are agreed upon in advance between human and machine, utilizing the concept of working agreements, in order to minimize automation surprises and to ensure the fact that the machine does not take on too much responsibility.

This framework allows a gradual delegation of responsibility to the machine while keeping the human in firm control of those tasks and objects that are regarded important. The human can delegate complete autonomy to the machine or express a wish of being informed about decisions of the machine. Alternatively, the human can keep control over critical items while being optionally informed about potential problems. The levels of automation will be maximized either by responsibility (reflecting the responsibility the users or higher authorities are willing to delegate to the machine) and the comprehensiveness and quality of the software.

Responsibility for the adaptation mechanism itself is delegated according to the same hierarchy; while some researchers keep the human firmly in command (Miller & Parasuraman, 2007), others are more inclined to let the machine balance the workload (Kaber & Riley, 1999; Wilson & Russell, 2007). Thanks to the fine-grained working agreements made possible by the object-oriented framework, a justification for a highly autonomous adaptation process is advocated. Because the circumstances under which the machine adjusts itself and the limitations of these adjustments are proscribed in advance, the human knows what to expect. When the human starts getting overwhelmed by the situation, the machine should step in without forcing the human to think about a better distribution of work or, worse, start a negotiation with the machine with respect to the amount of adjustment. Such an approach stands or falls with a good estimate of the workload, however, and this remains a topic for research and is discussed in the next chapter of this dissertation (chapter 3).

An additional advantage with the proposed form of adaptive automation and the underlying working agreements is the fact that trust in the machine can be built by starting the automation at relatively low levels, with mainly advice and warnings coming from the machine. When humans start to see the benefits (and limitations) of the machine, the levels of automation can be increased to the point where the operators are comfortable with the level of support at each configuration.

What makes all this possible is the fact that the machine continuously updates its own view of the situation (in the system world view) and is only restricted in the expression of this view, be it by having an advice ready, signaling, or taking autonomous action. The major assumption here is that the machine is indeed capable of calculating its own view of the situation. If this is not the case, the LoA for the corresponding tasks must be set to low (advice or consent) or automation must be considered absent (manual). Of course, this applies to adaptive or adaptable automation in general: one cannot have adaptive automation without automation (the underlying algorithms) to start with.

Upon reflection, an object-oriented framework toward adaptive system behavior seems quite natural. The notion that the amount of work relates in some way to the number of objects and that work can be divided by distribution of objects seems intuitive. The approach has the additional advantage that it tends to closely match the domain terminology because the objects form the natural core of the domain. The idea of the machine continuously updating its own view of the situation independently from the human perhaps seems controversial as ordinarily updates would take place only in response to a request or because the human has delegated some of the work to the machine. In this chapter, both ideas are taken leading to a comprehensible, holistic, and consistent framework for adaptive automation.

The proposed framework shows that it is possible to define an adaptive framework using the concept of separate system and user spaces and an object-oriented framework on top of the model of Parasuraman et al. (2000). Besides its use in a naval command and control system, the framework looks applicable across a wide range of complex systems, both military and civilian, because the inherent assumptions about the problem domain are limited in scope. Obviously, the domain should be able to put into automation to some extent, be able to model in objects (preferably large numbers of similar ones), and a user view must be represented in the system. The need for a user view may seem to limit the generalizability of the framework as some domains do not have such a view represented within the machine. Whether this lack means a true limitation to the framework or signals a shortcoming in human machine ensembles remains open for discussion. However, Klein et al. (2004) argue that maintaining common ground is a key requirement of joint activity of human machine ensembles. Embedding a user world view is one approach to maintain common ground.

The proposed framework therefore presents other researchers and developers with a fine-grained framework that they can use in their own work to get adaptive automation up and running in an operational setting. It enables a granular way of allocating work between humans and machines while speaking in terms that match the operational knowledge of the end-user.

This chapter discussed the implementation of adaptive automation for the C2 identification task requiring an elaboration on *how* adaptive automation is implemented using an object-oriented task model. The question *when* adaptation should take place (i.e., which conditions should trigger the automation to adapt) will be discussed in the next chapter (chapter 3). Chapter four of this dissertation discusses the effects of such an adaptive automation system by comparing a number of measures (e.g. performance, workload, accuracy and timeliness of decisions) when working with adaptive automation to working without such a mechanism using both traditional (symmetric) and asymmetric naval scenarios.

3

TRIGGERING ADAPTIVE AUTOMATION

Abstract – One of the challenging factors in the application of adaptive automation concerns the question of when changes in the level of automation must be effectuated. Literature catalogs five approaches to triggering adaptive automation and the aim of this chapter is to develop triggering models that link closely with the object-oriented task model that are applicable in real-world settings. A short literature review summarizes successful triggering mechanisms. In addition, theoretical mechanisms to trigger global & local adaptation are discussed. Subsequently, the operator performance model and the operator cognition model are proposed. An experiment was conducted to test the validity of the operator cognition model with 18 naïve subjects. Results revealed that the number of objects and associated complexity of those objects serve as a predictor of the subjective workload. The experiment thus supports the validity of the operator cognition model. More importantly, the model can be applied to trigger adaptive automation in real-world settings.

This chapter is based on the following conference papers:

de Greef, T. E., & Arciszewski, H. F. R. (2007). *A Closed-Loop Adaptive System for Command and Control.* Paper presented at the Foundations of Augmented Cognition - Third International Conference, FAC 2007, Held as Part of HCI International 2007, Beijing, China.

de Greef, T. E., & Lafeber, H. (2007). *Utilizing an Eye-Tracker Device for Operator Support.* Paper presented at the 4th Augmented Cognition International (ACI) Conference being held in conjunction with the HFES 51st Annual Meeting, Baltimore, Maryland.

de Greef, T. E., Lafeber, H., van Oostendorp, H., & Lindenberg, J. (2009). *Eye Movement as Indicators of Mental Workload to Trigger Adaptive Automation.* Paper presented at the Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience, 5th International Conference, FAC 2009 Held as Part of HCI International 2009, San Diego, CA, USA.

3.1 Introduction

While chapter 2 of this dissertation discussed how to implement adaptive automation in a real world setting, this chapter focuses on the question when adaptive automation must be triggered. One of the challenging factors in the development of successful adaptive automation concerns the question of when changes in the level of automation must be effectuated. Said differently: what signals and/or responses should be used by the automation to make a proper decision to support the human. Frequently, a reference is made to the concept of 'the workload being too high or too low'. However some acknowledge (cf. Gopher & Donchin, 1986) the fact that it remains difficult to give the concept a concrete form. The purpose of this chapter is to bring into focus which signals and responses are effective triggers for adaptive automation using the object-oriented framework (discussed in chapter 2) and how these measures relate to real world operational settings (naval C2 is taken as an example, discussed in more detail in chapter 2).

The next section provides an overview of the mechanisms discussed in the literature. The following three sections provide a theoretical framework discussing subsequently mechanisms to trigger global and local adaptation, the operator performance model, the operator cognition model, and the combination of both in a hybrid triggering model. Lastly, this chapter discusses an experiment linking workload with the operator cognition model.

3.2 Previous Work

The success of the application of adaptive automation depends in part on the quality of the automation and the support it offers to the human. The other critical part constitutes when changes in the level of automation are effectuated. Workload generally is the key concept to invoke such a change of authority. However, workload is a multidimensional, multifaceted concept that is difficult to define. It is generally agreed that attempts to measure workload objectively relying on a single representative measure are unlikely to be of use (Gopher & Donchin, 1986). The definition of workload as "*… an intervening variable similar to attention that modulates or indexes the tuning between the demands of the environment and the capacity of the operator…*" (Kantowitz, 1988) seems to capture the two main aspects of workload, i.e., the capacity of humans and the task demands made on them. The workload increases when the capacity decreases or the task demands increase. Both capacity and task demands are not fixed entities and both are affected by many factors. Skill and training, for example, are two factors that increase capacity in the long run whereas capacity decreases when humans become fatigued or have to work under extreme working conditions for a prolonged period.

Wilson and Russell (2007) define five strategies based on an earlier division by Parasuraman et al. (Parasuraman, Mouloua, & Molloy, 1996). They state that adaptation triggers can be based on critical events, operator performance, operator psychophysiology, models of operator cognition, and hybrid models that combine the other four techniques. First the occurrence of critical events is the result of automation that monitors the environment for incidents that could endanger the goals of the mission (Scerbo, 1996; Inagaki, 2000a, 2000b; Scerbo, 1996). Inagaki, for example, has published a number of theoretical models (Inagaki, 2000a, 2000b; Moray et al., 2000) where a probabilistic model was used to decide who should have authority in the case of a critical event. The philosophy behind this method is that a critical event requires too much cognitive effort that information processing and decision making is seriously hampered. Second, a decline in operator performance is widely regarded as a potential trigger. Such an approach measures the performance of the human over time and regards the degradation of the performance as an indication of a high workload. Many experimental studies (Figure 3.1) derive a performance indicator from the execution of a secondary task (Clamann, Wright, & Kaber, 2002; Kaber & Riley, 1999; Kaber, Wright, Prinzel, & Clamann, 2005; Kaber, Perry, Segall, McClernon, & Prinzel, 2006; Clamann & Kaber, 2003). Although this approach works well in laboratory settings, the addition of an artificial task to measure performance in a real-world setting is unfeasible to extract performance measures. Third, operator psychophysiology measures human psychological load reactions using physiological measures such as galvanic skin response or heart rate variability (cf. Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006; Byrne & Parasuraman, 1996; Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2000; Veltman & Gaillard, 1998; Wilson & Russell, 2007). The capability of human beings to adapt to variable conditions, however, may distort accurate measurements (Veltman & Jansen, 2004). There are two reasons why physiological measures are difficult to use in isolation. First of all, the human body responds to an increased workload in a reactive way. Physiological measurements therefore provide the system with a delayed cognitive workload state of the operator instead of the desired real-time measure. Second, it is possible that physiological data indicate high workload but that these are not necessarily commensurate with poor performance. This is the case when operators put in extra effort to compensate for increases in task demands. At least several (psycho-physiological or other) measurements are required to get rid of such ambiguities. The fourth approach uses normative models of operator cognition. These models are approximations of human cognitive processes for the purpose of predicting the operator cognitive state. The winCrew tool (Archer & Lockett, 1997), for example, is based on the multiple resource theory (Wickens, 1984) that could serve to trigger adaptive automation. Alternatively, the human's interactions with the machine can be monitored and evaluated against a model to determine when to change levels of automation. In a similar approach, Geddes (Archer & Lockett, 1997) and Rouse, Geddes, and Curry (Wickens, 1984) base adaptive automation on the human's intentions as predicted from patterns of activity. A number of studies (Hilburn, Jorna, Byrne, & Parasuraman, 1997; Hilburn, Molloy, Wong, & Parasuraman, 1993; Kaber & Endsley, 2004; Parasuraman, Mouloua, Molloy, & Hilburn, 1993) used such models to anticipate points in time at which the adaptive automation system should support the human. The fifth approach follows Gopher and Donchin (1986) in that a single method to measure workload is too limited. Hybrid models therefore combine a number of triggering techniques because the combination is more robust against the ambiguities of each single model.

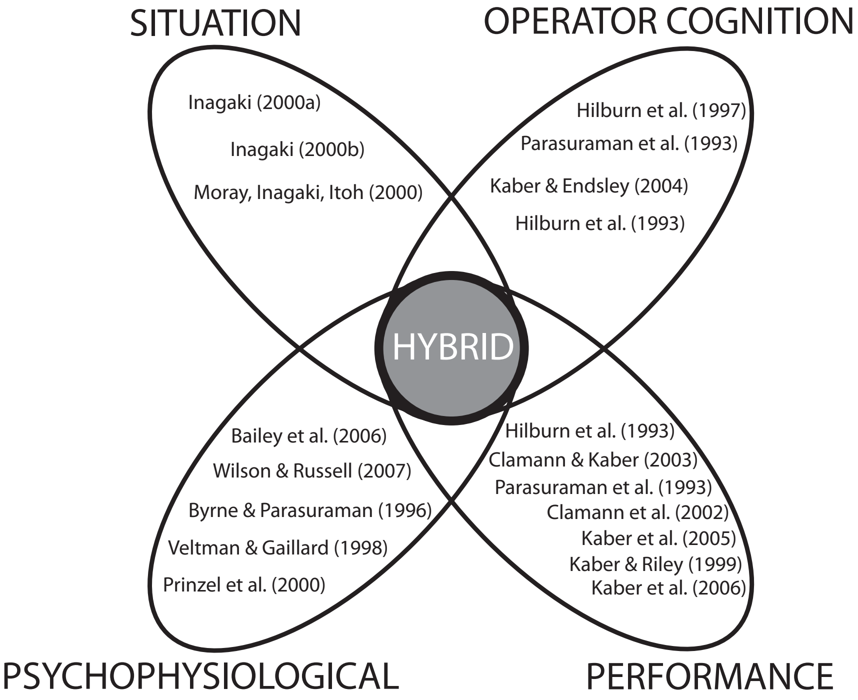


Figure 3.1 – An overview of studies on adaptive automation categorized by triggering strategies

Figure 3.1 shows from a number of adaptive automation studies which trigger method is used. However, the foremost interest of this chapter was a triggering model capable of working outside the laboratory in the real operational world. Future and current military operations are difficult and the environment should be described as highly complex. This chapter proposes that a single triggering method proves too limited to capture the complexity of the environment and the human response correctly. Consequently, this chapter studies which models are usable to trigger adaptive automation in a highly complex environment. A hybrid approach is believed to be more robust because if there are more independent indicators pointing to a high workload, the operator is more likely to be overloaded in fact. A multipronged approach is thought to alleviate such artifacts as a bias in the performance of the primary tasks.

3.3 Global and local adaptation

Two different types of adaptation triggers are identified. The distinction between the two types can be interpreted as that between local and global adaptation (de Greef & Lafeber, 2007). Global aiding is aimed at the relief of the human from a temporary overload or underload situation by taking over parts of the work. On the other hand, if the human misses a specific case that requires immediate attention in order to maintain safety, local aiding comes to the rescue. Figure 3.2 shows both cases in an example. The left shows a naval tactical situation prior to any form of adaptation. The middle picture shows the situation after local aiding when the operator has missed a contact in the east flying directly towards the navy frigate while crossing a safety range. The automation identifies this track as hostile (black). The right figure shows the situation after global aiding has been triggered. The automation aids the operator globally by identifying six tracks as neutral using predefined working agreements. In both cases work is shifted from the human to the machine, but during global aiding this is done in order to avoid the overwhelming of the human (i.e. overload), whereas local aiding offers support in those individual local cases the human misses things possibly due to for example cognitive tunneling. A human is in principal not overloaded in cases where local adaptation is necessary. Particular instances may be missed or decisions are delayed with potentially far-reaching consequences. A further distinction is that local aiding concerns itself with a specific task or object whereas global aiding takes away from the operator that work that is least detrimental to his or her situational awareness. According to this line of reasoning a local case ought to be an exception and the resulting actions can be regarded as a safety net. The safety net can be realized in the form of separate processes that check safety criteria. In an ideal world, global adaptation would ensure that local adaptation is never necessary because the human always has enough cognitive resources to handle problems. But things are not always detected in time and humans are sometimes distracted or locked up so that safety nets remain necessary.

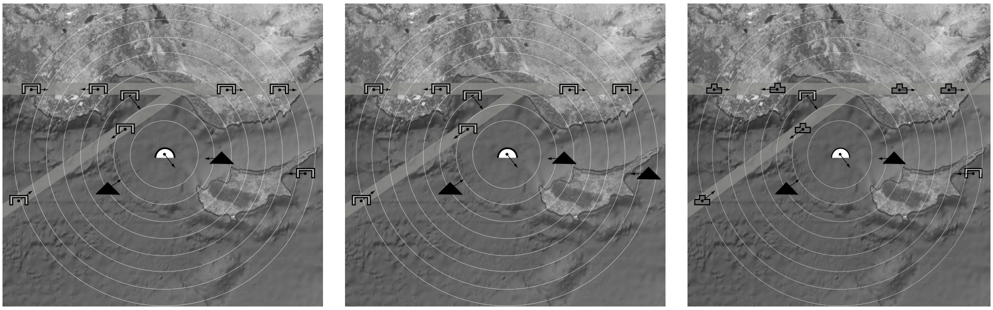


Figure 3.2 – Left: shows a navy environmental representation prior to triggering any form of adaptive automation. The human has identified two tracks a hostile (black triangles) and its own ship in the center as friendly (white). Middle: the operator has 'missed' a track that required identification and adaptive automation was triggered locally and has identified this track in the east as hostile (black triangle). Right: Global adaptation has taken place and using predefined working agreements (see chapter 2) the automation identified six tracks as neutral.

3.4 Triggering local adaptation

Local aiding is characterized by a minimum time left for action required to avoid endangering mission goals. Activation of such processing is through triggers that are similar to the critical events defined by Scerbo (1996). These local triggers are indicators of the fact that certain predefined events endanger mission goals and that action is required shortly. In the case of naval C2 a critical event is usually due to a predefined moment in the state of an external entity. The time left as a way to initiate a local aiding trigger can usually be translated to range from the ship or unit to be protected. In most cases therefore triggers can be derived from the crossing of some critical boundary. Examples are tracks that are not yet identified at a critical range called the identification safety range. Similarly, the successful and safe usage of weapons is limited by a number of critical ranges as well. It is especially the minimum range, at which the weapon is no longer usable, that can be earmarked as a critical one. Within the air traffic domain, the crossing of a minimum separation distance between two airplanes could serve as a critical event.

Typically, local aiding occurs in situations where either the human misses something due to a distraction by another non-related event or entity or to the fact that the entity has so far been unobserved or been judged to be inconsequential (e.g., tunnel vision).

3.5 Triggering global aiding

One of the advantages of the object-oriented framework outlined in chapter 2 is that it offers two very useful hooks for the global adaptation approach. The first hook is the difference between human world-view and machine world-view and is elaborated upon in the operator performance model. The second hook is based on the character of the objects present and is utilized for estimating the workload imposed on the human by the environment (the operator cognition model).

3.5.1 The Operator Performance Model

Performance is usually defined in terms of the success of some action, task, or operation. Although many experimental studies are able to define performance, real world settings are more ambiguous and lack an objective view of the situation (the 'ground truth') that could define whether an action, task, or operator is successful or not. Defining performance in terms of reaction times is another popular means although some studies found limited value in utilizing performance measures as a single way to trigger adaptive automation (cf. de Greef & Arciszewski, 2007).

As explained in chapter 2, the object-oriented framework includes the principle of separate workspaces for man and machine. This entails that both the machine and the human construct their view of the world and store it in the system. For every object (i.e., track) a comparison between the two world-views can then be made and significant differences can be brought to the attention of the human. This usually means that new information has become available that requires a reassessment of the situation as there is a significant chance that the human's world-view has grown stale and his or her expectations may no longer be valid. These significant differences are used in two ways to model performance. First, an increase in the number of differences between the human world-view and the machine world is seen as an indication of a decline in performance. Although differences will inevitably occur, as the human and the machine do not necessarily agree, an increasing skew between the two views is an indication that the human has problems with his or her workload. Previous work suggested that the subjective workload fluctuated in proportion to the density of signals resulting from skew differences (van Delft & Arciszewski, 2004). Therefore, the second performance indicator is the average reaction time.

Utilizing either skew or reaction times as the only trigger mechanism is problematic because of the sparseness of data due to the small number of significant events per time unit in combination with a wide spread of reaction times (de Greef & Arciszewski, 2007). The combined use of skew and reaction times provides more evidence in terms of human cognitive workload. This in turn is enhanced by the operator cognitive model discussed below.

3.5.2 The Operator Cognition Model

While the operator performance model is aimed to get a better understanding of the performance in terms of the human response to the situation, the operator cognition model aims at estimating the cognitive task load the environment exerts on the human operator. A more complex or busy environment logically requires more (cognitive) effort and attention in comparison with an easy and quiet situation.

The operator cognition model is based on Neerincx's (2003) Cognitive Task Load (CTL) model and is comprised of three factors that have a substantial effect on the cognitive task load. The first factor, percentage time occupied, has been used to assess workload for time-line assessments. Such assessments are based on the notion that people should not be occupied more than a certain amount of the total time available. The second load factor is the level of information processing. To address cognitive task demands, the cognitive load model incorporates the skill-rule-knowledge framework of Rasmussen (de Greef & Arciszewski, 2007) where the knowledge-based component involves the highest workload. To address the demands of attention shifts, the model distinguishes task-set switching as a third load factor. It represents the fact that a human operator requires time and effort to reorient himself to a different context.

These three factors present a three-dimensional space in which all human activities can be projected as a combined factor. Specific regions indicate the cognitive demands activities impose on a human operator. Figure 3.3 shows the three dimensional space where the middel area is the optimal bandwidth in which the operator is most comfortable while the top area represents an overload situation and the lower space demonstrates an underload situation.

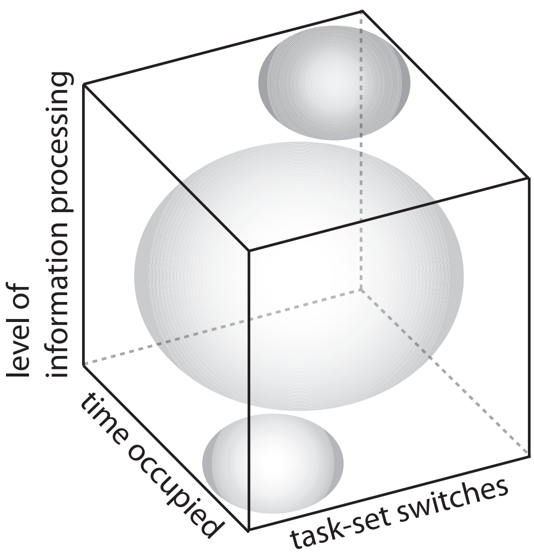


Figure 3.3 – The three dimensions of Neerincx's (2003) cognitive task load model: time occupied, task-set switches, and level of information processing. Within the cognitive task load cube several regions can be distinguished: an area with an optimal workload displayed in the center, an overload area displayed in top vertex, and an underload area displayed in the lower vertex.

Applying Neerincx's CTL model leads to the notion that the cognitive task load is based on the volume of objects requiring information processing (reflecting time occupied), the number of different objects and tasks (task set switching), and the complexity of the objects in the situation (level of information processing). First, as the volume of information processing is likely to be proportional to the number of objects (tracks) present, the time occupied factor will be proportional to the total number of objects. The second CTL factor is the task switching factor. Two different types of task switching are recognized, each having a different effect size. The human operator can change between tasks or between objects (tracks). The first switch relates to the attention shift that occurs as a consequence of switching tasks, for example from the classification task to the engagement task. The second type of switch deals with the required attention shift as a result of switching from object to object within the same task. The latter type of task switch is cognitively less demanding because it is associated with changing between objects in the same task and every object has similar attributes, each requiring similar information-processing capabilities. Finally, a command and control context can be expressed in terms of complexity (i.e., level of information processing). The complexity of an object (i.e., a track in C2) depends mainly on the identity of the track. For example, 'unknown' tracks result in an increase in complexity since the human operator has to put cognitive effort in the process of ascertaining the identity of tracks of which relatively little is known. The cognitive burden will be less for tracks that are friendly or neutral. The unknown, suspect, and hostile tracks require the most cognitive effort for various reasons. The unknown tracks require a lot of attention because little is known about them and the operator will have to ponder them more often. On the other hand, hostile tracks are considerable cognitive more demanding because their intent and inherent danger must be decided. Especially in current-day operations, tracks that are labeled hostile do not necessarily attack and neutralization might only be required in rare cases of clear hostile intent. Suspect tracks are somewhere between hostile and unknown identities, involving too little information to definitely identify them and requiring continuous threat assessment as well. The relation between the theoretical complexity of tracks is summarized in Table 3.1.

Table 3.1 – Complexity in relation to the identity of track

Low Complexity	Neutral, Friendly
Medium Complexity	Unknown
	Suspect
Complexity High	Hostile

In summary, the proposed operator cognition model relates the three variables of Neerincx's CTL model to the object-oriented framework and distinguishes two factors that have an effect on cognitive task load. First, the volume of objects has an effect on both the time occupied and the task set switches. The latter due to different contextualization of each object. Secondly, the identity of each object defines the complexity of objects (Table 3.1).

While the cognitive task load model stems from psychological theoretical validated theories, the proposed operator cognition model requires validation to prove a relation between the experienced workload and the characteristics of the object-oriented task model. Therefore, the next section describes an experiment that validates the relation between the operator cognition model and experience subjective workload.

3.6 The Operator Cognition Model Validated

In order to see whether the proposed model of operator cognition is a descriptor for cognitive workload the relation between the object-oriented framework and cognitive task load was studied in an experiment. More specifically, this experiment attempted to answer the question whether the factors in the operator cognition model properly predict cognitive workload. According to the proposed theoretical framework manipulation of the volume and complexity of tracks has an effect on the cognitive task load factors. Consequently, the hypotheses are that:
(1) the number of tracks (i.e., objects) have an effect on the subjective workload, and that (2) the complexity of tracks have an effect on subjective workload.

3.6.1 Apparatus & Procedure

The subjects were given the role of human operators of (an abstracted version of) a combat management workstation aboard naval vessels. The workstation comprised a schematic visual overview of the nearby area of the ship on a computer display, constructed from the data of radar systems. On the workstation the participants could manage all the actions required to achieve mission goals. Before the experiment, the participants were given a clear description of the various tasks to be executed during the scenarios. Before every scenario, a description about the position of the naval ship and its mission was provided. The experiment was conducted in a closed room where the participants were not disturbed during the task. During the experiment, an experimental leader was situated roughly two meters behind the participant to assist when participants had questions or ran into problems.

3.6.2 Participants

Eighteen subjects participated in the experiment and were compensated financially for their participation. The test subjects were all university students, with a good knowledge of English. The participant group consisted of ten men and eight women. They had an average age of 25 (SD = 5.1).

3.6.3 Experimental tasks

The goal of the human operator during the scenarios was to monitor, classify, and identify every track (i.e. airplanes and vessels) within a 38 nautical miles range around the ship. Furthermore, in case one of these tracks showed hostile intent (in this simplified case a dive toward the ship), they had the mandate to protect the naval vessel and eliminate the track.

To achieve these goals, the participant was required to perform three tasks. First, the classification task gained knowledge of the type of the track and its properties using information from radar and communication with the track, air controller, and the coastguard. The participant could communicate with these entities using chat functionality within the software. The experimental leader responded to such communications. The second task was the identification process that labeled a track as friendly, neutral, or hostile. The behavioral characteristics of tracks were manipulated to differentiate between identities. The last task involved weapon engagement in case of hostile intent as derived from certain behavior. To use the weapons, the participant was required to follow a specific procedure that was elaborated upon at the start of the experiment.

3.6.4 Scenarios

There were three different scenarios, each implying a different cognitive task load. The task load was manipulated, according to the proposed operator cognition model, by manipulating the number of tracks and the complexity of the tracks. The task loads conditions were under-load, normal load, and an overload achieved by manipulating two factors and Table 3.2 provides an overview per scenario.

Table 3.2 – Total number of tracks and the number of tracks with hostile behavior per scenario.

	Total number of track within 38 nautical miles	Track with hostile behavior
Under-load scenario	9	1
Normal workload scenario	19	7
Overload scenario	34	16

3.6.5 Experimental Design

A within-subjects experimental design was applied where each participant performed all tasks for each scenario (independent variable). The presentation of scenarios was counterbalanced over participants using a Latin square design to compensate for possible learning effects.

In order to verify whether the manipulated items affected mental workload, the subjects were asked to indicate their workload. Every 100 seconds subjects had to rate his or her perceived workload on a Likert scale (one to five). Level 1 indicated low workload, level 3 normal workload, and level 5 high workload. The levels in between indicate intermediate levels of workload.

3.6.6 Results

A repeated-measures ANOVA revealed a significant effect in subjective workload between the three scenario's ($F(2,32) = 190.63$, $p < .001$, Figure 3.4). Least square difference post-hoc analysis reveals that all three means were significantly different ($p < .05$). Compared to the under-load scenario ($M = 1.28$, $SD = 0.36$), the perceived mental workload was significantly higher in the normal workload scenario ($M = 2.87$, $SD = 0.60$). In turn, the perceived mental workload in the overload scenario ($M = 4.14$, $SD = 0.56$) was significantly higher again than in the normal-workload scenario.

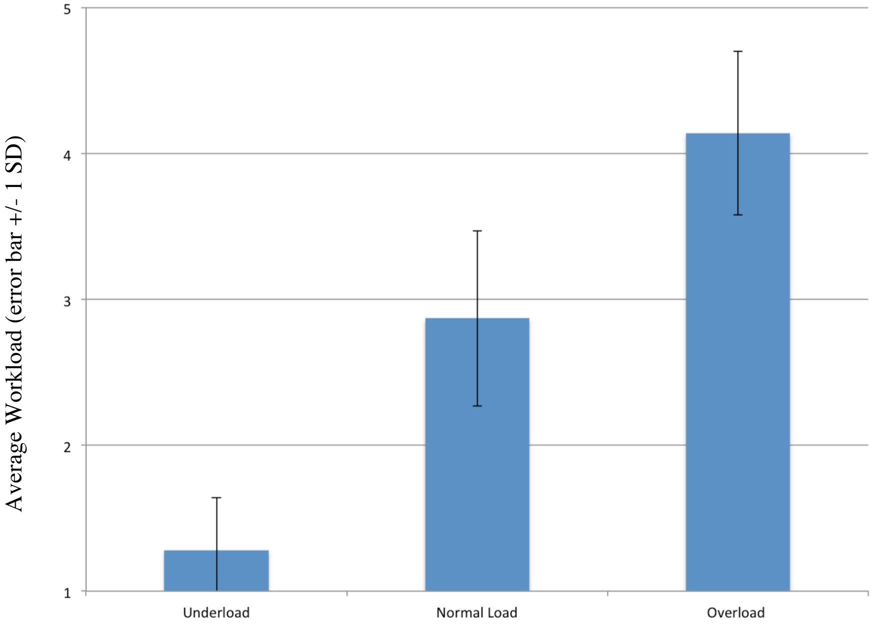


Figure 3.4 – The subjective workload per scenario as indicated every 100 seconds on a five point Likert scale. Note: for the mental workload verification, N = 17 as the data of one subject was missing due to a failure in logging.

3.7 Experimental Conclusion

The data from the experiment reveal that manipulation of the CTL factors using numbers and types of domain objects had a significant effect on the subjective workload. It is therefore concluded that the total number of tracks and the number of tracks with extraordinary behavior is an indicator of the difficulty the environment poses on a human operator. The data therefore supports the proposed model of operator cognition.

3.8 Discussion

Adaptive automation is a concept where autonomy of tasks is being shifted between the human and the machine based on a machine's estimation of the operator's workload. Such changes in authority are beneficial to performance because it navigates the human out of less optimal cognitive states (i.e., overload, underload, vigilance) and keeps the human in a bandwidth of optimal workload. Cognitive overload situations manifest whenever a competition for the users' attention is going on between numerous different information items. This overload originates in the limitations of human attention and constitutes a well-known bottleneck in human information processing. On the other hand, prolonged periods of low activity lead to performance degradations because the operator gets out of the information processing loop as he or she becomes a passive monitor (Endsley & Kiris, 1995).

Requiring the machine to estimate the workload of the operator is a challenging prerequisite and therefore this chapter provided insight to the question *when* adaptation should take place when the paradigm of adaptive automation is applied in a operational setting. More specifically, this chapter proposes a hybrid triggering model yielding a robust solution because triggering only takes place when there are more independent indicators pointing to a high workload. A multi-pronged approach is thought to alleviate such artifacts as a bias in the performance of tasks.

The hybrid trigger mechanism models both the performance (i.e., operator performance model) and the workload imposed on the human by the environment (i.e., operator cognition model). Both the operator performance model and the model of operator cognition show potential to be used as triggering mechanisms for adaptive automation. The operator performance model describes a relation between performance and 1) average response time and 2) skew between the human view and the machine view of the situation. On the other hand, the operator cognition model describes a validated relation between the environment and the cognitive task. It is expected that the combination of the operator performance model with the operator cognition model is more robust because the two models allow to correlate performance fluctuations with the cognitive demands posed on the operator by the environmental situation. This needs to be validated in a future experiment however.

This chapter provided insight to the question *when* adaptation should take place when the paradigm of adaptive automation is applied in a real world operational setting. Providing an answer to the when question is not enough to successfully implement the concept of adaptive automation and therefore a framework was proposed in chapter two of this dissertation. The next chapter (chapter 4) discusses the effects of an adaptive automation system by comparing a number of measures (e.g. performance, workload, accuracy and timeliness of decisions) when working with adaptive automation to working without such a mechanism using both traditional (symmetric) and asymmetric naval scenarios.

4

EVALUATION OF ADAPTIVE AUTOMATION USING AN OBJECT-ORIENTED TASK MODEL IN A REALISTIC C2 ENVIRONMENT

Abstract – Manning reduction initiatives and more complicated military operations lead to a higher cognitive workload in command & control (C2) environments. Extending automation with adaptive capabilities can aid the human to overcome cognitive workload challenges. At present, most adaptive-automation research has focused on laboratory experiments and only limited research has aimed to implement and validate adaptive automation in a real-world setting. The objective of the present study was to investigate the effects of adaptive automation in precisely such a setting, extending the scientific knowledge base of adaptive systems with an evaluation of a real-world adaptive task. Implementing adaptive automation in a real world C2 setting required extending current adaptive automation theories with an object-oriented task model and a hybrid triggering mechanism. The extended model was evaluated with eight naval officers using a high-fidelity C2 environment and showed an overall efficiency effect of 56%. Furthermore, no negative side effects of adaptive automation have been found and the data show that the scenarios were manipulated correctly. In addition, the positive efficiency effects appear most strongly in the more complicated asymmetrical scenarios (65%). This latter conclusion shows that adaptive automation can be a valuable contribution to future C2 systems.

This chapter is based on*:

de Greef, T.E., Arciszewski, H.F.R, and Neerincx, M.A. (2010). Adaptive Automation based on an Object-Oriented Task Model: Implementation and Evaluation in a Realistic C2 Environment. *Journal of Cognitive Engineering and Decision Making*, Vol 4, 152-173.

** The first and second author equally contributed to this research and have agreed to alternate first and second authorship for this journal publication and the one that is referred to in chapter 2.*

4.1 Introduction

Today, military teams are challenged by increasingly complex operations. Large amounts of information, with an ambiguous nature, and legislative constraints tax their abilities (Grootjen, Neerincx, & Weert, 2006). In naval command and control (C2), for example, operations in the littoral, asymmetric threats, and restrictive legislative rules of engagement all contribute to this increased complexity. An asymmetric threat is characterized as a civilian entity having a hostile intention, requiring an increased cognitive effort to properly distinguish it from a non-threat. The task of characterizing entities as being friendly, neutral, unknown, suspect, or hostile is coined identification and is an important task in military C2 as an incorrect identification might lead to an incorrect decision to use weapons. Military teams are constrained in their actions by so-called rules of engagement. These rules severely constrict the use of active sensors types and weapons in determining the intent of an unknown contact or to warn a contact from an unwanted course of action. They thus increase the cognitive effort of the teams, which have to assess the situation with limited means to probe and adjust the situation. In addition, the teams have to continuously weigh the danger to themselves against possible errors in a reaction.

Whereas the environment poses additional challenges, a tendency to operate at lower costs leads to restrictions on the operational teams by manning-reduction initiatives and diminished training. In addition, teams are expected to operate with multiple organizations, requiring them to fold in seamlessly in a variety of international organizations. In short, today’s teams are challenged to do more with less in an ad-hoc ambiguous international setting.

Both the environmental conditions and the pressure to reduce operational costs increase the risk to the human operator of becoming overloaded. This in turn reduces the safety margins by limiting the degree of redundancy between systems within the same organization. An overload case manifests itself as a result of having to undertake too many complex tasks (cf. Neerincx, 2003), when different ambiguous information items compete for the human’s attention. A decreasing redundancy results in a diminishing flexibility in the organization, leading in turn to a reduced capability to reconfigure itself in order to remain responsive to a rapidly changing and complex environment.

Decreased human effectiveness and diminished organizational flexibility are unwanted and the possibilities to have automation assist the human operator in such situations offer a rewarding object of study. When a human is starting to get overwhelmed by the situation, automation capable of autonomous decision-making could intervene and reallocate part of the work to itself so that the workload of the human operator is lessened and he or she regains the capacity to execute his or her tasks effectively, efficiently, and satisfactorily.

Simply increasing the level of automation, however, has its own problems. Research has indicated that offering high levels of automation suffers from pitfalls such as over-reliance (Parasuraman & Riley, 1997), skill degradation (Billings, 1997), and reduced situation awareness (Endsley & Kiris, 1995). All these pitfalls are related to making the human operator a passive monitor, in essence pushing him or her out of the loop.

The approach to a dynamic division of work between the human and automation is called adaptive automation (Hancock et al., 1985; Parasuraman et al., 1992; Rouse, 1988; Scerbo, 1996) aiding human operators to manage their workload in real-time. Literature suggests that this dynamic behavior represents the best match between task demands on one side and the available cognitive resources of a human on the other hand (Parasuraman et al., 1996; Wickens & Hollands, 2000). A number of studies have shown that adaptive automation can regulate workload, improve performance, and enhance situation awareness (Bailey et al., 2006; Hilburn et al., 1997; Kaber & Endsley, 2004; Kaber et al., 2006; Moray et al., 2000; Prinzel et al., 2003).

The object-oriented framework extends current adaptive automation models by overcoming three shortcoming that surface when bringing the paradigm of adaptive automation to real world settings (discussed in more detail in chapter 2). The first shortcoming is that current adaptive automation implementation are ambiguous to explain how to divide work within a single task. The object-oriented task framework therefore allows a fine grained division of work that links closely with how humans delegate work to a colleague. Secondly, the object-oriented framework allows to allocate parts of the work that are less critical in terms of severity or responsibility. And last, third, the object-oriented framework allows end-users to make a final say as to what levels of automation the system is authorized to reach as the end-user is the domain expert and can weight decisions best.

This chapter focuses on the human-machine interaction effects introduced by the fact that the automation is capable of reassigning work to itself, thereby lowering the workload of the human. Consequently the research interests relate to the effects of such an adaptive automation system on response times to warnings, the number of warnings, the accuracy and timeliness of decisions, the subjective workload, and the number of communicative acts of naval officers performing a real-world task in a high-fidelity simulation environment. Its primary purpose was to evaluate these measures when working with adaptive automation in an operational setting with those without such a mechanism using both traditional (symmetric) and asymmetric naval scenarios. This study discusses also the experimental results in relation to the effects observed in other studies (e.g. Clamann, Wright, & Kaber, 2002; Kaber & Riley, 1999; Kaber et al., 2006; Parasuraman et al., 1996).

Chapter two of this dissertation discusses the implementation of adaptive automation for the C2 identification task requiring an elaboration on *how* adaptive automation is implemented using the object-oriented framework. The subsequent chapter within this thesis discusses the question *when* adaptation should take place (i.e., which conditions should trigger the automation to adapt) including empirical data. However, this study opted for a hybrid triggering scheme and why this is appropriate is discussed in the next section.

4.2 Triggering Models

In addition to the challenge how and to what extend to shift control (see chapter 2), another challenge relates to the question of *when* changes in authority must be effectuated and chapter three of this dissertation elaborates on the different strategies that can be applied to trigger automation on an appropriate moment.

Table 4.1 shows from a number of adaptive automation studies which trigger method were applied. No study explicitly applied a hybrid task model. However, the purpose of this study was to utilize a triggering model capable of working outside the laboratory and in the real world. Future and current military operations are difficult and the environment should be described as highly complex. Chapter three of this dissertation discusses that a single triggering method proves too limited to capture the complexity of the environment and the human response correctly. Following chapter three, a hybrid task model was utilized to trigger adaptive automation because the combination was expected to be more robust: if there are more independent indicators pointing to a high workload, the operator more likely experiences a high workload. It was expected that a multi-pronged approach should alleviate artifacts such as a bias in the performance of the tasks.

For the prototype a hybrid trigger scheme was opted for based on operator performance in combination with a model of operator cognition. First, the operator performance is based on a widening disparity between the automation's interpretation of the world and the human's view, reflected in an increase in the number of signals (per time-unit) as well as the response times to these signals. An increase is assumed to be an indication that the human is struggling with his workload, especially if other indicators prove likewise. The second indication uses Neerincx's cognitive task load model (2003). The total number of tracks provides an indication of the time occupied and task set switches whereas the complexity of objects (expressed in terms of their identity) is an indication of the level of information processing. The most difficult tracks are hostile, suspect, and unknown (the latter two requiring continuous scrutiny as to their real identity), whereas neutral and friendly tracks are mostly processed using rule-based algorithms.

4.3 Testing with Domain Experts

The naturalistic decision-making proponents (Zsambok & Klein, 1999) argue that domain experts in the field and naïve subjects in the laboratory utilize different cognitive mechanisms in order to solve problems at hand. They advocate studying decision-making *in the wild* in order to understand how experts perform complex cognitive tasks in demanding situations.

Most research on adaptive automation, however, has focused on low-fidelity simulations in laboratory experiments using naïve participants (e.g. Clamann et al., 2002; Kaber & Riley, 1999; Kaber et al., 2006; Parasuraman et al., 1996)(Table 1). Only a few studies use domain experts with high-fidelity simulations such as air traffic controllers using a simulation of the Amsterdam airspace (Hilburn et al., 1997). The low-fidelity simulations utilize abstracted real-world tasks such as *Multitask* and the *Multi Attribute Battery (MAT) Task*. Although the low-fidelity experimental studies provide valuable general and fundamental knowledge applicable to a wide arena of users, the complex matters of (asymmetric) warfare require knowledge, expertise, and experience in the domain thereby excluding naïve participants. These facts support validating the principles of adaptive automation with experts and high-fidelity simulations.

4.4 Current Research

Table 4.1 summarizes a number of studies in the field of adaptive automation. The table shows which model is used to share a task between the human and the automation, what type of trigger is used, the participants (professional or naive), and the fidelity of the simulation environment. The table reveals that only one study used professionals with a related high-fidelity simulation environment and no study utilized a hybrid trigger model. Furthermore most studies base their division of levels of automation on a generalizable model. However, none of these models allow for a fine-grained division of work. It seems therefore that only limited research has so far dealt with evaluating adaptive automation: 1) using professionals with real-world tasks in a high-fidelity command & control simulation, 2) with a generalized fine-grained task model to divide the work between the human and the automation, and 3) a hybrid trigger model. The objective of the present study was to investigate the effects of adaptive automation in precisely such a setting, extending the scientific knowledge base of adaptive systems with an evaluation on a real-world adaptive task.

The proposed combination of an object-oriented approach (described in chapter two) and a hybrid triggering model (described in chapter three) is set to avoid some of the potential problems related to adaptive automation (cf. Lee, 2006). Automation surprises (Sarter & Woods, 2000), for example, relate to unanticipated events by the automation. The application of a hybrid trigger mechanism seems more robust to these negative effects because the occurrence of trigger moment is better correlated to workload leading to less non-anticipated triggering of adaptive automation. Furthermore, the automation-human coordination asymmetry (Woods, Tittle, Feil, & Roesler, 2004) is minimized because working agreements define an a-priori explicit contract what and what not to delegate to the automation at times of high workload. Lee & Sanquist (2000) and Grabowski & Sanborn (Grabowski & Sanborn, 2003) worry that poorly designed automation might reverse beneficial effects of automation in non-routine situations. The object-oriented framework, however, allows for a fine grained way to remain in control of the difficult or complex tracks while the automation deals with the easy routine items in non-routine stressful situations.

Hypotheses in three areas are generated. The first area concerns improved efficiency in relation to beneficial effects of earlier adaptive automation studies. These benefits are in the area of performance and workload (Bailey et al., 2006; Hilburn et al., 1997; Kaber & Endsley, 2004; Moray et al., 2000; Prinzel et al., 2003). Consequently, it is expected that adaptive automation results in improved reaction times to difficult tracks (hypothesis 1a), less outstanding warnings requesting operator attention (hypothesis 1b), and a lower experienced workload (hypothesis1c) when compared to working with static levels of automation. Furthermore, the adaptation supports the operator in the first stages of the C2 information processing loop (i.e., identification) and not to the decision-making stage. Given that negative results of the automation aren't anticipated (e.g., automation surprises), no effects on the decision-making are expected resulting in an equal accuracy and timeliness of decisions (hypothesis 1d) in comparison with static levels of automation[1]. Likewise, no difference in communicative acts (hypothesis 1e) are expected.

The second area concerns hypotheses on the *type* of scenario as motivated at the beginning of the introduction. It is expected that scenarios with an asymmetric threat will show a longer reaction time to warnings (hypothesis 2a) but with an equal accuracy and timeliness of decisions (hypothesis 2b) in comparison with traditional symmetric ones.

Finally, it is expected to find an interaction effect between the automation mode and the scenario type given the increased cognitive demands due to asymmetric conditions. It is hypothesized that the reaction time improvements thanks to adaptive automation manifest themselves most strongly in the asymmetric scenarios (hypothesis 3) when compared to traditional scenarios and static levels of automation.

To test these hypotheses, an experiment was conducted in which eight naval experts participated. In addition to executing tasks in the C2 domain, the participants had to navigate the ship. The major C2 tasks were identification and weapon engagement. These tasks are part of a full C2 information-processing loop, particularly information analysis and decision-making. In support of these tasks, the participants were offered a synthetic tactical view of the situation to aid their situational awareness and a chat console to gather and distribute information among team members. In addition a simulated helicopter was flying around in support of the C2 task execution. The participants could communicate with and assign new tasks to the helicopter using the chat console.

In the next section the experiment is described in terms of methods and materials used. Following the description of the experiment, the results will be reported. The results will be elaborated in the discussion. Finally, a concluding section will discuss the results and elaborate on future extensions.

---

[1] Some studies claim beneficial situation awareness effects (Kaber & Endsley, 2004; Kaber et al., 2006) yet this study didn't incorporate situation awareness effects in its

Table 4.1 - An overview of studies on adaptive automation

Study	Trigger	Participants	Environment
(Wilson & Russell, 2007)	Psycho-physiological measures	Semi-Naïve	Monitor UAV's and visual search task
(Bailey et al., 2006)	Psycho-physiological measures or Adaptable (user instigated change of authority)	Naïve	Low Fidelity, MAT Battery task
(Kaber et al., 2006)	Operator Performance (STL)	Naïve	Low Fidelity ATC task (MultiTask)
(Kaber & Endsley, 2004)	Model of operator cognition (time)	Naïve	Low fidelity RM & TE task (MultiTask)
(Clamann & Kaber, 2003)	Secondary performance based	Naïve	Low fidelity ATC task (Multitask)
(Clamann et al., 2002)	Operator Performance (STL)	Naïve	Low Fidelity, ATC task (MultiTask)
(Moray et al., 2000)	Situation Based	Naïve	Experimental Microworld, Low Fidelity
(Kaber & Riley, 1999)	Operator Performance (STL)	Naïve	Low Fidelity RM & TE Task (MultiTask)
(Hilburn et al., 1997)	Model of operator cognition (traffic load)	Experts	High Fidelity ATC Task
(Parasuraman et al., 1996)	Operator Performance (STL)	Naïve	Low Fidelity MAT (Modified) Task
(Hilburn et al., 1993)	Model of operator cognition or Operator Performance	Naïve	MAT task (piloting task)
(Parasuraman et al., 1993)	Model of operator cognition or Operator Performance	Naïve	Tracking, fuel-management, and system monitoring

Note. RM & TE stands for radar management and target elimination, MAT for multi attribute task battery (Suite of flight related tasks including system monitoring, compensatory tracking, and resource management). Semi-naïve subjects of Wilson & Russel have had an average of 10.6 hours of training adding to their experience. STL stands for secondary task load performance measurement.

4.5 Methods and Materials

4.5.1 Participants

A bottleneck of using domain experts in experiments is their scarcity and limited availability. First of all, there may just be a limited number of experts for specific tasks such as air & surface warfare officers in the Royal Netherlands Navy (RNLN). Secondly, due to their scarcity these domain experts usually have a very busy work schedule and therefore have limited time to participate in research activities. These two reasons constrained our group to eight naval officers participating in the evaluation.

The participants were four warfare officers and four assistant warfare officers of the RNLN with several years of operational experience using naval combat systems on naval ships. At the time of the trials the subjects were either involved in the development of new combat management systems for the RNLN or engaged in the training of officers at the operational school of the RNLN.

In addition, one Royal Netherlands Navy anti-air and one anti-surface warfare officer with extensive experience served as a subject matter expert. At each trial one of these two was present and was asked to judge the workload and decision making of the participant every two minutes. They had seen the scenarios beforehand and were familiar with the presented tactical angles and ambiguities of the scenarios.

4.5.2 Apparatus

The participants worked with a workstation called the Basic-T (Figure 4.1) attached to a simulated combat management system. The Basic-T consisted of four 19-inch touch screens arranged in a T-shaped lay-out driven by two PCs (Dell Precision 370 Pentium 4 machines containing a nVidea Quadro FX video card). The Basic-T functioned as a future operational workstation in the command centre of a naval ship and was connected by means of a high-speed data bus to a simulated combat management system running on a simulated naval ship (Figure 4.2). Both the ship and the environment of the ship (terrain, friendly, hostile, and neutral ships and aircrafts) were being simulated in a simulation program called JROADS. Originally designed to evaluate the air defense capabilities of naval ships, JROADS has been largely extended, among other things with functionality to enable high-fidelity human-in-the-loop experiments. The high-speed data bus is based on an in-house implementation of the High-level Architecture and represents the data structures present in state-of-the-art combat management systems (Figure 4.2). The ship closely resembles a modern navy air defense and command frigate in its sensor and weapon suite and additionally incorporates the proposed adaptive automation support.

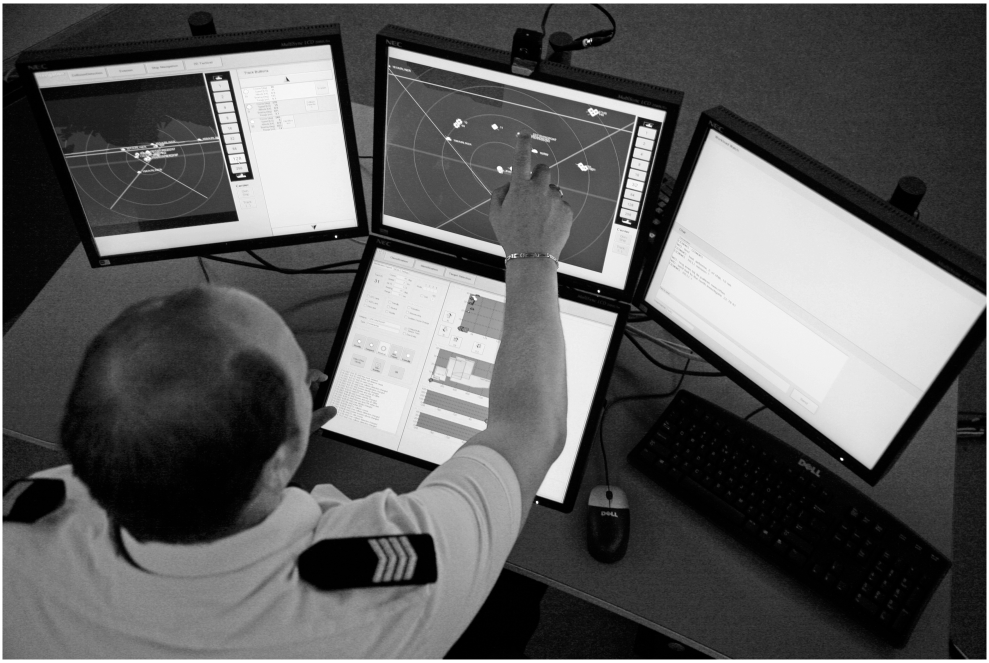


Figure 4.1 – The prototype contains real world identification task using a high-fidelity simulation environment

4.5.3 Tasks

The participants were given mission goals and were instructed to defend the frigate against any threats. In all cases the primary mission goal was to build a complete tactical picture in the operational area of the ship, identifying all platforms present, and to neutralize hostile entities. As the sensor reach of a modern naval ship extends to many miles around the ship, this represented a full-time job. In addition, the participants were responsible for the short-term navigation of the ship, steering it toward whatever course was appropriate under the circumstances and avoiding potential collisions with other ships. The navigation responsibilities were limited for experimental reasons. The participants had a helicopter at their disposal to investigate the surrounding surface area. Although the use of a helicopter greatly extended their surveillance capabilities, it also increased task workload due to increased data volume and the need to assign new tasks to the helicopter.

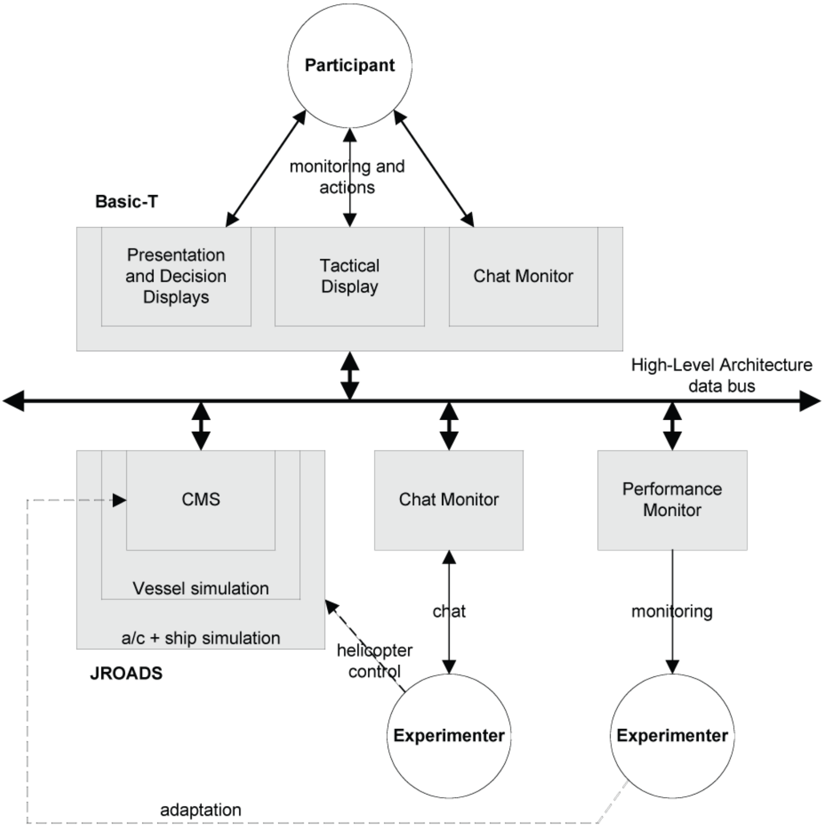


Figure 4.2 – An overview of the setup of the apparatus and the experimental controllers involved.

The static automation was equal to the lowest level in the adaptive mode. The operator could request advice (in this mode the system communicated both arguments for and against a certain identity) and the automation drew attention to tracks where system and user identities differed. In the adaptive mode the automation would acquire more responsibility in handling tracks autonomously when the human's workload increased. At the medium level, neutral tracks were autonomously identified by the system; at the highest level all tracks were identified autonomously but suspect and hostile ones were signaled to the user so that they could be vetoed.

Adaptive automation was applied to the identification task. One experimental leader simulated the adaptive trigger mechanism (Figure 4.2) using the wizard-of-Oz experimental paradigm in which automation functioning is mimicked by a person. The wizard-of-Oz paradigm was chosen to reduce costs and safe time that was otherwise required to implement the triggering mechanism. The experimental leader had a real-time visualization displaying three trigger-elements in order to trigger adaptive automation. First, the number of outstanding warnings provides an indication of the overall information processing load that is asked from the human operator. Secondly, the warning response times provide a performance indicator. Third, the environmental complexity is operationalized by calculating the complexity of all tracks (e.g., a number of suspect tracks is considered more complex than an equal number of neutral tracks). The complexity is represented by the total number of tracks identified by the automation as unknown, suspect, or hostile.

The hybrid approach was triggered to a higher level of autonomy whenever two of the described trigger-elements showed an increase in the same direction allowing to make an unbiased trigger decision. More specifically, the automation would scale up if minimally two of the following criteria were met: 1) more than six outstanding warnings, 2) an average response time to outstanding warnings exceeding 60 seconds, and 3) more than 15 complex tracks as identified by the automation. In other words, the automation scales up whenever, for example, the warning response time increases in combination with an increase in environmental complexity. However, the automation scales down using only one ‘negative’ trigger-element (in a situation where the others remain equal). This means that 1) less than two outstanding warnings, 2) an average response time to outstanding warnings lower than 15 seconds, or 3) a complexity less than five tracks each yielded a decrease in automation.

Each of these three trigger-elements were clearly visible in a separate display that also contained controls to increase or decrease the automation. The display allowed the experimental leader to observe each element fluctuate in time allowing the experimental leader to make an unprejudiced decision to scale up or down in relation to each threshold criteria. Anytime the mechanism adapted, a message was put on the chat communicating what level the automation was going to work. After-action analysis of the results showed that the experimental leader followed the triggering criteria correctly as each increase or decrease of automation was based on these criteria and no other triggering opportunities were identified.

Another experimenter worked the simulation ‘kitchen’, controlling the helicopter, executing commands to on-board personnel (using the fire control radar, gunnery, communications, etc.) and injecting and responding to chat (Figure 4.2) according to navy communication protocols.

4.5.4 Procedure

The participants participated in the experiment for two days. The first day was divided into two parts. In the first part of the day the participants were informed about the general goals of the experiment and the theoretical background of the research. The second part of the first day was used to familiarize the participant with the workstation, the various tasks, and the adaptive behavior. This stage consisted of an overall demonstration of the system and three training scenarios. In the first scenario the experimental leaders guided each participants through the system. The remaining scenarios were executed sequentially by the participant. The offered scenarios showed an increasing complexity and the last scenario more or less approached the complexity of the evaluation scenarios.

The evaluation took place on the second day. Prior to the experimental trials each participants was offered a scenario to refresh their memory on the ins and outs of the workstation. After this warm-up period, the trials commenced. After each run a debriefing session with the participant was held in order to discuss experiences.

4.5.5 Scenarios

Four scenarios were developed in cooperation with training experts of the Royal Netherlands Navy. All scenarios included various threats or suspicious-looking tracks that contributed to the workload. Two of the four scenarios were developed around more or less traditional air and surface warfare in a high-tension peace-enforcing situation while the other two scenarios were situated against a civilian smuggling background. The latter two scenarios were made more challenging by the introduction of more ambiguous tracks (i.e. an asymmetric threat) and by providing the participant with the information of a possible terrorist attack. All scenarios took about 20 minutes to conclude. Because of the relative freedom of the participants to operate their ship, differences in the actual runs of the scenarios occurred. This freedom of action, however, did not result in significantly different outcomes.

4.5.6 Design

The experiment was a two (adaptation mode: adaptive or static) by two (scenario type: traditional or smuggle) repeated measurement design where the independent variables were varied within subjects. Participants were offered four scenarios in total, equally divided over the traditional and the smuggling scenarios. Furthermore, the trials had two scenarios with the adaptive automation and two scenarios in the static automation mode. The order of scenarios and adaptation modes was counterbalanced across participants. The interfaces between the conditions were equal thereby minimizing interface effects between conditions.

4.5.7 Dependent variables

Five dependent variables were measured. First the *human performance* was measured in terms of overall track identification time and a subset of these identification times for the suspect tracks. Secondly the *human-automation performance* was measured by logging the number of warnings on tracks requiring human attention, and the number of pending tracks. Warnings are defined as tracks that are brought to the attention of the human because there is a problem with their identity from the automation's point of view. A new track receives a temporary identity called 'pending' as long as the human has not paid attention to the track. When the automation adapts, some of these tracks are identified by it and disappear as warnings and pending tracks. Consequently, these two variables define the performance of the human-automation combination. Third, the participant *subjective workload* ratings were recorded during each scenario on a one-dimensional rating scale from one to five, one meaning heavy underload and boredom, three a comfortable and sustainable workload, and five an overload of the human. Fourth, *naval expert ratings* estimated the participant's workload, the correctness of actions taken, and the timeliness of actions taken on a handheld device every two minutes. The expert was unaware of the condition and utilized his or her experience and knowledge to judge the participant's workload. Furthermore, the expert also rated the decision making process in terms of correct and timely decisions with regard to the engagement of weapons, the navigation of the frigate, the routing of the helicopter, and the radio warnings issued to suspect/hostile tracks. The expert used the same one-dimensional rating scale as the participants for workload and a similar scale for the other variables (one meaning 'bad', five 'excellent'). In addition, the observer workload estimates served as a cross-check on the participants workload ratings. Fifth, the *communication* was logged and analyzed in terms of number of messages and type of speech act (cf. Bowers, Jentsch, Salas, & Braun, 1998; Kanki, Folk, & Irwin, 1991). Any chat messages communicating a change of authority was filtered out in order to honestly compare communication between the adaptive and the static conditions. Each message was categorized as one of six types of speech act: a command, a query, an information-sharing act, an acknowledgement of a command, a response to a query, and other.

4.6 Results

The measured (dependent) variables were contrasted to the independent variables using a repeated measurement ANOVA with an alpha level of .05 to determine statistical significance. Due to problems with the logging system, 6.25% of the variable identification times and 9.38% the identification times for the suspect tracks were missing and therefore replaced by the mean of all values in the corresponding variable.

4.6.1 Adaptation

All scenarios had three periods requiring intensified cognitive effort, potentially providing a window of opportunity to trigger adaptive automation. The first occurred when the frigate went from silent (i.e., no usage of sensors) to active use of its sensors. At times a navy frigate sails silently to minimize radar emission potentially revealing its presence and position. This switch generated a large flow of information concerning new tracks. Secondly, the human operator utilized the helicopter to inspect suspicious tracks. At the start of each scenario the helicopter was close to the frigate and was flying outward not using its radar system. After five minutes, the helicopter switched on its radar thereby increasing the information flow. Third, two developments on the surface and in the air overlapped near the end of the scenario requiring increased and divided cognitive attention.

Adaptive automation kicked in on average 2.5 times per scenario and the system lowered its assistance 1.3 times per scenario. In 25% of the trials, the automation increased its assistance to the maximum level. Each increase or decrease of automation was based on the rules as expressed in the task subsection and no other triggering opportunities were identified when the log files were re-analyzed.

4.6.2 Human Performance

The human performance in terms of overall track identification time was significantly better ($F(1,7) = 26.93$, $p < .01$) in the adaptive condition (M = 89 sec, SD = 47) in comparison with static automation (M = 202 sec, SD = 104). The overall track identification time was significantly higher ($F(1,7) = 8.13$, $p < .05$) in the more complicated smuggle scenarios (M = 176 sec, SD = 125) in relation to the traditional scenario (M = 114 sec, SD = 47). Furthermore, an interaction effect was found (Figure 4.3 left) between the adaptive mode and the scenario type ($F(1,7) = 29.99$, $p <$ .01). Tukey's post-hoc analysis showed that the identification times for both scenario types differed significantly ($p < .05$). The left side of Figure 4.3 shows clearly that the difference in identification time between automation modes manifested most strongly in the -more complex- smuggling scenarios (from $M = 261$ to $M = 92$ seconds, $SD = 117$ and $SD = 59$ respectively) in comparison with the traditional scenarios (from $M = 142$ to $M = 86$ seconds, $SD = 41$ and $SD = 36$ respectively).

A very specific subset of these tracks is the set of tracks recognized by the automation as threats: the most important and cognitive demanding set as these might put the mission or the ship at risk. Threats encompass suspect and hostile tracks. Not every participant identified adversary tracks as 'hostile' leading to too little data to analyze statistically. However, 'suspect' identification times were significantly lower ($F(1,7)=9.78$, $p < .05$) in the adaptive condition ($M = 137$ sec, $SD = 96$) compared to the static condition ($M = 239$ sec, $SD = 157$). The data also showed an interaction effect ($F(1,7)=10.02$, $p < .05$) with respect to scenario type and automation mode (Figure 4.3 right). Post-hoc analysis using Tukey's HSD test showed that the identification time in the smuggle scenario using the adaptive mode was significantly lower to all other combinations ($p < .05$). The effect of adaptive automation manifested only in the smuggle scenario ($M_{STATIC} = 320$, $SD_{STATIC} = 180$; $M_{ADAPTIVE} = 129$, $SD_{ADAPTIVE} = 102$) in comparison with the traditional scenario ($M_{STATIC} = 157$, $SD_{STATIC} = 71$; $M_{ADAPTIVE} = 145$, $SD_{STATIC} = 95$).

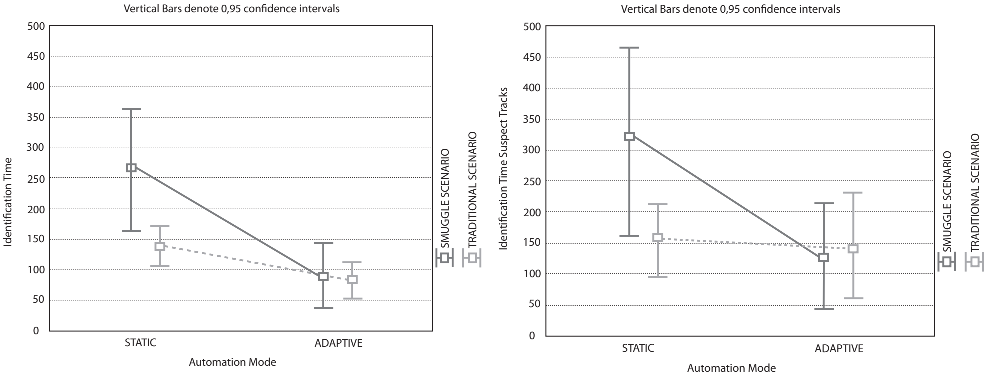


Figure 4.3 – The left graph displays the interaction effect of the identification times and the right graph shows the interaction effect of the identification times of the suspect tracks.

4.6.3 Human-Automation performance

The human-automation performance revealed effects for both the number of warnings requiring human attention and the number of pending tracks. The number of warning per time unit was significantly ($F(1,7)=21.97$, $p < .01$) larger with the static

(M = 10.21, SD = 5.89) than with the adaptive automation (M = 3.36, SD = 1.92). The pending tracks showed a significant main effect ($F(1, 7)=20.25$, $p < .01$) in that the number of pending tracks was much larger with the static (M = 8.49, SD = 5.58) when compared to the adaptive mode (M = 1.43, SD = 0.79).

4.6.4 Subjective workload

The participants were asked to rate their perceived workload on a regular basis. These data revealed no effects on the independent variables automation mode ($F(1,7)=0.003$, $p = .96$), scenario type ($F1,7)=3.33$, $p = .11$), or the interaction between these variables ($F(1,7)=0.13$, $p = .73$). The subjective workload was on average a 3.3 on a five-point scale.

4.6.5 Expert ratings

Analysis of the expert ratings of workload, accuracy of taken decisions, and timeliness of taken decision revealed no effect of the independent variable automation mode (averages respectively $M_{WORKLOAD} = 3.3$, $M_{ACCURACY} = 3.7$, $M_{TIMELINESS} = 3.8$ on a scale from 1 to 5). The other variables showed no significant results between the scenario types nor significant interaction effects. The subjective and the expert workload ratings were correlated ($r(32)=.516$, $p < .01$).

4.6.6 Communication

The communication was analyzed using the absolute number of chat messages and the type of communicative act. The number of messages showed no differences in the independent variable automation level ($M_{STATIC} = 21.19$, $SD_{STATIC} = 7.40$; $M_{ADAPTIVE} = 22.38$, $SD_{ADAPTIVE} = 4.92$), scenario type ($M_{SMUGGLE} = 24.13$, $SD_{SMUGGLE} = 5.99$; $M_{TRADITIONAL} = 19.44$, $SD_{TRADITIONAL} = 5.67$), or the interaction between these two variables. On average 21.78 messages were sent. Furthermore, the data revealed no significant difference in the independent variable automation level on communicative acts (i.e., number of commands $M_{STATIC} = 11.89$, $SD_{STATIC} = 3.38$; $M_{ADAPTIVE} = 12.94$, $SD_{ADAPTIVE} = 3.43$; posed queries $M_{STATIC} = 2.88$, $SD_{STATIC} = 2.99$; $M_{ADAPTIVE} = 2.69$, $SD_{ADAPTIVE} = 1.78$; information sharing acts $M_{STATIC} = 1.19$, $SD_{STATIC} = 1.87$; $M_{ADAPTIVE} = 2.00$, $SD_{ADAPTIVE} = 2.28$; acknowledgements $M_{STATIC} = 3.38$, $SD_{STATIC} = 2.96$; $M_{ADAPTIVE} = 2.94$, $SD_{ADAPTIVE} = 2.41$; responsive acts $M_{STATIC} = 0.50$, $SD_{STATIC} = 0.82$; $M_{ADAPTIVE} = 0.13$, $SD_{ADAPTIVE} = 0.34$; or others $M_{STATIC} = 1.44$, $SD_{STATIC} = 1.90$; $M_{ADAPTIVE} = 1.69$, $SD_{ADAPTIVE} = 1.82$).

The data did reveal a significant difference in the communication data in relation to scenario type. A significant effect was revealed in the number of posed queries in the scenario type ($F(1,7) = 6.02$, $p < .05$). It appeared that the more ambiguous smuggle scenarios led to more posed queries ($M = 3.88$, $SD = 2.75$) when compared to the traditional scenarios ($M = 1.69$, $SD = 1.41$).

4.7 Discussion

We mocked-up an implementation of adaptive automation using a realistic task from a C2 setting and evaluated it using a high-fidelity simulator with the extensive aid of RNLN officers. The purpose of this study was to evaluate response times to warnings, the number of warnings, the accuracy and timeliness of decisions, subjective workload and communicative acts when working with adaptive automation in an operational setting and compare them with trials without such a mechanism in both traditional and asymmetrical scenarios. This section discusses the results.

4.7.1 Hypotheses

The human-performance results demonstrate improved identification times of tracks when the automation adaptively aids the human, confirming hypothesis 1a. This is the case both for the reaction times to all tracks and for the subset of suspect tracks. The results show an improvement of 42% (101 seconds) when dealing with suspect tracks and a general identification improvement of 56% (113 seconds).

The results of the human-automation performance confirm hypothesis 1b in showing both a lower number of warnings requiring operator attention and a lower number of pending tracks requiring human attention in the adaptive condition. This shows that the automation does take over work that would normally require attention and that this reduction results in a lower demand on attentional resources. The participants indeed reported that the tactical display was more “quiet” in the adaptive mode and that it enabled them to focus more on the important issues. The adaptive mode took away some of the ‘less important’ work in the case of heavy loads, thereby resolving cognitive limitations.

The results did not reveal an effect of the automation mode on the subjective workload and consequently did not support hypothesis 1c. The increase in performance was not reflected in a lowered workload as experienced by the participants or as observed by the expert observers (the correlation between these two estimates was ‘large’ according to Cohen (1992). Failing to find a subjective workload effect but finding a significant performance effect is explained by a restricted focus of attention of the participants on the more important objects, in combination with an acceptance of a larger risk due to the diminished attention to the other objects. We believe that the experts maintained their subjective workload at a 'comfortable' level (i.e., around level three on a five-point scale) while accepting an increased risk due to not finishing or delaying tasks. This explanation matches the adaptive-operator effects described by Sperandio (1971) and later by Veltman & Jansen (2004). Sperandio showed that air traffic controllers alter task strategies and reduced communication to keep the workload at a constant manageable level. Veltman & Jansen argue that experienced human operators can utilize two strategies to cope with increased demands. The first strategy involves investing more mental effort and the second involves reducing the task goals. For this second strategy Veltman & Jansen state that "*...operators will slow down the task execution, will skip less relevant tasks, or accept good instead of perfect performance*" (p 9). In our case the participants seemed to accept the larger risk of delaying identifying contacts by limiting their attention to a smaller area around the ship in order to maintain their mental effort at a reasonable level.

No differences were found in expert ratings on the accuracy and timeliness of decisions taken (in support of hypothesis 1d) and the data did not reveal any differences in communication between automation modes (in support of hypothesis 1e). The consistency in decision making led to the conclusion that the type and timing of the aid did not cause problems. In some cases participants stated that the automation "seemed to read their mind" when taking over some of the work, while other participants felt there was no clear correlation between workload and time of adaptation. Although more experimental studies are clearly required, these remarks demonstrate the value of this study in relation to getting a grip on the workload of the human operator and reassigning work in order to optimize overall performance.

Hypothesis 2a stated a longer reaction time to warnings in the more complicated smuggle scenarios. The general reaction time to warnings showed an increase of 65% when participants worked in the smuggle scenarios, thereby validating hypothesis 2a. The number of questions posed by the participant was 2.3 times higher in the smuggle scenarios, contributing to the hypothesis that the situations were more difficult to comprehend. No support was found for hypothesis 2b as the results show no difference in the expert report of accuracy or timeliness of decisions taken.

The interaction effect between the scenario type and the automation mode for the general track identification times and the suspect identification times demonstrates the effect of adaptive automation to be the strongest in the smuggle scenarios, validating hypothesis 3. This means that the paradigm of adaptive automation can prove valuable in future military missions as it is expected that these will become more complex (asymmetrical).

In summary, the evaluation did show a positive efficiency effect of adaptive automation while possible negative effects of adaptive automation did not surfaced. However, using naval domain experts resulted in a low number of participants giving low power to these results. The performance effects of adaptive automation are large both for identification times in general (60%) and for the difficult suspect tracks (42%). Both are a good indication of the beneficial effects of adaptive automation in real-world domains. Although no negative effects have surfaced in this experiment, more (naturalistic) evaluations are required to positively determine whether these negative effects indeed do not occur.

4.7.2 Transparency of the Automation

Two potential hazards of adaptive automation in terms of increased cognitive workload are: (1) unexpected automation behavior confusing the human about the automation's operation and (2) mode awareness causing humans to miss a change in automation level resulting in conflicts with the automation. In an early paper, Billings & Woods (1994) condemn adaptive automation because the potential beneficial effect would be balanced by unpredictable behavior leading to confusion. Although these concerns are valid, no evidence for these negative effects was found. Although more evaluations are clearly required to draw strong conclusions about the absence of these negative effects, none of the participants made remarks concerning mode awareness. It is the assumption that this is probably because changes in identity (the result of the task partially delegated to the automation) were clearly visible on the tactical displays as both the color and the shape of updated tracks changed.

4.7.3 Applied Methods

The observer ratings of workload and the subjective workload assessment are subject to discussion (i.e., sensitivity). Workload estimates by both the participant and the expert were measured because a crosscheck on both estimates was desired. The correlation between the two was 'large' according to Cohen (1992), fortifying the results. The observer estimate of the decision making process provides an easy way to get a grip on this process. The method used is not very accurate but decision-making was not the focus of this study and this study merely wanted to make sure that the adaptive automation was not influencing other stages of the military command & control information processing loop. Future studies might consider alternative methods, however. The measurements of the communication process were inspired by studies from Kanki, Fork, and Irwin (Kanki et al., 1991) and Bowers, Jentsch, Salas, and Braun (1998). Although the process to analyze communication requires a lot of effort, it does shed light on the communication process and the possible effects of different conditions.

4.7.4 Additional Requirements

Observations and remarks from the participants revealed two new requirements. The first requirement states that evidence should be shown when the human asks the machine for its opinion. This means that the machine should explain the rationale of its decision (cf. Harbers, van den Bosch, & Meyer, 2011), allowing the human operator to understand why the machine comes to a specific conclusion (increases trust) and most likely increases critical thinking of the human. The second requirement expresses that the machine should explain what it has done on objects when the machine transfers objects back to the human.

4.8 Conclusions

Manning reduction initiatives and more complicated scenarios lead to a higher cognitive workload with human operators in command & control environments. Simply increasing the level of automation, however, suffers from pitfalls related to making the human operator a passive monitor, in essence pushing him or her out of the loop. Extending automation with adaptive capabilities can aid the human to overcome cognitive workload challenges and the pitfalls related to high-level of automation.

This chapter described the evaluation of an adaptive combat management system using realistic naval scenarios. An object-oriented framework was applied (elaborated upon in chapter 2) and a hybrid triggering mechanism was applied (elaborated upon in chapter 3). The evaluation utilized experienced domain experts as participants. These elements distinguish this study from other work. Furthermore, the object-oriented framework and hybrid triggering limit some risks that are associated with both (highly) automated systems and adaptive systems.

In order to get an adaptive system that could be used operationally, we needed a triggering system that was not susceptible to biases in the measurements was required. Therefore a hybrid system was opted for. To trigger the adaptive automation, an algorithm was used that is based on a combination of operator performance and an operator cognitive model. The algorithm was applied using the Wizard-of-Oz paradigm but it could have been implemented and in a fully automatic manner. Although the reactions of subjects to the trigger moments were mixed and the system is not perfect yet, it is proposed that a combination of measuring methods is more robust and leads to less automation surprises. However, more work is required to fully support this claim (i.e., to quantify the value of multiple triggers over single triggers).

This investigation examined the potential improvements claimed in the literature for adaptive automation: improved performance and decreased workload. The results did reveal a large performance improvement, especially in the more complex scenarios. This latter finding makes adaptive automation a likely candidate to incorporate in future combat management systems as military scenarios are expected to become complex (asymmetrical) while manning reduction initiatives stress the military system. Surprisingly, no improvement in workload were found. This discrepancy is explained by adaptive-operator effects (Sperandio, 1971; Veltman & Jansen, 2004) that state that domain experts adjust their task goals in order to cope with variable and highly demanding situations. More research with respect to the boundaries of this adaptability, especially under larger workloads and longer periods of underload seems worthwhile. Decision making was not affected either, providing an indication that the situational awareness of the subjects did not improve or deteriorate much due to the adaptive automation. The effects of the object oriented framework on situation awareness should be investigated more closely and this is a direction for future research.

The application of adaptive automation is a step forward in making automation act as a partner in joint activity (Hollnagel & Woods, 2005; Klein et al., 2004). Klein et al. (2004) describe ten challenges in making automation a team-player in joint activity. Adaptive automation seems to fit CHALLENGE 8 as "…*every element of an "autonomous" system will have to be designed to facilitate the kind of give-and-take that quintessentially characterizes natural and effective team-work among groups of people*" (p 93). Although this study shows the beneficial effect of applying adaptive automation to the military command & control domain, additional research is required in order to improve adaptive automation and to make automation a better team player in fields like the inference of operator intention, the expression of agent intention, and goal negotiation.

As an ePartner, being adaptive is important aspect and chapter two, three, and four of this dissertation have provided theoretical and empirical evidence that machines can be adaptive and that being adaptive have positive effects for human machine collaboration. But being adaptive is only one important aspect of ePartners and this dissertation puts forward that ePartners should also support coordination of activities while working remotely. The following chapters therefore propose observability as a solution to the problems related to working at a distance (chapter 5) and evaluate the effects of observabilty displays (chapter 6, 7, and 8).

5

OBSERVABILITY TO COMPENSATE FOR PROBLEMS RELATED TO DISTRIBUTED OPERATIONS

ABSTRACT – This chapter proposes observability as a solution to problems that are associated with coordinating activities while working across temporal or geographical boundaries. The common denominator in distributed settings is the failure by participating actors to directly observe actions or responses, and sense states of the other actors. Such observations help to anticipate information processing needs and create an awareness of the weak spots in the team. To aid with such observations, observability displays displaying performance, behavior, intention, task progression, and condition information of the remote actors are proposed. The situated Cognitive Engineering methodology is discussed in relation to the design of observability displays. The background of joint activities is discussed followed by a domain analysis of the Netherlands Urban Search and Rescue organization. Lastly, an observability display is proposed for the urban search and rescue domain. Observability displays serve as explicit tools to monitor remote actors, having a positive effect on coordination and increasing resilience to unexpected events.

This chapter is based on:

de Greef, T. E., van der Kleij, R., Brons, L., Brinkman, W. P., & Neerincx, M. A. (2011). Observability Displays in Multi-Teams. Paper presented at the 10th International Conference on Naturalistic Decision Making (NDM 2011) *.

de Greef, T. E., Oomes, A. H. J., & Neerincx, M. A. (2009). *Distilling Support Opportunities to Improve Urban* Search and Rescue Missions. Paper presented at the 13th International Conference on Human Computer Interaction, Interacting in Various Application Domains, San Diego.

de Greef, T. E., & Oomes, A. H. J. (2008). *Facilitating Synchronization and Coordination within Dispersed Emergency Management Teams.* Paper presented at the European Conference on Cognitive Ergonomics 2008.

**parts of this paper is also discussed in chapter 7*

5.1 Introduction

In correspondence with other human-computer interaction studies, this dissertation promotes a focus shift from ***automation extending*** the human to ***automation partnering with*** the human using the ePartner analogy. This dissertation therefore discusses two essential capabilities important for effective human machine cooperation, namely adaptivity and coordination. While the previous three chapters elaborated on adaptivity, the next four chapters study in which way an ePartner can assist in coordinating activities in distributed settings.

Coordination is defined as "…the deployment of team resources, activities, and responses to ensure integration, synchronization, and completion within temporal constraints…" (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995, p. 345, table 10.1) but also as "…the process of managing dependencies among activities…" (Malone & Crowston, 1994, p. 90). Both definitions are valid but the latter definition is consistent with the simple intuition that without one or more dependencies there is nothing to coordinate. Feltovich et al. (Feltovich et al., 2008) note that the essence of joint activity is interdependence, which is why the coordination of activities through time and/or space is required. Joint activity is defined as an activity "…*that is carried out by an ensemble of people acting in coordination with each other*" (Clark, 1996, p. 4). While Clark's book focused on humans, this dissertation follows the joint cognitive systems paradigm and focuses on actors that engage in joint activities in systems consisting of both humans and machines.

Team work is affected negatively when a team of actors is distributed geographically or over time (Powell, Piccoli, & Ives, 2004). Teams that are separated by a spatial or temporal boundary experience complications in the development of effective interpersonal relations, experience more frequent communication mishaps, and have a lower awareness of team members' endeavours (Thompson & Coovert, 2006). The common denominator in distributed settings is the failure to directly sense the states of remote actors and observe their actions or responses, where both actions help anticipating information processing needs and becoming aware of the weak spots in the team. An actor, for example, observing the actions of its co-actor anticipate the next step in their joint activity (Heath & Luff, 1992) and also see whether the co-actor is coping or requires assistance (cf. backing-up behavior). Observing actions leads to anticipation of the next step without requiring explicit communicative calls to coordinate activities, thereby optimizing teamwork (Figure 5.1). On the other hand, the possibility of verifying whether the co-actor copes or requires assistance leads to a team that is highly flexible in a variety of unexpected situations (Figure 5.1).

Observations, which can range from quick observations of social contextual cues to monitoring performance, serve team related processes well (Figure 5.1). However, the actions, responses, and states of other team members are not directly observable when teams encounter temporal or spatial boundaries. The goal of this chapter is to propose an observability display that overcomes challenges related to distributed work. The ePartner uses this display to support the human when coordinating joint activities in distributed settings.

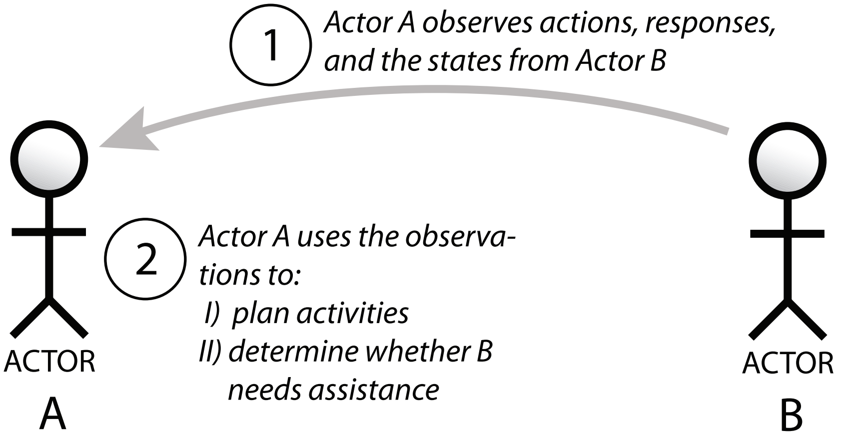


Figure 5.1 – Actor A and B coordinate their joint activities. In co-located settings, actor A observes actor B and uses the observations to either plan its own activities or determine whether B is coping with the situation and requires assistance in terms of backing-up behavior.

The concept of shared situation awareness comes to the fore in reference to distributed teamwork. Situation awareness is defined as "…*the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future*" (Endsley, 1995, p. 36). For example, an air traffic controller has good situation awareness when he or she is fully aware of all the airplanes under his or her responsibility –i.e. the aircraft in his or her sector- in such a way that he or she can anticipate what those airplanes will do in the future. Where two sectors overlap, there is a shared area between two air traffic controllers where planes are handed over. This shared area determines the overlap in responsibilities. Shared situation awareness is the degree of similarity in awareness of the aircraft in the shared area. Figure 5.2 explains, using a short scenario, the concept of (shared) situation awareness in the air traffic domain.

As an area *air traffic controller*, Anthony is responsible for making sure that flights are separated on the traffic routes according to horizontal and vertical separation rules. Further, Anthony is responsible for handing over flights to *approach controller* Josh as soon as an airplane arrives in the approach sector of an airport. The six dotted circles in the picture below represent the planes under Anthony's responsibility. One of the six planes is in the shared area and has to be handed over to Josh. As *approach controller*, Josh is responsible for safely managing the air traffic from the different traffic routes to the airport and vice versa. Josh needs to hand over planes leaving the approach area to Anthony. Of the four planes that are under the responsibility of Josh, (black circles) one needs to be handed over to Anthony.

When Anthony and Josh are able to monitor and predict the trajectories of the aircraft under their responsibility, they have a high level of situation awareness. They have a high level of **shared** situation awareness when Anthony and Josh have a similar awareness of the two planes that are in the shared area, especially if both are aware of the fact that aircraft will soon be handed over to the other.

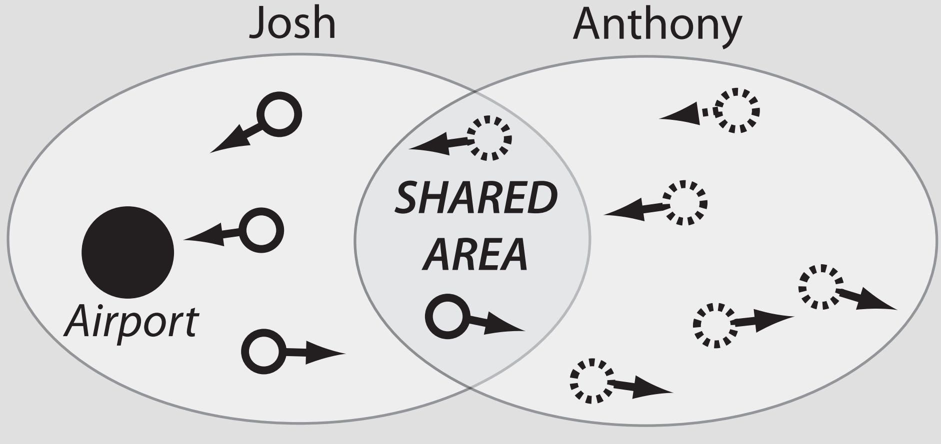


Figure 5.2 – Scenario that demonstrates the difference between situation awareness and shared situation awareness in the air traffic domain

The scenario exemplifies that both actors have a high level of situation awareness and a high level of shared situation awareness of the elements (i.e. aircraft) under their (shared) responsibility. Having a high-level of awareness allows Anthony and Josh to determine the impacts on the safety of the airplanes (i.e., one of the (shared) goals).

Supporting (shared) situation awareness requires a proper understanding of the task. Often this is achieved using a goal-directed task analysis to identify the goals, decisions, and requirements (Endsley, Bolte, & Jones, 2003, p48). However, a complicating factor in joint activity is that the actors are part of each other's environment and should be included as part of each other's situation awareness. A goal directed task-analysis insufficiently expresses such a requirement. A goal-directed task analysis helps to define situation awareness requirements and the analysis starts with goals that need to be satisfied. However, only limited information requirements related to coordination surface from a goal-directed task analysis because coordination itself is not necessarily goal on its own and mostly a hidden process. Coordination is not part of the air traffic control domain, but rather the "air traffic controller" domain and the teamwork this entails. In other words, a goal directed analysis fails to explicitly distill awareness requirements about the other actors. In co-located settings, it is the richness of cues that help team-members to be aware of and comprehend what the other actors are doing, why it is being done, and whether actors are coping. This awareness, in it its turn, promotes the coordination process and determines whether assistance is required that is beneficial to teamwork. The concept of 'observability' is proposed to address the requirement to communicate awareness about remote actors in case actors are part of each other's environment (which is when dependencies exists between the tasks the actors work on).

Observability allows actors to monitor remote co-workers in the shared environment, to comprehend what they are doing and how this impacts the joint tasks. Observability is defined as "*the perception what 'others' in your environment are doing and how they are operating allowing to determine the impact on joint activities. This requires information about the performance, behavior, intention, task progression, and mental and physical condition of the other actors*". Observability thus communicates information that isn't primary to the (shared) operational goal (cf. goal directed task analysis) but secondary allowing actors in a distributed team to operate in a variety of changed conditions and unexpected events. It takes a wider view in relation to shared situation awareness on what information is needed by participants (cf. an organization tailored situation awareness perspective, Neerincx et al, 2011).

Using an observability display has several advantages over traditional ways of working. It is expected that such displays improve the coordination process of distributed teams leading to a better performance. Secondly, it is expected that observability displays make a team more resilient to unexpected events because team-members are aware of each other facilitating the detection of performance degradations and problems with actors. In other words, we expect that a team recovers quicker (time wise) and with less of a drop in performance due to unexpected events because actors back each other up or because the team adapts.

Adaptable teams are, theoretically, envisioned as optimal (Burke et al., 2006). An adaptable team allows swift reconfiguration of the team's resources or compensatory behavior in response to changed circumstances or events that impacted the performance and/or safety significantly. Being adaptable allows a team to bring the performance back to acceptable levels and/or operate within safety boundaries. The adaptability of a team depends partly on the capability and capacity of the team to provide backing-up behavior. Backing-up behavior allows actors to provide assistance when cognitive or physical resources become depleted (e.g. in high-stress situations, unknown/dangerous situations). However, engaging in back-up behavior can only take place when actors monitor each other and check whether and how performance criteria are met.

The central point of observability is to observe others in the environment using the ePartner allowing 'you' to know what your co-actors are doing and how that impacts the joint tasks. Following the definition of observability, this requires communicating performance, behavior, intention, task progression, and condition indicators. However, how these are communicated and in which form depends on the joint activities, dependencies, and the working environment and the next section describes how the situated Cognitive Engineering methodology can be used to design for observability.

5.2 Design for Observability

The situated Cognitive Engineering (sCE) methodology (Neerincx & Lindenberg, 2008) promotes a highly iterative design process with incremental top-down development of functions and requirements. This applies the sCE methodology to the design of observability displays and discusses which ingredients to consider when designing observability displays.

The *derive* phase of sCE has a human factors perspective, a technological design space, and an operational demands viewpoint (Figure 5.3). Analyzing each of these serve as an input to the design rationale. The design rationale is an assembly of requirements, use-cases, and claims organized such that the use-cases provide context for the requirements and the claims provide justifications for the requirements.

Observability is typically a human factors aspect (block 1 in Figure 5.3), leading to improved coordination and resilience to unexpected events. Within this context, it is important to understand and define observability but also elements that are linked to observability such as coordination, backing-up behavior, team adaptation, and coping strategies. This is discussed in more detail in section 5.3. The operational demands (block 2 in Figure 5.3) provide a description of the domain, the tasks, actors involved, and the constraints of the domain. The operational demands are discussed in section 5.4. The technological design space (block 3 in Figure 5.3), finally, is envisioned by a mobile display capable to connect to other displays that communicate using ad-hoc wireless networks and possibly satellite communication technologies. In this light it is also proposed to consider ontologies and semantic web solutions allowing data to be shared and reused over organizational boundaries.

The derive phase leads to a design rationale which can be reviewed by subject matter experts and a prototype that can be evaluated, possibly using simulation technology (Figure 5.3). A design rationale and a prototype will be presented in respectively section 5.5 and 5.6.

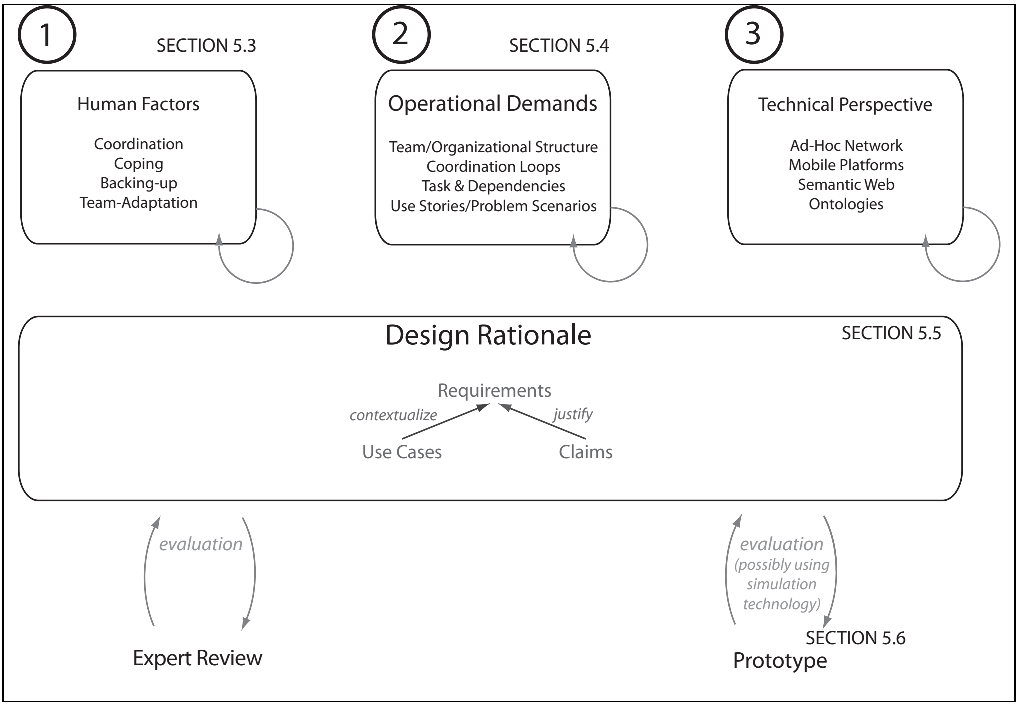


Figure 5.3 - The situated Cognitive Engineering (sCE) methodology promotes an iterative design process with incremental top-down development of functions and requirements. The human factors perspective, the operational dimension, and the technological design space serves as important input for the design rationale.

5.3 Human Factors

The starting point of observability lies in the joint activities of actors and the need to coordinate activities. Coordination has been defined as the management of dependencies (Malone & Crowston, 1994) and as such it is proposed to analyze the tasks in the domain and the dependencies that exist between the tasks and the involved (group of) actors. Prior to doing this, a solid understanding of the team and organizational structure, including the paths used to coordinate activities, is essential. The concept of coordination loops (Voshell, Woods, Prue, & Fern, 2007) fits well to structure the coordination paths. In addition, user stories and problem scenarios can be used to improve the understanding of the coordination processes in the domain. Subsequently, joint activities need to be explicated in terms of the dependencies in order to fully understand the relationships between the actors and the tasks. This is an essential step in the process and in the design of an observability display. Of course, what needs to be known is highly specific and context dependent, but analyzing the dependencies within joint activities provides a solid backbone. Table 5.1 summarizes a number of possible dependencies. A flow dependency, for example, is a dependency between two or more actors that describe that actor B can only start its activity after actor A has finished. Therefore, it is important for actors to know which task is being executed and how far they are in executing that task in order to coordinate their activities. Likewise, a simultaneous constraint dependency is a dependency where two or more actors have to be present at a specific location at a specific time.

Table 5.1 – A list of dependencies and their definitions (adapted from Malone & Crowston, 1994)

Dependency	Definition
Flow dependency	Activities have to be performed in a specific order; also known as a producer/consumer dependency
Sharing dependency	A limited amount of for example time, people, or money that has to be divided; also known as shared resource dependency
Fit dependency	Multiple activities collectively produce a product of which the pieces should fit together
Task-subtask dependency	A group of activities are all subtasks of an overarching task to achieve some goal
Simultaneous constraint	Activities need to occur at the same time

5.4 USAR Domain – Operational Demands

The previous sections have elaborated upon distributed teamwork and observability displays are proposed to aid actors in overcoming problems related to working over spatial or temporal boundaries. This section focuses on a domain that is largely distributed, namely Urban Search And Rescue (USAR). The USAR domain provides a natural setting where problems related to distributed work can be studied. This section therefore describes the USAR domain in general using observational material from two real-world trainings session of the Netherlands USAR (USAR.nl) team, operational reports, and lessons-learned documentation.

5.4.1 Method

The description of the USAR work domain is based on operational reports, lessons learned from previous missions, and a two-phase observational study. The first phase involved a study at a training session of the Netherlands USAR team (USAR.nl). USAR.nl trained for a short deployment mission in which 24-hours operations continued for five days in the Czech Republic. During two days, five research associates observed the mission, took pictures, made notes, and interviewed team members. In this setup, one research associate was always present at base camp while two research associates were with the rescue teams that worked at remote sites. Afterwards, an analysis was conducted using the gathered data and personal observations, leading to seven coordination loops and a total of eight high-level tasks that are important in the USAR domain (de Greef, Oomes, & Neerincx, 2009). The analysis was conducted to understand the coordination problems associated with distributed USAR operations. The second phase aimed to better understand the attributes of coordination bottlenecks. As such, two subsequent training deployment missions of four days in Dubai were used to annotate and understand coordination bottlenecks triggered by seven events. At each of the deployments, a team of three research associates was present to record the functioning at various locations during each of the events. Also, a daily debriefing with the commanders and the training leaders yielded substantial understanding about the reasoning of the commanders and the problems they encounter.

5.4.2 Short description of Work Domain – Structure, coordination loops, and tasks

The Urban Search And Rescue (USAR) mission's goal is to excavate victims trapped in voids after a man-made or natural disaster (e.g. an earthquake, a typhoon, a flooding) while maintaining safety boundaries for its operating members. USAR operations are characterized by extremely difficult working conditions caused by the ambiguity and uncertainty of the situation and by the physically (24 hour operations) and emotionally challenging working conditions. Furthermore, the workers are characterized by excellent competencies and a high level of motivation. The distributed setting in which USAR missions take place, however, put severe pressure on the coordination process and increase the risk of coordination breakdowns. Figure

5.4 shows how an USAR team is organized and which teams coordinate their activities.

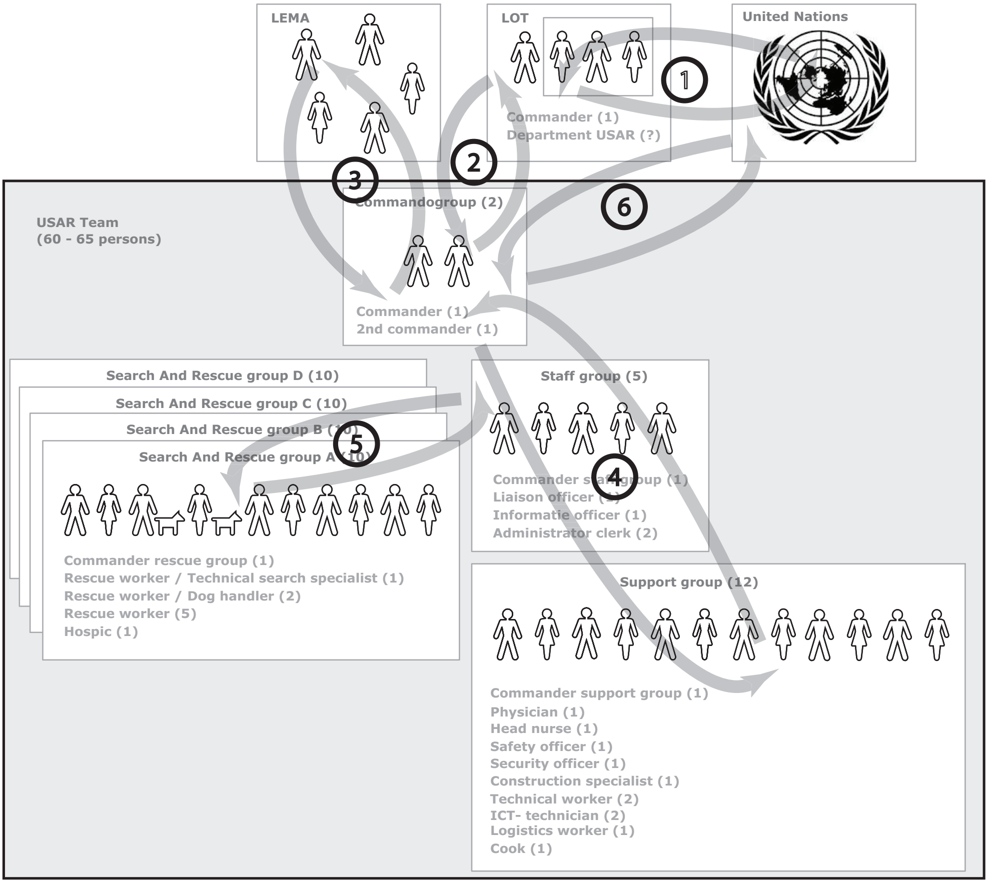


Figure 5.4 - A typical USAR organization deals with a United Nations office, a local emergency management authority (LEMA), a local operational team (LOT), a command group, a staff group, a support group, and four search and rescue groups. These groups coordinate activities using six coordination loops.

5.4.3 Observed Problem Scenarios

This section describes four coordination problem scenarios as observed during field observations (Table 5.2). The scenarios are generalized and described in such a way that they cannot be traced back to a specific team or actor, thereby insuring that they meet ethical guidelines.

The first two problem scenarios describe a typical observability problem while the last two problem scenarios highlight a typical situation awareness problem. Scenario 1 provides an example where rescue group Alpha requires a resource (i.e., drinking water) and the support team provides this resource. The staff group's unawareness of the progression of the drilling and the physical activities in relation to the time spent fails to lead to an anticipative act to, for example, call the commander of Alpha and ask whether they need anything. On the other hand, the commander of team Alpha is unaware that the requested water is on its way (and where the transporting team is) so that (s)he remains unaware about the progress of the water delivery. The second scenario also highlights an observability problem. Scenario 2 describes a rescue team that needs to be replaced by another rescue team but where the staff is unaware of the change of context (i.e. a rapid progress toward the rescue of the victim). The staff thus fails to observe the progression, thereby failing to adapt the original plan (i.e., let the rescue team finish the rescue and let the other team rest).

Scenarios 3 and 4 highlight typical situation awareness problems. Observability displays serve as an extension to situation awareness displays and as such these problems can be tackled using such a display. The third scenario describes a situation where the commander needs to actively gather information instead of inspecting an overview of the situation that updates his or her awareness of the mission. The commander needs this information in order to properly understand the state of the involved actors and this information is required to understand whether the USAR mission is operating within safety boundaries and is achieving its goal efficiently. The fourth scenario also describes a situation awareness problem as actors are unaware of available resources and fail to use these resources.

Table 5.2 – Two problem scenarios showing observability problems (scenario 1 & 2) and two scenarios highlighting SA problems (scenario 3 & 4)

Problem Scenario 1
While operating in extreme weather conditions (over 40 degrees Celsius) rescue group Alpha works in a rescue site to excavate a victim. In order to do so, the group has to drill a hole in a concrete structure and dig towards the adjacent room to reach the victim. Given the extreme working conditions and the physical demands of this rescue, the commander of the team keeps a record of the drinking behavior of the team on an hourly base in order to avoid dehydration. At some point in time the commander notices that the available water is running low in relation to the remaining time of their shift. On this notion, the commander asks the staff for additional drinking water using the satellite telephone. While the staff group accepts the call, the commander of Alpha does not receive any explicit feedback that water is being transported to the rescue site. Given the current intensity of working, the commander estimates that the supply of drinking water is finished within 30 minutes. On this notion, the commander decides to lower the intensity of activity in order to continue working on the assumption that a lower activity of the rescue workers leads to a lower consumption of water. However, within 15 minutes the water is being delivered. Although the water would have been in time easily, the commander was unaware whether water was on its way and how long it would take before the water was being delivered leading to inefficient use of human resources.

Problem Scenario 2
Rescue team Charlie is working hard to excavate a victim that previously has responded to signals. However, team Charlie needs to be replaced by team Delta according to the predefined schedule (a responsibility of the staff group). The staff group is not aware of the latest progress of team Charlie and consequently sticks to the predefined plan (low level of observability). Therefore, team Delta is awoken. However, team Delta is still tired due to their deployments at night (it is hard to sleep during the day given the high temperatures) and immediate deployment at arrival (the staff was unaware of these activities). The staff team should have decided otherwise because a tired team is replacing a team that is feeling fit. More importantly, team Charlie is very excited because they are close to a reward they have worked hard for: the rescue of a victim.

Problem Scenario 3
After a difficult meeting with the LEMA representative, the USAR commander returns to base camp and checks upon the men at base camp. Moments later, the commander enters the staff group tent and requests an update on the number of victims and the rescue work currently being done by rescue group Bravo and Charlie. While all information is available, it sits on different computers and resides with different people, some whom have gone to rest/bed. This requires the commander to utilize the satellite phone to gather information heavily interrupting the commander of Bravo and Charlie.

Problem Scenario 4
Rescue team Bravo works at a rescue site and is required to drill a hole through roughly 50 cm of fortified concrete. This requires special drilling equipment and a water to cool the drill. Four of the ten rescue workers start a search for an appropriate water supply (e.g. a pond, an underground water tank, a river). The team can transfer the water from quite some distance as they have hoses, tanks, and pumps available. After 20 minutes, a team member finds a source 100 meters away while yesterday’s rescue team that worked close by spotted an underground tank only 10 meters from the drilling spot. The team wasted time looking for a valuable water source and could have commenced drilling earlier.

5.5 Design Rationale

Table 5.3 describes a first iteration of a design rationale consisting of claims and requirements following the definition of observability in that performance, behavior, intention, task progression, and mental and physical condition information should be communicated through the observability display. Claims are defined for each requirement. Note that positive and negative claims are defined to explicate the positive but also the negative impacts of requirements.

Table 5.3 - The requirements of observability displays with claims

Requirement 1	The performance, behavior, intention, task progression, and mental and physical conditions of each remote actor should be communicated through the observability display	
Claim	**+**	The awareness of remote actors is high allowing actors to easily judge the state and activities of remote actors
	-	The display requires an additional activity effecting workload
	-	Micromanagement

Requirement 2	When actors engage in a joint activity, a visual link should be visible	
Claim	+	Better understanding which joint activities are carried out by which actors
	-	Cluttering of display

Requirement 3	Task progression shall be visible allowing all actors to observe the progress towards goal accomplishment	
Claim	+	Increased awareness of task progression allowing to spot deviations and act accordingly
	-	Insight in the progression might lead to more re-directive statements and micromanagement

Requirement 4	The physical and cognitive condition representation of each actor should be clearly indicated	
Claim	+	Allows to engage in backing-up behavior or reassignment of actors in trouble
	-	Micromanagement

Requirement 5	The task executed by an actor should be presented on the observability display	
Claim	+	Increased awareness what the other actor is doing
	-	Insight in the task might lead to more re-directive statements and micromanagement
	-	Insight in the task might lower trust

Requirement 6	When tasks and the domain have a close link to a geographical area, the actors should be plotted on the map	
Claim	+	Little effort required to map the location of actors with their state
	-	Cluttering

Requirement 7	Whenever an actor is outside the resolution of the screen, the actor is presented at the border clearly indicating that the location on the screen isn't the location of the actor	
Claim	+	The actor is aware of the state of the remote actor and can determine the effect is has on its shared task
	-	Misinterpretation of the actual location of the remote actor

5.6 An Observability Display for the USAR Domain

This section proposes an observability display for the urban search and rescue domain. The performance, behavior, intention, task progression, and mental and physical conditions information of each remote actor should be communicated through the observability display. The USAR domain is highly geographically of nature and therefore the basis of the display is a map on which several items, including the iconic representation of actors, are presented. Figure 5.5 shows the proposed iconic representation of an actor. The iconic representation shows in the center an identification label (ID tag) matching the identity of the actor. The task the actor is currently engaged in is represented below the ID tag. Figure 5.5 shows that the actor is engaged in a RECON task, which is short for reconnaissance. In addition to the current task, the progression of the current task is expressed as progression toward goal accomplishment. The goal with a reconnaissance task, for example, is to scan the area and determine an operational plan for search and rescue activities. The goal is accomplished when a specific area is scanned which can be measured using GPS coordinates and navigational analysis. In the example of Figure 5.5, the actor is roughly at ¾ of finishing the RECON task and the blue progression bar folded around the icon shows the progression towards goal accomplishment. Furthermore, the iconic representation communicates the mental and physical condition of the actor by visualizing the state laterally of the icon. In this case the actor is fit from a physical point of view (green) but (s)he has a hard time from an emotional and mental perspective (red). Intention is communicated using the directional arrow pointing in which direction the actor is moving and could easily be extended by showing an area that the actor is intending or commanded to scan. Behavior and performance can be determined using the 'tail' that represents the traversed navigated path.

The iconic representations such as presented in Figure 5.5 are to be displayed on a display. Figure 5.6 shows a map of the area around the campus of the Delft University of Technology. Two iconic representations are presented on top of this map are, namely actor C-07 and C-01.

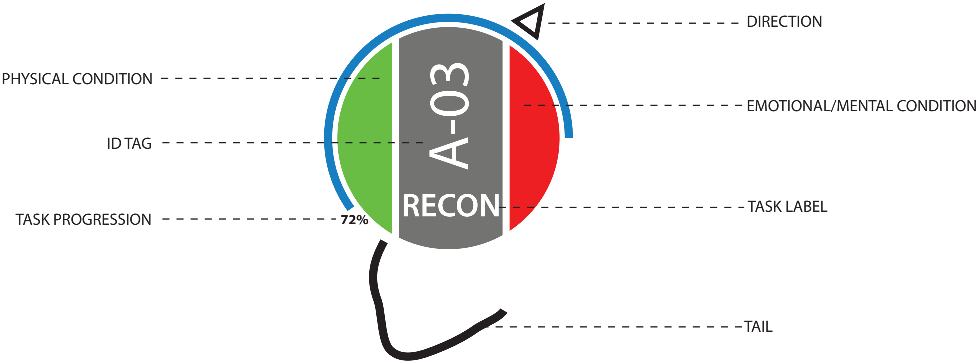


Figure 5.5 – The iconic representation of an actor shows performance, behavior, intention, task progression, and mental and physical conditions.

Figure 5.6 – The proposed observability display with two iconic representations.

5.7 Conclusions

Actors that work in close proximity benefit from observing each other. These observations help to anticipate information processing needs and generate awareness of the weak spots in the team. The lack to directly observe actions, responses, and states of actors is a key problem in dispersed teams; this chapter has proposed 'observability' as a way to make performance, behavior, intention, task progression, and conditional information of the remote actors visible using human interaction technology. Observability allows actors to monitor remote co-workers in the shared environment, leading a better understanding of what they are doing and how their actions impact the joint tasks. In this way it addresses a requirement to communicate awareness about remote actors in case actors are part of each other's environment.

Observability should be added to requirements derived for situation awareness displays. The development of situation awareness displays starts using a goal directed analysis leading to information requirements about who needs what information at what moment. However, the coordination process is seldom regarded as an explicit goal in goal directed analysis and therefore does not lead to information requirements from the perspective of the coordination process. For that reason, observability explicitly starts with the joint activities that binds actors and explicates the dependencies between actors and tasks using coordination loops and a dependency analysis.

Two beneficial effects are envisioned using observability displays. First, the use of an observability display should lead to increased anticipation on what will happen next, in this way lowering the costs of coordination while performance remains at a high level. The second benefit is the fact that the team will be more resilient to unexpected events. A quick reshuffle of resources or an adjustment of the plan is always preferable in response to an unexpected event. An observability display allows the team members to easily comprehend how well everybody is doing and why they are doing what, allowing them in turn to use this information to reshuffle resources or adjust a pre-defined plan.

While this chapter conceptualizes observability displays as explicit tools to monitor remote actors, the next three chapters describe evaluations of observability displays. The next chapter (chapter 6) studies the effect of team experience and task complexity on the use of observability displays while chapter 7 studies the effect of observability displays on backing-up behavior. Chapter 8 applies observability displays to the urban search and rescue domain and investigates whether observability leads to better responses to deviations in the predefined plan and to a better performance.

6

THE EFFECT OF TEAM EXPERIENCE AND TASK COMPLEXITY ON THE FREQUENCY OF USE OF OBSERVABILITY DISPLAYS

ABSTRACT – The aim of this chapter is to test whether coordination and the frequency of use of the observability display changes when a team gains experience and when the task is more or less complicated. Task complexity and team experience are two factors that may influence the use of observability displays. Complexity is known to inflict conflicts in goals and tasks, which has a negatively impact on the coordination process. On the other hand, team experience leads to knowledge structures that facilitate the coordination process, reducing the need to fall back on coordination tools such as observability displays. An experiment was set up that systematically controlled team experience and task complexity for a dispersed team of three persons and measured the use of the observability display, performance outcome, and the coordination process. The teams had to coordinate the sequence and the color of Soduku puzzles. The results revealed no effect of team experience on the dependent variables. However, the data revealed significant effects of task complexity. Coordination and performance decreased in the complex task conditions. More importantly, the display use frequency was significantly lower for complex tasks, corresponding with effects found in other studies. In conclusion, the display use frequency was dependent on task complexity and no evidence was found that display use changes with experience.

6.1 Introduction

Actors that work in close proximity benefit from observing each other's actions, feedback responses, and states (cf. Heath & Luff, 1992). These three factors help to anticipate information processing needs and generate awareness of the weak spots in the team. The lack of directly observed actions, responses, and actor state is a key problem in dispersed teams; observability aims to bring back those specific elements using human computer interaction technology (see chapter five).

Task complexity and team experience are two factors that potentially influence the use of an observability display. Complexity is known to generate conflicts in goals and tasks that negatively impact the coordination process. On the other hand, team experience leads to developed knowledge structures that facilitate the coordination process, in turn reducing the need to fall back on coordination mechanisms such as an observability display. Therefore it is interesting to systematically study task complexity and team experience in a controlled experiment and measure its influence on the use of an observability display and factors related to coordination and joint performance.

Team experience is essentially a learning process where team-members get accustomed to and develop knowledge of each other's habits, skills, and customs. In addition, actors learn how the interactions between the different actors evolve. Such shared knowledge is most often referred to as the 'shared mental model' of the team. Shared mental models allow teams to coordinate implicitly which is, in turn, a performance enabler (allowing teams to continue operating when things get difficult) (Cooke, Salas, Cannon-Bowers, & Stout, 2000). It is expected that more experienced teams have developed shared mental models that facilitate the coordination process, thereby reducing the need to utilize the observability display to coordinate activities.

Task complexity negatively influences the coordination process because more complex tasks generate conflicts in goals and tasks. Xiao et al. (Xiao, Hunter, Mackenzie, Jefferies, & Horst, 1996) describe a positive relation between task complexity and a focus on task-related activities at the cost of team-related activities such as coordination. It is therefore likely that increased task complexity interferes with the coordination process. The main objective of an observability display is to support the coordination process in a distributed setting and therefore it is important to examine the relation between task complexity and use of the observability display.
The aim of this chapter is to understand whether team related processes and the use of the observability display changes when a team gains experience and when the task

gets more or less complicated. Therefore, an experiment was setup that systematically controlled *team experience* and *task complexity* for a dispersed team of three persons measuring the use of the observability display, joint performance, the coordination process, and shared mental models.

The next sections will give an overview of the relevant literature and the subsequent section describes the experimental setup and how team experience and task demands were varied. The described results of the experiment are discussed in paragraph 6.4. Finally, conclusions are discussed in section 6.5.

6.2 Background

6.2.1 Shared Mental Models

Shared Mental Models are mental constructs that, in addition to various types of knowledge, also contain knowledge about the goals of the team, the individual qualities of the team members, and the agreements/procedures in place to achieve these goals. Shared Mental Models help team members to coordinate joint actions implicitly thereby reducing the amount of explicit communication (Zaccaro, Rittman, & Marks, 2001). With shared mental models, team members are able to interpret information and anticipate actions (Cannon-Bowers, Salas, & Converse, 1993). A well developed shared mental model decreases the need to fall back on alternative coordination mechanisms (Bolstad & Endsley, 1999). A mental model serves as a "*dynamic, simplified, cognitive representation of reality that team members use to describe, explain, and predict events*" (Burke, Stagl, Salas, & Pierce, 2006, p. 1199) and is seen as a basic structure of the human cognitive system (Johnson-Laird, 1983). Salas, Sims & Burke (2005) describe a shared mental model as one of three elements that benefit the coordination process. In other words, the better the shared mental model is developed, the more easily the coordination goes.

6.2.2 Task Complexity

Task complexity relates to the complexity associated with executing a task. Rasmussen (1986), for example, defines three levels of complexity. Skill based reasoning is the least complex way of reasoning and is based on developed skills that take little effort to process. A bit more complex is the rule-based reasoning mechanisms were general rules apply to a variety of situations. The most complex form of reasoning relates to knowledge based reasoning as very general concepts allow humans to cope in unfamiliar situations.

More complex tasks require more mental effort meaning that actors assign more mental effort to process the task leading to differences in subjective workload. Mental effort is often measured using a scale on which participants respond to a question about how much mental effort a task had cost them. However, it should be noted that increased task demands also have effects on variables other than workload. Veltman and Jansen (2004), for example, reported an increase mental effort or a reduction of task goals when task demands increase. Correspondingly, Sperandio (1971) showed that an increase of task demands led to altered task strategies and reduced communication in order to keep the workload at a manageable level.

6.3 Method

6.3.1 Hypothesis

The goal of this chapter is to study the effects of *team experience* and *task* on the use of an observability display, performance, and the coordination process. It is hypothesized that team experience and complexity have effects on, respectively, the shared mental model and the mental effort and are therefore measured. Given that observability displays support the coordination process, effects can also be expected on the performance and the coordination process. More specifically, seven hypotheses were generated, namely:

1. An increase of team experience results in an improved shared mental model,
2. An improved shared mental model is negatively associated with the frequency of use of the observability display,
3. An improved shared mental model results in improved coordination between the team members,
4. Complex tasks have a lower performance in comparison with low complex tasks,
5. The use of the observability display is lower in when performing high complexity tasks,
6. An increase of use of the observability display results in improved coordination between the team members, and
7. Improved coordination between the team-members results in an improved performance.

6.3.2 Design

Team experience and *task complexity* were tested systematically in a within-subject repeated measurement design. Two levels of *team experience* was distinguished by repeating similar blocks of trials (i.e. training block, block 2 and block 3). A general training session preceded the training block. The training block served to additionally train teams on the coordination of activities. The subsequent two blocks (i.e. block 2 and block 3) were the actual experimental trails where data were gathered. *Task complexity* was manipulated so that each block had a high complexity trial and a low complexity trial. A pilot test had been executed in order to select appropriate levels for the high and low task demand setting. Each block consisted of 1 trial with high task demand and 1 trial with low task demand (Figure 6.1) The order of low and high complexity trials within block 2 and block 3 was counterbalanced over teams while the sequence of complexity was fixed in the training block in order to optimize the training experience.

	TRAINING Block		Block 2		Block 3	
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
Team A	Low Complex	High Complex	Low Complex	High Complex	Low Complex	High Complex
Team B	Low Complex	High Complex	High Complex	Low Complex	High Complex	Low Complex
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Team X	Low Complex	High Complex	Low Complex	High Complex	Low Complex	High Complex

Figure 6.1 – Team experience and task complexity were the independent variables in the described experiment. Team experience increased over the blocks and task complexity was varied within a block between high and low complexity in the task.

6.3.3 Task

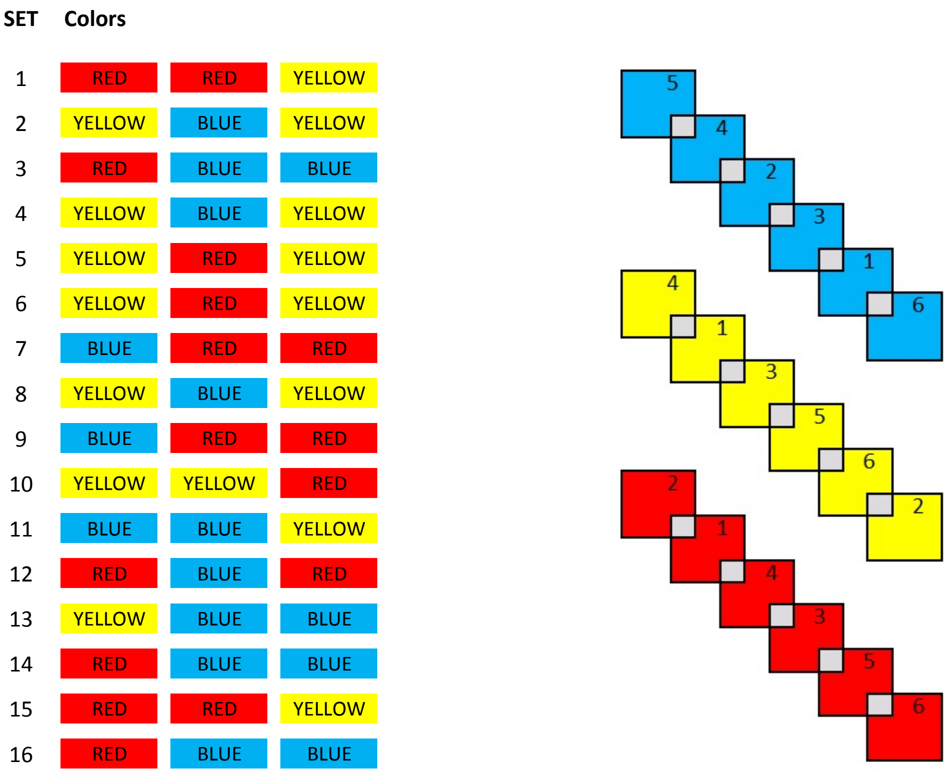
The central premise of observability displays is that it helps in coordination activities within distributed teams. Malone & Crowston (1994) defines coordination as a process of managing dependencies and therefore a team-task was developed in which team members had to manage dependencies while being in different locations. Moreover, the study applied a repeated-measurement design focusing on the longitudinal effect of *team experience* and a variation of two *task complexity* conditions.

The basis of the tasks required solving, within a distributed team of three, as many Sudoku puzzles as possible within a specified timeframe of 20 minutes. The team could earn two points for each solved red Sudoku puzzle and one point for each blue or yellow puzzle. Each puzzle was unique and could be identified by a number that was printed alongside the puzzle. The Sudoku puzzles were printed on paper and located centrally in front of the rooms of the participants, sorted by color. Participants were allowed to look through the puzzles and select a particular one.

Sudoku puzzles are puzzles that consist of a 9x9 array of cells, which can be divided in nine blocks of 3x3 cells. Some of these cells are pre-filled with numbers and the goal is to fill in the missing numbers according to the rule that in each row, column, and block the numbers 1 to 9 are to be placed exactly once.

Solving these puzzles could be done in isolation without teamwork, but coordinating (i.e., managing dependencies) the sequence and color of Sudoku puzzles led to significant benefits. The sequence and color of puzzles could be managed optimally when two dependencies were managed leading to an optimal team performance. The first dependency describes a fit dependency: each Sudoku puzzle had a different color (red, blue, or yellow) and the team could earn bonus points by handing in solved Sudoku puzzles in sets of three matching a specific and unique color combination (e.g., one red and two yellow puzzles). For each correct set of puzzles, the team earned three bonus points. At the beginning of each trial each team member received a sheet on which the color combinations were displayed (see Figure 6.2 Left). In order to hand in correct sets of Sudoku puzzles (bonus points), participants had to manage the sequence of colors in order to create sets that followed a provided color combination sheet. In other words, the team-members needed to coordinate who was working on what color puzzle. The second dependency was a flow dependency in that a part of the previous puzzle could be used in a next puzzle. The upper-left block of cells (i.e. 3x3 cells) could be copied from the lower-right block of an earlier puzzle thereby saving time because less cells need to be filled in. The numbers of lower-right nine cells could be copied to the next Sudoku puzzle provided that sequence of number matched a copy sequence sheet as provided at the start of each trial (an example is displayed in the right side of Figure 6.2). It was still possible to solve a Sudoku puzzle without a previous puzzle but participants were then forced to devote more time to find the appropriated numbers that were otherwise copied. In other words, within each color of Sudoku puzzles, there was a sequence in which the Sudoku puzzles could be made leading to a reduction of time to solve because cell could be copied. Within each trial the participants received a new sheet with the order information (Figure 6.2 right).

Within each trial, the complexity of the Sudoku puzzles was varied by manipulating the complexity of the red puzzles. In other words, the blue and yellow puzzles had the same complexity level in the high and low complex condition but the complexity of the red puzzles was varied to manage complexity.



SET	Colors		
1	RED	RED	YELLOW
2	YELLOW	BLUE	YELLOW
3	RED	BLUE	BLUE
4	YELLOW	BLUE	YELLOW
5	YELLOW	RED	YELLOW
6	YELLOW	RED	YELLOW
7	BLUE	RED	RED
8	YELLOW	BLUE	YELLOW
9	BLUE	RED	RED
10	YELLOW	YELLOW	RED
11	BLUE	BLUE	YELLOW
12	RED	BLUE	RED
13	YELLOW	BLUE	BLUE
14	RED	BLUE	BLUE
15	RED	RED	YELLOW
16	RED	BLUE	BLUE

Figure 6.2 - Left: an example of a color combination sheet distributed to participants before each trial. Handing in a particular sequence resulted in additional bonus points. Right: an example of a copy sequence sheet allowing participants to copy the upper-left part of a puzzle from another puzzle. A unique puzzle identifier identified each puzzle.

6.3.4 Observability Display

Ultimately, an observability display reflects performance, behavior, intention, task progression, and condition information of the remote actor (see chapter five). Figure 6.3 shows the observability display used in the experiment. The three participants were labeled A, B, and C and the adjacent row visualized observability information in relation to the participant. The row adjacent to participant B, for example, shows a yellow and a red block, each representing a puzzle. The color of the block matched the color of the puzzle (i.e. behavior information) and the numbers in the puzzle represented the time spent on the puzzle (i.e. performance information). The number above the block (situated on the top-left) represented the unique number of the puzzle and this number could be used in the copy sequence sheet. The overall progression could be determined by the number of finished puzzles in relation to the overall time spend on the task. Intention was communicated by the buttons on the right side of the display positioned below the label 'NEXT SUDOKU'. Participants were instructed to communicate their intention by pressing a button that corresponds to the color of the Sudoku puzzle each is going to solve next. Team members could only indicate their own intention and were not able to select the intention for another team member. Information about the physical and emotional condition of the team-members was not manipulated in the experiment and as such not represented actively but could be integrated easily using the background color of the participant identifier (now green).

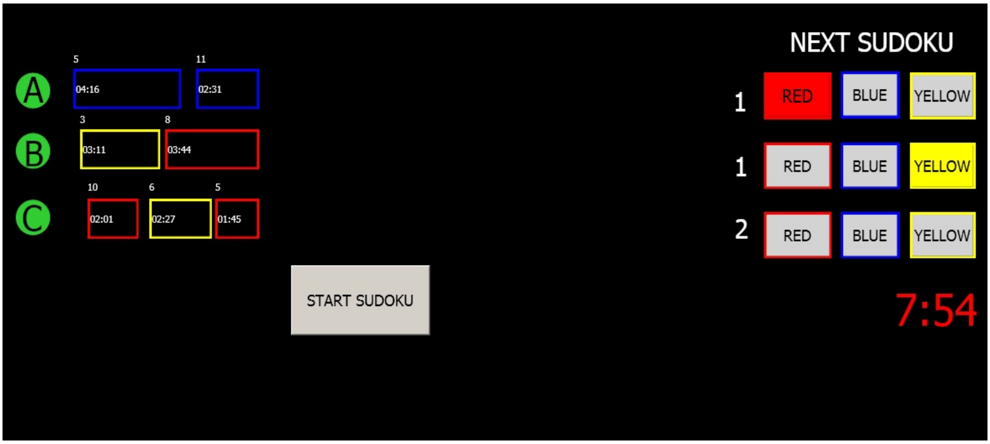


Figure 6.3 – The Observability Display that is used in the experiment. The row adjacent to the participant identifier shows the color of the puzzles worked on by the participant. The buttons on the rights side of the screen communicate which puzzle color the participant is going to work on next.

Participants had to provide input for the observability screen on two additional occasions. First, participants had to press the 'start Sudoku' button when they started a Sudoku puzzle. If the participants pressed the start-button, they were requested to fill in the number of the puzzle resulting in a new (puzzle) bar appearing on the screen where the color of the bar was identical to the color selected in the intention-part of the display. The bar became white when the participant previously had selected no intention. The 'start sudoku' puzzle changed to 'stop sudoku'. As long as a participant was working on a Sudoku puzzle, the bar became wider, until the 'stop sudoku' button was pressed. Secondly, the participants had to click with the mouse in order to make the display visible for 20 seconds. After 20 seconds the observability display blanked-out automatically. The number of occasions to make the display visible was logged automatically and used as measure for frequency of use of the observability display.

The observability display allowed participants to view specific information about the other team-members allowing anticipation of which color puzzles were to be worked on by whom. In this way, participants could adapt their work to the activities of the other team members (i.e., to coordinate activities) and combine color sequences. For optimal performance, participants had to know the colors of puzzles that were completed and which puzzle number was done to determine the next color and puzzle number. The past colors were reflected in the observability display by coloring the puzzle that was worked on per team member. The next puzzle number could be determined using the unique puzzle identifier (as represented on the display) in combination with the copy sequence sheet.

6.3.5 Participants

Forty-eight participants (20 men, 28 women, $M_{AGE} = 23.44$, $SD = 3.02$) took part in the experiment. Because of the complexity of the used task, participants were selected from a pool of bachelor and master students. The participants were randomly assigned to one of sixteen three-person teams. It was assured that none of the team members knew each other prior to the experiment. Participants were compensated financially for their participation and a significant financial bonus was awarded to the best performing team in order to enhance motivation. As teams did not change significantly on the background variables (education, age, team experience, Sudoku experience), these variables are not discussed further.

6.3.6 Procedure

Upon arrival, the participants received a general oral instruction about the experiment in which they were told that the purpose of the experiment was to study teamwork in distributed settings and that the goal of the participants was to maximize performance. After participants filled in an informed consent form, a detailed written explanation of the task was given and the observability display was explained to the participants. Subsequently, the participants were situated in different rooms to practice with the observability display and requested to fill in a background information questionnaire (age, gender, education, experience with Sudoku puzzles, teamwork, and so forth). Subsequently, participants were given some time to practice a Sudoku puzzle and the observability display. After this initial training, participants trained one block (see Figure 6.1) in order to gain experience with the observability display and the coordination process.

The experimental trials commenced after the training block. At the beginning of each trial, the experimental leader gave each participant a sheet with new order-information and new color combinations for the puzzles in that trial; each trial started with a new set of Sudoku puzzles. Participants also received an envelope that contained a short questionnaire concerning the activity of team-members. During each trial, there was a freeze moment in which the puzzle task was paused, the display blanked out, and the participants had to fill in the activity awareness questionnaire. The freeze moment was between 7 and 13 minutes after the start of the block, indicated to the participants with an auditory signal. The order of moments at which the task was frozen, was counterbalanced among the teams limiting the participants to guess the timing of the freeze moment. When all participants had filled in the questionnaire, the task continued. At the end of each trial the participants were required to fill in a questionnaire pertaining to task demand, team coordination, and shared mental model. At the end of each trial, the experiment leader collected the Sudoku puzzles, information forms, questionnaires, and noted the performance of the team.

6.3.7 Apparatus

Four rooms were used for the experiment. Three rooms were used to situate a participant and the experimental leader used the fourth room. Each participant room contained a desk with a standard desktop computer (Figure 6.4). The computer screen showed the observability display and the participant could use the mouse and keyboard to provide input. In the experimenter room, the experimenter was able to monitor the participant rooms with use of video cameras and the session was also recorded on tape. The computers were connected through a standard TCP/IP network and an observability display was developed in C# on top of the Planning tasks for Teams (PLATT) environment (Kamphuis & Houttuin, 2010) facilitating the data transfer between the various computer displays and the start and end of each trial.

Figure 6.4 – The participant room contained a standard desktop computer on which the observability display was shown. Each participant was located in a different room and the experimental leader was situated in another room to control the experiment and monitor the participants using camera feeds.

6.3.8 Dependent measures

The perceived *complexity* was measured in two ways in order to validate whether the manipulation of task demands succeeded. The task demand was measured by a questionnaire derived from Maynard and Hakel (1997) and the rating scale mental effort (RSME). The questionnaire contained three items and participants had to rate the items 'I found this to be a complex task', 'This task was mentally demanding', 'This task required a lot of thought and problem-solving' on a 7-point Likert type scale labeled from *strongly disagree* to *strongly agree'*. The RSME (Zijlstra, 1993) is a one-dimensional scale with ratings from 0 to 150 on which participants have to respond to the question 'how much effort did it cost you to fulfill the task?'. The scale has nine descriptive indicators along its axis (e.g., 12 corresponds to not effortful, 58 to rather effortful, and 113 to extremely effortful).

Coordination was measured after each trial with an adjusted version of the Inter-team Coordination Questionnaire of Hoegl, Weinkauf, and Gemuenden (2004). Participants had to respond to four items ('Activities were well coordinated with other team members', 'Coordination with other team members went smoothly', 'Double and overlapping activities were avoided' and 'Conflicts with other team members were settled quickly') on a 7-point Likert type scale labeled from '*strongly disagree*' to '*strongly agree*'.

The *shared mental* model was measured based on a questionnaire defined by Austin (2003). After each trial, participants were asked to judge the skills of themselves and of the other team members on a 7-point Likert scale labeled from '*very bad*' to '*very good*'. Participants were required to rate three skills, namely '*solving Sudokus*', '*coordinating tasks*', and '*contribute to making color combinations*'. The shared mental model was calculated (following Austin, 2003) on basis of the variance in ratings on each item that were calculated per trial. This means that the lower the variance between skill judgments is, the better developed the shared mental model is.

The awareness about the team-member's activities (activity awareness) was measured at a specific moment within each trial by freezing the task and asking participants to fill in an activity awareness questionnaire. The activity awareness questionnaire consisted of five items that were derived from a larger pool of questions. Six different versions of the questionnaires were used in counterbalanced order and the full list of questionnaires can be found in Appendix A. Examples of observability questions are '*how long is person A working on his puzzle?*', '*what is the color of the puzzle person B is working on?*', and '*how many puzzles has person C solved?*'. The number of correct answers serves as a measure of activity awareness.

Two *performance* measures were used. The number of complete puzzles per team per trial and the total score served as performance measures. The total score is the amount of points per team per trial. For each blue or yellow puzzle, the team received one point, for each red puzzle the team received two points. The team received three points for each handed in set of puzzles that consisted of the right colors.

The *usage of the observability* display was recorded by logging the amount of times that the participants had to reinitiate the observability display. The observability display would blank-out after 20 seconds and participants had to click the display to see the information (i.e. re-initiation).

Table 6.1 – Overview of the six dependent variables that were measured and how these were measured

Dependent Variable	How	Cronbach a
Complexity	1. Questionnaire	a = 0.88
Mental Effort	2. RSME	n/a
Coordination	3. Questionnaire	a = 0.92
Shared Mental Model	4. Variance on skill judgment	a = 0.73
Activity Awareness	5. Questions at freeze moments	n/a
Performance	6. Number of puzzles	n/a
	7. Score	n/a
Display Use Frequency	8. Logging frequency of re-initialization	n/a

6.4 Results

A 2 x 2 General Linear Model Repeated Measures design was used for data analysis, with *team experience* (block 2, block 3) and *task complexity* (low, high) as within-group variables. Data were aggregated and analyzed at the team level.

Main effects of team experience, task demand, and the interaction effect were examined. An alpha level of .05 was adopted to report significant results. Post-hoc multiple comparisons were done using Least Significant Differences comparisons. Although this test does not correct for the number of tests performed, this test was used because of the limited power as 16 teams participated in the experiment. Correlations between dependent variables were calculated using Spearman's rho (Table 6.4). Table 6.2 and Table 6.3 show respectively a summary of the repeated measures ANOVA, the mean values and standard deviation of the dependent variables, and the correlation matrix.

As the use of display could be influenced by the number of puzzles completed, it was tested whether the number of puzzles should be used as covariate when analyzing the use of the observability display. MANOVA Wilks's lambda showed that there was no significant relation between the use of the observability display and the number of puzzles completed, $F(36,20) = 1.16$, $p > 0.5$.

6.4.1 Complexity

Complexity was measured using a questionnaire and the RSME. Repeated-measures ANOVA showed no significant effect of team experience on task complexity questionnaire scores, $F(1,15) = 0.47, p = .50, \eta_p^2 = 0.30$. However, the scores on the complexity questionnaire were significantly higher after in the high complexity trials ($M = 5.19$, $SD = 0.63$) in comparison to the low complexity trials ($M = 4.78$, $SD = 0.66$, $F\ (1,15) = 19.60, p < .01, \eta_p^2 = 0.57$). The data revealed no interaction effect of team experience and complexity on the complexity questionnaire ($F(1,15) = 0.88, p = .77, \eta_p^2 = 0.01$).

Repeated-measures ANOVA showed no significant effect of team experience on RSME, $F(1,15) = 3.12, p = .10, \eta_p^2 = 0.17$. Likewise, no effects were found of task complexity on RSME ($F(1,15) = 1.39$, $p = .26$, $\eta_p^2 = 0.09$) nor on the interaction between team experience and task complexity ($F(1,15) = 0.66, p = .43, \eta_p^2 = 0.04$).

6.4.2 Coordination

The results revealed no effect of team experience on the coordination process, $F(1,15) = 4.31$, $p = .06$, $\eta_p^2 = 0.22$. However, a significant effect of task complexity on coordination was found ($F(1,15) = 10.95$, $p < .01$, $\eta_p^2 = 0.42$). The coordination process was rated higher in the low complexity trials ($M = 5.27$, $SD = 0.77$) in comparison to high complexity trials ($M = 4.84$, $SD = 0.73$). The data revealed no interaction effect ($F(1,15) = 0.00, p < .95, \eta_p^2 = 0.00$).

6.4.3 Shared Mental Model

The *shared mental* model was based on a questionnaire that was offered to participants after each trial and asked participants to rate the skill of team-members. In a subsequent step, the variance between these measures was calculated and served as an indicator of the shared mental model. A lower variance equals a better shared mental model. The shared mental model measure shows no significant influence of team experience ($F(1,15) = 0.15, p = .70, \eta_p^2 = 0.10$). However, a significant effect of task complexity was found, $F(1,15) = 9.34, p < .01, \eta_p^2 = 0.38$. The shared mental model was higher in the high complexity condition ($M = 1.46$, $SD = 0.86$) compared to the low complexity condition ($M = 1.08$, $SD = 0.62$). The data revealed no significant interaction effect of team experience and task demand ($F(1,15) = 0.01, p = .95, \eta_p^2 = 0.00$).

6.4.4 Activity Awareness

The data on activity awareness of team-members measure showed no effect on team experience ($F(1,15) = 2.74$, $p = .12$, $\eta_p^2 = 0.15$). However, task complexity did reveal significant differences ($F(1,15) = 7.54$, $p < .05$, $\eta_p^2 = 0.33$). The awareness about the activities of team-members was higher in the high complexity condition ($M = 3.50$, $SD = 0.73$) compared to the low complexity condition ($M = 3.15$, $SD = 0.66$). No significant interaction effect was found, $F(1,15) = 0.41$, $p = 0.53$, $\eta_p^2 = 0.03$.

6.4.5 Performance

Performance was measured in two ways: first, the number of puzzles was recorded and analyzed. The data revealed no effect of team experience on the number of puzzles made ($F(1,15) = 0.92$, $p = .35$ $\eta_p^2 = 0.06$). However, the number of puzzles in low complexity trials ($M = 3.68$, $SD = 0.96$) was significantly higher compared to puzzles made in high complexity trials ($M = 1.98$, $SD = 0.59$, $F(1,15) = 174.78$, $p < .01$, $\eta_p^2 = 0.92$). No significant interaction effect of experience and task demand was found, $F(1,15) = 0.53$, $p = .82$, $\eta_p^2 = 0.00$.

Secondly, the total score was measured. There was no effect of team experience on the scores, $F(1,15) = 0.72$, $p = .41$, $\eta_p^2 = 0.05$ but the score in low complexity trials differed significantly from the score in high complexity trials, $F(1,15) = 166.89$, $p < .01$, $\eta_p^2 = 0.92$. The score was higher in the low complexity condition ($M = 23.91$, $SD = 0.6.81$) compared to the high complexity condition ($M = 12.22$, $SD = 4.21$). No significant interaction effect of experience and task demand was found, $F(1,15) = 0.18$, $p = .68$, $\eta_p^2 = 0.01$.

6.4.6 Use of the observability display

The use of the observability display was measured by logging how often participants clicked on the screen to make the display visible. The amount of clicks were not influenced significantly by team experience, $F(1,15) = 0.03$, $p = .87$, $\eta_p^2 = 0.00$. However, a significant effect of task demand on the use of the observability display was found, $F(1,15) = 69.37$, $p < .01$, $\eta_p^2 = 0.82$. In the high complexity trials, the observability display was used less ($M = 12.49$, $SD = 2.26$) than in the low complexity trials ($M = 16.00$, $SD = 2.69$). No significant interaction effect of experience and task demand was found, $F(1,15) = 0.91$, $p = .36$, $\eta_p^2 = 0.06$.

6.4.7 Correlations

The dependent variables were analyzed for correlations and this section reports the significant findings (see Table 6.4). First of all, complexity was measured in two ways and both measures correlated significantly ($r_s = 0.42$, $p < .01$). Secondly, the performance was also measured in two ways and these two measures also correlated significantly ($r_s = 0.98$, $p < .01$). Third, there was a negative significant relation between the two complexity measures and the two performance measures (all $p < .05$, see Table 6.4). Fourth, the coordination measures correlated negatively with the RSME ($r_s = -0.25$, $p < .05$) but not with the complexity questionnaire. Likewise, a significant positive correlation existed between coordination and both performance measures ($r_{s,\, puzzles} = 0.42$, $r_{s,\, score} = 0.43$, both $p < .01$). Fifth, the shared mental model correlated positively with the complexity as measured with the task complexity questionnaire ($r_s = 0.35$, $p < .01$) but not with the RSME. Sixth, the display use frequency correlated positively with both performance measures ($r_{s,\, puzzles} = 0.56$, $r_{s,\, score} = 0.53$, both $p < .01$).

Table 6.2 - The results of the repeated measurement ANOVA

	Team Experience			Task Complexity			Experience x Task Complexity		
	F(1,15)	*p*	η_p^2	*F(1,15)*	*p*	η_p^2	*F(1,15)*	*p*	η_p^2
Complexity	0.47	0.50	0.30	19.60	**<0.01**	0.57	0.88	0.77	0.01
Mental Effort	3.12	0.10	0.17	1.39	0.26	0.09	0.66	0.43	0.04
Coordination	*4.31*	0.06	*0.22*	10.95	**0.01**	0.42	0.00	0.95	0.00
Shared Mental Model	0.15	0.70	0.10	9.34	**0.01**	0.38	0.01	0.95	0.00
Activity Awareness	2.74	0.12	0.15	7.54	**0.02**	0.33	0.41	0.53	0.03
Performance - Puzzles	0.92	0.35	0.06	174.87	**<0.01**	0.92	0.53	0.82	0.00
Performance - Score	0.72	0.41	0.05	166.89	**<0.01**	0.92	0.18	0.68	0.01
Display Use Frequency	0.03	0.87	0.00	69.37	**<0.01**	0.82	0.91	0.36	0.06

Table 6.3 – Mean values and standard deviations of the dependent variables

	Team Experience				Task Complexity			
	Block 2		Block 3		Low		High	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Complexity	5.03	0.74	4.96	0.60	4.78	0.66	5.19	0.63
Mental Effort	76.54	11.27	73.67	10.42	74.21	10.79	76.01	11.03
Coordination	*4.94*	0.79	*5.17*	0.76	5.27	0.77	4.84	0.73
Shared Mental Model	1.30	0.69	1.24	0.84	1.08	0.62	1.46	0.86
Activity Awareness	3.21	0.74	4.44	0.68	3.15	0.66	3.50	0.73
Performance - Puzzles	2.75	1.16	2.91	1.17	3.68	0.96	1.98	0.59
Performance - Score	17.63	8.25	18.50	8.14	23.91	6.81	12.22	4.21
Display Use Frequency	14.29	3.08	14.20	3.03	16.00	2.69	12.49	2.26

Table 6.4 – The matrix listing the correlations between the dependent variables

Variable		1	2	3	4	5	6	7	8
1	Task Complexity	-							
2	RSME	.42**	-						
3	Coordination	.17	-.25*	-					
4	Shared Mental Model	.35**	.15	.03	-				
5	Activity Awareness	.14	.07	-.07	-.01	-			
6	Performance - Puzzles	-.44**	-.26*	.42**	-.21	-.20	-		
7	Performance - Score	-.47**	-.28*	.43**	-.23	-.21	.98**	-	
8	Display Use Frequency	-.15	-.02	.16	-.02	-.01	.56**	.53**	-

*$p < .05$, ** $p < .01$

6.5 Discussion

Task complexity and team experience are two factors that potentially influence the use of an observability display. Complexity is known to generate conflicts in goals (Xiao et al., 1996) and tasks that negatively impact the coordination process and team experience leads to developed knowledge structures that facilitate the coordination process. An experiment was conducted to study the effects of task complexity and team experience on the use of an observability display: this section discusses the results.

6.5.1 Hypotheses

The complexity questionnaire and Rating Scale Mental Effort were used to examine whether the manipulation of task complexity succeeded. The task complexity questionnaire showed that the high complexity trials were perceived as more complex than the low complexity trials. However, the RSME did not show a significant difference between high and low complexity trials. Although the RSME aims to reflect the subjective mental effort, the manipulation of task complexity failed to show differences in subjective rating of mental effort. Given that mental effort is a multi-faceted construct (cf. Neerincx, 2003) it is not surprising that a direct measure of complexity shows an effect when complexity is manipulated while mental effort fails to do so. It was concluded that the manipulation of task demand was successful.

It was expected that the shared mental model, which contains knowledge about how the team operated and teamwork was executed, would develop with experience (hypothesis one). The shared mental model was measured using a shared mental model questionnaire that was administered after each trial. The results, however, fail to show an increase of a shared mental mode and therefore hypothesis 1 is not supported.

Hypothesis two expressed that an increased shared mental model is negatively associated with the use of an observability display and hypothesis three claimed a positive effect between shared mental models and coordination. The results show no effects of shared mental model on team experience, making it difficult to conclude anything about these two hypotheses. However, the effect of team experience on use of the observability display was compared given that team experience should theoretically lead to an improved shared mental model. Unfortunately, no effect of team experience on the use of the observability display was found, thus there is no data to support hypothesis two. With regard to hypothesis three, shared mental models are seen as enablers to the coordination process and therefore the results of team experience on coordination are compared, but failed to show an effect. Therefore, hypothesis three cannot be supported.

Hypothesis four states that more complex tasks lower the performance and the results provide support for this hypothesis. Both the number of puzzles finished and the score are higher in the low complexity condition. Hypothesis four is additionally supported by a lowered score of the coordination process in the more complex task condition.

According to hypothesis five, the use of the observability display is less when performing high complex tasks. The results reveal a large effect of task complexity on the display use frequency in that the display was used more in the lower complex task. It should be noted that more puzzles were finished in the low complexity condition, leading to a lowered frequency of display use as a result of the fact that less puzzles were completed in the high complexity condition. Recall that the test whether the number of puzzles should be used as covariate when analyzing the use of the observability display revealed no significant relation. Hypothesis five is therefore supported. This corresponds with Xiao et al. (1996) who report that an increased task demand leads to conflicts in goals and tasks and as such on the coordination process. Likewise, Sperandio (1971) and Veltman and Jansen (2004), describe altered task strategies in situations of increased demands. Team members have to divide their attention between the puzzle task and coordination process. In high complexity tasks, solving the puzzle had more priority, which resulted in a decreased use of the observability display.

It was hypothesized that an increase of the use of observability information would result in improved coordination (hypothesis six). The results showed no significant correlation between the display use frequency and the coordination process. However, a positive correlation between the display use frequency and performance was found. Given that improved coordination leads to improved performance, it seems that the increased display use frequency indirectly benefitted the coordination process and therefore supports hypothesis six.

The hypothesis concerning the positive effect of coordination on performance (hypothesis seven) can be supported by the results of the experiment. The correlation between both measures of performance (score as well as the number of puzzles) was very strong ($r = 0.98$) with the ratings of coordination.

A number of other effects were revealed in the data and follow expectancies as expressed in human factors literature or team literature. The negative correlation between the two performance measures and task complexity and the RSME follow the logic that more complex task(s) that require more mental effort lead to a lower performance. Also, the two performance measures correlate and the RSME and task complexity measures correlate.

6.5.2 Applied Measures

This section discusses some of the dependent measures used in the experiment that raise some concern with the validity of the measures.

A combination of unexpected results raised some questions with regard to use the activity awareness questionnaire and the shared mental model measure as a representative measures. In hindsight, it seemed that the amount of information on the display confounded with the number of puzzles. The questionnaire consisted of questions concerning the information on the observability display. The more information displayed on the screen, the more difficult it was to answer the questions correctly. As participants completed more puzzles in the low complexity trials, more information was shown on the display in this condition which resulted in less correct answers on activity awareness questions in the low task demand condition, raising doubt as to the validity of this measure. Because of this, it was decided to reject the activity awareness questionnaire.

The results of the shared mental model measure cannot confirm hypothesis one as the results failed to reveal an increase of team experience on the shared mental model. Remember that the deviation in participants' judgments of each other's skills was used as a measure meaning that a lower deviation is seen as a better shared mental model. It is arguable that the measure used corresponds with a shared mental model. In other words, it might be that the shared mental model did develop but that the applied measure failed to record such models accurately. Literature reports that measures of shared mental model are very comprehensive and time-consuming (Mathieu, Goodwin, Heffner, Salas, & Cannon-Bowers, 2000; Stout, Cannon-Bowers, Salas, & Milanovich, 1999). Because of the limited time available during the experiment, it was decided to use a more basic and less time consuming construct. Further, one could argue whether the four hour duration of the experiment, was enough to develop a shared mental model. It could be that a period longer than four hours is needed to develop a shared mental model, which resulted in a lack of an increases in the shared mental model in this study. Because of these two reasons, it is more plausible to conclude that the used measurement was not adequate for measuring the shared mental model than it is to conclude that the shared mental model did not develop over time.

6.6 Conclusion

A common denominator in distributed settings is the failure to directly observe actions, responses, and states of remote co-workers, which is a bottleneck in anticipating information processing needs and managing resources adequately and promptly. Chapter 5 of this dissertation claimed that observability displays increase the efficacy of distributed teams and lead to increased responsiveness to unexpected situations.

The main findings of this study were that the display use frequency is dependent on the task complexity and no evidence is found that the frequency of display use changes when a team gains experience. Moreover, the results show clear effects between task complexity, coordination, and performance.

The goal of the present study was to examine the effects of team experience and task demand on the use of an observability display. In this way, more insight was gained into how an ePartner supports human actors in a distributed team. The present study has shown that the use of the observability display, coordination, and performance within distributed teams dependents on the complexity of the situation. The diminished use of observability displays in a complex setting requires the consideration that observability displays should potentially adapt different notification styles based on the complexity of the environment (cf. Streefkerk, Esch-Bussemakers, & Neerincx, 2007).

The aim of an observability display is to improve the coordination of joint activities. Coordination can manifest in many different forms: the most common effect of an improved coordination process manifests as an increase in performance. However, there are a number of alternative aspects of improved coordination. McIntyre and Salas (1995) emphasize the importance of backing-up behavior as a component of teamwork and that backing-up is an important aspect of the coordination processes. Backing-up behavior relates to providing assistance whenever the capacity of one team is being surpassed. Therefore, the next chapter will study the effect of observability displays on backing-up behavior.

7

OBSERVABILITY WITHIN DISTRIBUTED SUB-TEAMS

ABSTRACT – The objective of this chapter is to examine the effects of an observability display in a distributed setting on backing-up behavior and team performance using a complex information-processing task. Backing-up behavior concerns assistance whenever the capacity of one team member is being surpassed and is a key component of team adaptiveness. A team that adapts quickly in response to unexpected events is resilient and can cope with a variety of situations. In an experimental between-subject design, 20 teams processed and distributed information and needed to solve Sudoku puzzles in another phase. Performance, coordination, communication, and workload were measured. The observability display did not affect performance but showed beneficial effects on backing-up behavior and communication. The benefits of backing-up behavior and communication provide a strong cue of the beneficial effects of observability, as both are important elements in teamwork. However, no evidence was found that observability displays decreased mental effort.

This chapter is based on:

de Greef, T. E., van der Kleij, R., Brons, L., Brinkman, W. P., & Neerincx, M. A. (2011). Observability Displays in Multi-Teams. (S. M. Fiore & M. Harper-Sciarini, Eds.)*10th International Conference on Naturalistic Decision Making (NDM 2011).* Orlando, FL, U.S.A.: University of Central Florida.

7.1 Introduction

The lack to directly observe actions, responses, and actor state is a key problem in dispersed teams and chapter five of this dissertation proposed observability as a way to bring back those specific elements using human computer interaction technology. The aim of an observability display is to improve the coordination of joint activities and being increasingly resilient to unexpected events while working distributed.

The previous chapter demonstrated that the complexity of the task has an influence on the display use frequency and this chapter questions in what way backing-up behavior emerges when using an observability display. Backing-up behavior relates to providing assistance whenever the capacity of one team is being surpassed. McIntyre and Salas (1995) emphasize the importance of backing-up behavior as a component of team adaptiveness and that backing-up is a manifestation of the coordination processes.

The main objective is to examine the effects of an observability display on backing-up behavior and team performance using a complex information-processing task. In addition, communication and mental effort are measured and compared. Coordination is studied because communication allows actors to coordinate joint activities and it is expected that the observability display provides a more natural way to 'keep an eye' on things and thereby lower the need to communicate verbally. On the other hand, workload describes the ration between task demands and the capacity of the actor. It is expected that the task demands related to coordination when using an observability display are lower for two reasons. First, the display shows progress information that is otherwise communicated verbally. Perceiving information compared to verbally communicating information is expected to be less demanding. Secondly, the display represents tasks on an abstract way following results of Dabbish & Kraut (2008) who reported least cognitive demands when a the environment and tasks were represented on an abstract way.

More specifically, four directional hypothesis were generated, namely that:

1. more backing-up behavior occurs with teams using the observability display,
2. the performance is higher when working with the observability display,
3. less communication takes place with teams that use the observability display, and
4. the mental effort is lower for team that use the observability display.

The subsequent section discusses the methods and materials utilized throughout the experiment. Section 7.3 describes the results and section 7.4 attributes conclusions to these results. Section 7.5 considers a number of practical considerations and the concluding section highlights future research opportunities.

7.2 Methods & Materials

7.2.1 Participants

The experiment was conducted with 20 teams and each team comprised two sub-teams that were situated in different rooms (Figure 7.1). Each sub-teams consisted of two participants. A total of 80 participants (38 men and 42 woman, $M_{AGE} = 23.98$, $SD = 3.04$) participated in this experiment. The participants were selected from a population of bachelor and master students from a non-technical university and received €40 for their contribution. In addition, a €120 bonus was awarded to the best performing team to enhance motivation.

7.2.2 Experimental Design

A between-subject experimental design was applied with *display type* as the only independent variable. The display type could either be 'on' or 'off' meaning that a team had an observability display available or that a team was working without an observability display. Each team thus experienced one condition and each team executed in one trial (no repeated measures).

The experiment was constrained in two ways. First, the participants were not allowed to speak during the experiment. The participants could, however, communicate in a digital way (i.e. chat and e-mail) facilitating the analysis of the communication. Secondly, the digital communication was constrained such that within each sub-team only one person had the possibility to communicate with the collaborating sub-team (Figure 7.1). The communication constraint followed a classical military and USAR command structure in that the commander of a battle or rescue group serves as the communication hub.

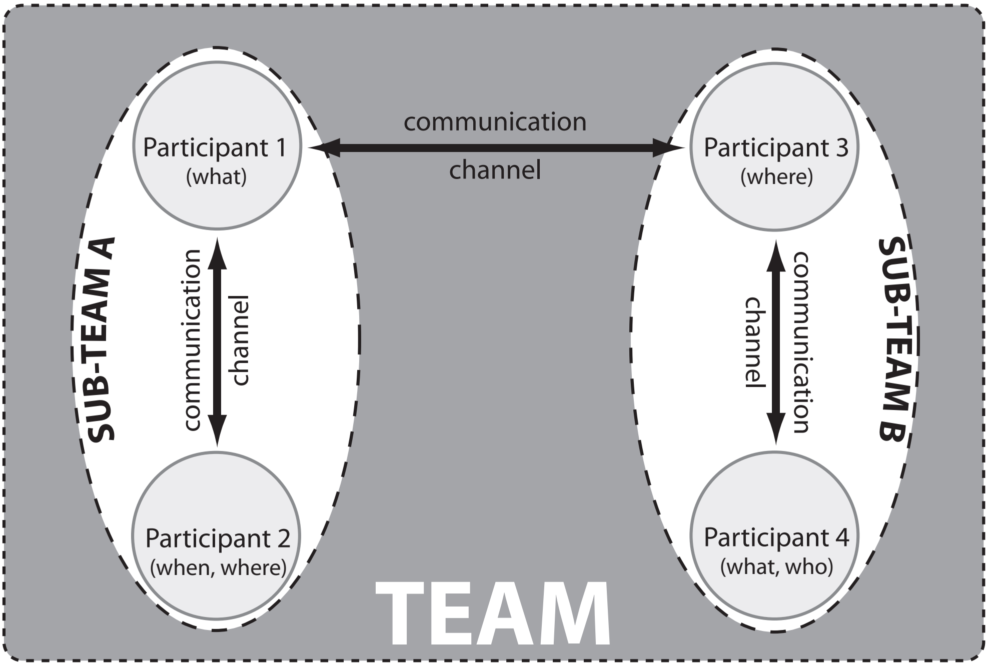


Figure 7.1 – A team comprised two sub-team and each sub-team contained two participants. Each sub-team was located in a different room. One of the two participants in a sub-team communicated with the other sub-team using chat or e-mail. Within a sub-team the participants communicated with each other using the same digital functionality. Verbal communication was prohibited. Dependent on the position within the sub-team, each participant was responsible to provide one or two answers to questions related to a terrorist attack (what, where, when, who).

7.2.3 Procedure

At arrival, the experiment leader confirmed that participants were not familiar with each other. After describing the house rules (switch off mobile phone, no smoking, when to use the restroom) the experiment leader explained the experiment. Subsequently, the participants received written en verbal instructions about the experiment and the task. Moreover, it was stressed that the team should win as a global team, achieve the best performance as a team and not regard the other sub-team as a competitor. Participants were invited to fill in a demographic questionnaire and a consent form. Next, participants were trained on the task and environment using a test scenario. After the training, the actual experiment commenced. After the trail, the participants filled out a paper-based questionnaire followed by a short debriefing session, which allowed participants to ventilate their experiences.

7.2.4 Task

The task used in this experiment comprised an information-processing task and a puzzle task that needed to be executed in sequence. Embedding a two-tasks-sequence was done in order to stress that teamwork often incorporated switching between different tasks. Intelligence agencies, for example, switch constantly between reconnaissance tasks and information sharing tasks. Likewise, USAR missions require reconnaissance of the environment before a staff can determine the optimal deployment of rescue groups. After deployment, this staff switches between a monitoring task and a re-planning task. The two-task-sequence within this experiment thus aimed to create a close link to real world team tasks in high-risk professional settings (chapter 1 contains an elaboration on high-risk professional domains).

The information processing task was based on the Experimental Laboratory for Investigating Collaboration, Information-sharing, and Trust (ELICIT) task (Lospinoso, 2007). In ELICIT, information factoids provided cues to four questions (Table 7.1) in relation to a terrorist attack. A factoid was an ambiguous statement requiring participants to combine several factoids in order to provide correct answers to questions. As an example, consider the following four factoids:

1. Terrorists can only attack in unprotected cities
2. The attack can occur in London, New York, Paris, or Tokyo
3. The attack will not take place in Europe
4. The city of New York is well protected

The combination of these four factoids allowed determining that the attack will occur in Tokyo.

A total of 34 factoids (Appendix B) were distributed randomly between participants and each factoid was distributed only once. Each participant needed to answer different questions, which was dependent on the position in the sub-team (Figure 7.1). Participant 2 of sub-team A (Figure 7.1), for example, needed to provide answers to the '*when*' and '*where*' questions while its co-worker (participant 1) needed to provide an answer to the question related to the '*what*' question in addition to serving as the communication hub between the sub-teams. A likewise distribution existed for sub-team B (Figure 7.1). Participants received a cheat sheet explaining which participant needed to answer which question. Because the factoids were randomly distributed and several factoids were required to remove the ambiguity, the participants were forced to share their factoids with their (distributed) team-members using the communication channel as shown in Figure 7.1. Participants received the factoids via an e-mail messaging system. Using the content of the factoid and the cheat sheet, the participant could forward the factoids to the appropriate participant. Participants were instructed not to forward everything to everybody because this would induce an information overload situation severely impacting performance.

While participants were responsible to correctly answer one or two of the questions related the terrorist attack, each sub-team was required to submit answers to all four questions using a paper-based answer form.

Table 7.1 – the four questions that need to be answered to correctly predict a terrorist attack

- Where will the attack be?
- What will be the target of the attack?
- When will the attack be?
- Who is behind the attack?

The puzzle task entailed solving a number Sudoku puzzles. A Sudoku puzzle is an old Chinese game with 9 blocks of 3x3 squares. These squares had to be filled with the numbers 1 to 9, in such way that each row, each column, and each block contains all these numbers only once. At the start of the game, some squares were already filled with numbers. The amount of empty squares could be seen as a complexity measure of the puzzle.

At the start of the puzzle task, the experiment leader distributed to each sub-team 12 puzzles and all these puzzles needed to be solved before a new information-processing phase would commence. The puzzles were presented on paper and the sub-team was told to solve puzzles individually. This meant that each participant grabbed each time a puzzle from the pile and when the puzzle was finished the participant needed to slip the puzzle under the door. The experiment leader would collect the finished puzzles.

Participants could engage in back-up behaviour by taking over puzzles from the other sub-team. Participants were instructed that in case they wanted to take over work, they should ask the other sub-team to offer them one or more puzzles. The transfer was performed by the recipient of the backing up behavior by walking to the room of their collaborators and passing the puzzles.

The team benefited from backing-up behaviour because the next information-processing step would commence only when all puzzles were finished. The earlier the next phase would commence, the earlier the overall task was accomplished thereby benefitting the temporal component of performance.

Participants were instructed to work as fast as possible which was encouraged by the bonus for the best performing team (performance is defined in terms of the number of right answers and the overall task time).

The task sequence (Figure 7.2) started with the information-processing task in which each participant received three factoids (a total of 12 factoids were distributed). This phase lasted five minutes allowing the participants to process and distribute facts (phase 1 in Figure 7.2). Recall that each factoid was distributed only once requiring participant to share and distribute the factoids. In the subsequent phase each sub-team received 12 Sudoku puzzles (on paper) making a total of 24 puzzles to be solved. The complexity of puzzles was manipulated between sub-team to increase to chance to engage in backing-up behaviour. The number of empty squares controlled the complexity of the puzzles (30 versus 35 for respectively easy and complex puzzles). In this phase sub-team A was provided with twelve easy puzzles and sub-team B with twelve difficult puzzles (sub-teams were unaware of the complexity differences). Recall that engaging in backing-up behaviour leads to solving the puzzles more quickly making a positive contribution the team performance. The next phase of the information-processing task would commence only after all the puzzles were solved. In phase 3 each participant received three new factoids that needed to be comprehended and distributed within five minutes. After these five minutes, a new puzzle phase initiated in which the difficulty of the puzzles was reversed between the sub-teams (see Figure 7.2). The last phase (phase 5) commenced after all the puzzles were completed. In this phase the remaining factoids (10) were distributed randomly among participant. The participants had another five minutes to provide per sub-team the answers to the four questions. Recall that participants were responsible for one or two of the questions related the terrorist attack but each sub-team required to submit answers to all four questions.

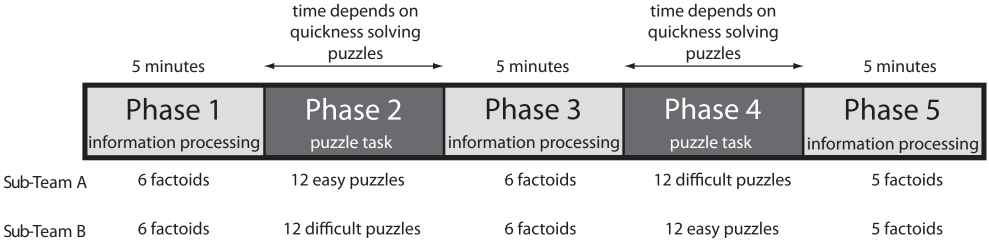


Figure 7.2 – The task required executing the information processing task and the puzzle task sequentially. The information-processing phase after a puzzle phase would only commence when all puzzles were finished. Engaging in backing-up behavior in the puzzle phase would safe time because one sub-team had more difficult puzzles to solve presumably requiring more time.

7.2.5 Display

Chapter five claimed that performance, behavior, intention, task progression, and condition information of remote team members should be displayed on an observability display. Chapter five also hypothesized that an observability display allowed a team to operate more effectively (i.e. how a team achieves its performance) and become more resilient. Figure 7.3 shows the observability display that was used in this experiment. The sub-team (A and B) and team-member (1-4) identifier was used to organize this information allowing participants to easily link observability information to participants. Information on remote team-members or sub-team could be found in the row adjacent to the appropriate identifier. Moreover, each tasks was represented in a different color block. The information-processing task was represented in dark grey and the puzzle task was represented in light grey. For example, the puzzle block in Figure 7.3 is larger in comparison with the information-processing task that lasted five minutes.

The black circles in the information-processing task represent the e-mails (i.e., factoids) received. A circle with a white dot represents an unprocessed received e-mail containing one factoid. A team-member could open the e-mail/factoid only by selecting the e-mail/factoid and pressing an 'open e-mail' button. Whenever a participant opened the e-mail, the white dot would be replaced by a black dot signifying the processing of the factoid. During the training it was stressed that the opening of an e-mail indicated reading and processing the factoid. The amount of open circles served as an indicator of the progression on the information-processing task whereas the division of closed and open circles served as an indicator of mental workload (cf. condition observability information). On the other hand, puzzles were represented by a black square where the time spent on the puzzle was displayed. At the start of a new puzzle, a square appeared with the time spent on the puzzle. For the duration of the puzzle, the block would grow with the time spent on the puzzle and the time gauge communicating the time spent would also be updated. The number represented at the end of each sub-team (in the orange circle) indicated the total amount of solved puzzles per sub-team. As soon as a puzzle was solved, the time and the expanding of the square would stop and the number of solved puzzles would increase by one. In this way, the observability display showed the progression on the puzzle task and of the performance by viewing the times and the size of the squares. The experiment leader controlled the start and stopping of the puzzles using a camera feed to determine if and when participants started and stopped with a puzzle.

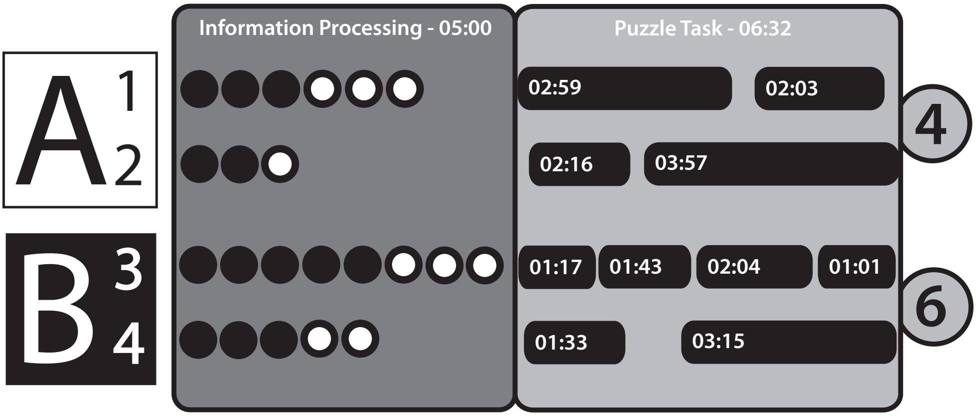


Figure 7.3 – The observability display after two phases during the experiment. The left side shows identifiers organizing information on the sub-teams (i.e. A and B) and the team-members (1-4). The information-processing task was represented in dark grey and the puzzle task was represented in light grey. Within the information-processing task, a black dot represents a factoid that was read and a white dot a factoid that is present in mailbox of the participant. The puzzle task shows black progression bars showing when a participant started a puzzle task and how much time is spend on a puzzle. The numbers at the right shows the total of puzzles that were finished per sub-team.

The observability display in Figure 7.3 doesn't show intent and behavioral information, as both were hard to retrieve given the task used in the experiment. As the task involved mainly cognitive work no physical information was communicated either by the display.

7.2.6 Apparatus

Each sub-team was placed in a separate room containing two desks with traditional desktop computers and a 190 by 120 centimeter screen for showing the observability display to the teams within the display condition. The desks and screen were set up such that the two participants in the sub-team could see the observability display without visual obstruction. Both rooms were setup in a similar fashion. A third room was used as a control room, from which the two other rooms were monitored with video cameras. The leader of the experiment also had access to a computer that controlled both the display and the computers of the participants. The factoids required for the information processing task were distributed using the Planning Task for Teams (PLATT) environment (Kamphuis, Essens, Houttuin, & Gaillard, 2010). In this experiment the participants could e-mail and chat with each other using the communication line represented in Figure 7.1.

7.2.7 Measures

Six dependent variables were measured (Table 6.1 provides an overview). First, self reported *observability* was measured directly after the experiment using a questionnaire. The questionnaire contained nine statements such as 'During the task I had a good overview of the activities of the other team' and 'During the experiment I could estimate how busy the other team was'. Their answer could differ from 'strongly agree' to 'strongly disagree' on a 7-point Likert-type scale. The average over the statements was calculated and used as a measure of observability. Secondly, the *performance* was measured in two dimensions. The total score and the total required time measured defined performance. Each correct answer to one of the four questions about the terrorist attack yielded three points maximizing score to 12 points per sub-team and 24 for the whole team. The total required time is calculated from the start of the task until the sub-teams handed in their answers to the questions and fluctuated with the time required to finish all the puzzles (see Figure 7.2). Third, the *backing-up behavior* was measured by counting both the absolute amount of transferred puzzles and the frequency of transfers. The absolute amount of puzzles is defined as the number of puzzles that was transferred in each trial while the frequency refers to how often sub-teams engage in backing-up behavior. This difference makes sure that it is recorded in case more puzzles are exchanged by one backing-up instance. It could thus be that the sub-team that was offered help said they needed help solving more than one puzzle. Fourth, the *perceived backing-up behavior* was measured using a questionnaire as defined by Smith-Jentsch, Kraiger, Collins, Cannon-Bowers, and Salas (2009). The questionnaire was translated into Dutch. Participants had to respond to three items ('the other team was willing to ask for help', 'the other team was willing to accept help', 'the other team was willing to accept feedback on his or her performance') on a 7-point Likert-type scale ranging from '*strongly disagree'* to '*strongly agree'*. The three items were averaged and served as a measure of perceived backing-up behavior. Recall that participants were instructed that in case they wanted to take over work, they should ask them to offer them one or more puzzles. Fifth, the *communication* was measured by counting both the amount of transferred messages and the number of words used in each message. A message is defined as the content within one chat or e-mail message. The message 'Please takeover two puzzles' communicated by chat counts as one message and four words. Sixth, the *mental effort* was measured by using the Rating Scale Mental Effort (Zijlstra, 1993). This is a one-dimensional scale with ratings from 0 till 150 on which participants responded to the question how much mental effort the task had cost them.

Table 7.2 – Overview of the six dependent variables and how these were measured

Number	Dependent Variable	How
1	Self Reported Observability	Questionnaire
2	Performance	Total Score Total Time
3	Backing-up Behavior	Absolute number of Transferred Puzzles Frequency of Transfers
4	Perceived Backing-up behavior	Questionnaire
5	Communication	Amount of Words Amount of Messages

7.3 Results

The data were aggregated to the team level by averaging individual scores to team scores. A 1-tailored t-test was used to compare dependent variables between conditions given that the hypotheses were directional and a between-subject experimental design was applied. An alpha level of .05 was adapted for the analyses. Cohen's *d* effect size has been reported for all t-tests. The reported *d*-value is considered large when *d* >= .8, medium when *d* is around 0.5 and small when *d* is between 0.2 and 0.3 (Cohen, 1988). Table 1 and Table 2 provide an overview of respectively the t-tests outcomes and the mean values and standard deviations.

One of the twenty teams had been excluded from the analysis because the experiment leader observed that one of the participants of that team was engaged in other tasks and this was afterwards confirmed by video analysis. This team was therefore removed from the dataset and the analysis was performed on the remaining 19 teams (76 participants). Two participants lacked to fill in the RSME scale, which have been replaced by the mean of the individual scores of the other 74 participants.

7.3.1 Observability

The results revealed a significant difference in observability ($t(17) = 1.85$, $p = .041$) with a large effect ($d = 0.89$). The participants using the observability display reported higher observability ($M = 4.23$, $SD = .86$) in comparison with the non-display condition ($M = 3.50$, $SD = 0.84$). Given that the observability display showed a higher observability, it must be concluded that the observability display indeed benefitted observability and that the results can be further analyzed.

7.3.2 Performance

Performance was measured in two dimensions (Table 6.1), namely the total score and the temporal dimension. No statistical significant effect was found on the performance score (*t*(17) = .12, *p* = .451) having a small effect (*d* = .05). Likewise, the performance in terms of total time did not show a statistical significant effect either (*t*(17) = .67, *p* = .257, *d* = .32) between the team using the observability display (*M* = 81.40, *SD* = 9.03) and those in the control condition (*M* = 78.56, *SD* = 9.54).

7.3.3 Backing-up behavior

A statistical significant effect was found for the *frequency* with which the teams engaged in backing-up behavior (*t*(17) = 3.59, *p* = .001) with a large effect (*d* = 1.74). The teams in the observability condition more frequently took over puzzles (*M* = 2.50, *SD* = 0.85) in comparison to teams without the observability display (*M* = 1.33, *SD* = 0.50). The absolute *amount* of puzzles exchanged between sub-teams did not reach statistical significance (*t*(17) = 1.50, *p* = .075, *d* = .73).

7.3.4 Perceived Backing-up behavior

The perceived backing-up behavior questionnaire contained three statements that were averaged given that Cronbach's alpha equaled 0.70. Nevertheless, the data failed to reveal a statistical significant difference on the answers of the perceived backing-up behavior questionnaire (*t*(17) = .92, *p* = .186, *d* = .45).

7.3.5 Communication

The communication took place digitally using chat or e-mail facilities and was analyzed in terms of the amount of words and the amount of messages. The amount of words in the communication between team-members differed significantly between conditions (*t*(17) = 1.74, *p* = .050) and this effect is large (*d* = .84). The teams with access to the observability display (*M* = 1858.70, *SD* = 383.15) used fewer words than the teams within the control condition (*M* = 2277.44, *SD* = 647.35). However, no statistical significant effect was found on the amount of messages send between the team-members (*t*(17) = 1.60, *p* = .064, *d* = .78).

7.3.6 Mental Effort

A statistical significant difference was found on the RSME scale (*t*(17) = 1.86, *p* = .040) with a large effect (*d* = .90) according to Cohen. However, the reported mental effort scores are in the opposite direction ($M_{Observability}$ = 77.63, $M_{Control}$ = 68.69) of the directional hypothesis (hypothesis 4).

Table 7.3 – Result outcomes of the t-tests;

	t(17)	*p*	*d*
Observability	1.85	**.041**	.89
Performance – score	.12	.451	.05
Performance – time	.67	.257	.32
Backing-up frequency	3.59	**.001**	1.74
Backing-up amount	1.50	.075	.73
Perceived backing-up	.92	.186	.45
Communication - words	1.74	**.050**	.84
Communication - messages	1.60	.064	.78
Mental Effort	1.86	**.040**	.90

Table 7.4 – Mean values and standard deviations of dependent variables

	Observability		Control	
	M	*SD*	*M*	*SD*
Observability	4.23	0.86	3.50	0.84
Performance – score	20.50	5.08	22.22	4.60
Performance – time	81.40	9.03	78.56	9.54
Backing-up frequency	2.50	0.85	1.33	0.50
Backing-up amount	4.10	0.99	3.33	1.22
Perceived backing-up	5.46	0.46	5.25	0.53
Communication - words	1158.70	383.15	2277.44	647.35
Communication - messages	249.90	58.52	214.11	34.73
Mental Effort	77.63	11.02	68.69	9.75

7.4 Discussion

The purpose of this experiment was to validate the effects of an observability display on performance, backing-up behavior, communication, and mental effort using a distributed information processing team task. The experiment compared working with an observability display to working without an observability display. More specifically, four directional hypothesis were generated, namely that:

1. more backing-up behavior occurs with teams using the observability display,
2. the performance is higher when working with the observability display,
3. less communication takes place with teams that use the observability display, and
4. the mental effort is lower for team that use the observability display.

Before continuing to discuss the results in relation to the hypothesis, it must be noted that the observability display indeed led to higher observability. The results of the self-reported observability increased almost 21% when using the display.

Interesting results were found in terms of backing-up behavior (hypotheses 1**)**. The frequency with which teams transferred the puzzles was significantly higher among teams in the observability display condition in comparison with the team that had no access to the observability display. The observability display obviously stimulated teams to engage in backing-up behavior more often (large effect). On average, teams using the observability display engaged in backing-up behavior 2.50 times versus 1.33 times for the teams that did not use the display.

In addition, the results failed to reveal evidence that a higher absolute amount of puzzles was being transferred. This means that the observability display condition was not found to increase teams to hand over more puzzles.

McIntyre and Salas (1995) have emphasized the importance of backing-up behavior as a component of teamwork. Shifting work between collaborators is a manifestation of the coordination processes and the results imply a beneficial effect of the observability display on coordination by showing backing-up behavior more frequently.

Contrasting the quantitative measure of backing-up behavior, the perceived backing-up behavior data failed to report a difference between the observability display condition and non-observability condition. In both conditions the teams reported a value in which they mildly agree with the propositions concerning the perceived backing-up behavior. The propositions in this questionnaire relate to the willingness to ask for help, willingness to accept help, and the willingness of accept feedback on their performance. Although this questionnaire is applied successfully in other studies (Hoegl, Weinkauf, & Gemuenden, 2004), the presence of the observability display failed to show an increase of the willingness related to the three propositions.

The second hypothesis expected performance benefits when using the observability display in comparison to not using such a display. However no statistical significant differences were found on the performance measures score and time and both effects were small. Participants in the observability condition failed to perform better at the task in terms of score and time (with a small effects). Therefore, hypothesis 2 is not supported by the performance results.

The inability to find a performance increase was quite striking. With regard to the performance score it was noted that 10 out of 19 groups had the maximum score of 24 points and 14 groups had a 80% score or higher. Although most of the teams reported they found the task quite strenuous, the number of teams that maximized the score provides an indication that the task might have been too easy. This notion provided a clue that a ceiling effect appeared possibly explaining the failure to find an effect on score.

Hypothesis 3 claimed that an observability display would lower the communication. Although the teams that used the observability display was not found to perform better, the results revealed that they used fewer words to communicate in the same amount of messages. The reported effects are large. The teams reduced the words in messages by more than 49% using the same amount of messages. The higher observability in the display condition lowered the amount of explicit communication.

The fourth hypothesis relates to mental effort. The results show an effect of the display condition on the mental effort; however, the means are opposing to the hypothesis thereby rejecting hypothesis four. The aim of the observability display was to lower the cognitive effort because the observability display stimulates implicit coordination (i.e., coordination that does not require explicit communicative acts). Hence the observability display incorporates various elements and the display was designed to be clear and easy to interpret. With hindsight, it might be the case that the observability display made teams more aware about the other sub-team requiring visual attention and cognitive effort to process the extra information and that this effort is larger in relation to the reduction of workload due to communication advantages. In essence, another task is introduced that surfaces more evidently requiring cognitive effort due to task set switching costs (Neerincx, 2003). In the non-observability condition this process was more on the background.

7.5 Conclusions

Chapter five of this dissertation discussed the complications of a team working in a distributed setting and proposed that observability displays improve effectiveness and resilience while working distributed. The goal of this study was to see how backing-up behavioral effects emerge when a team uses an observability display. Even though the observability display did not affect performance, effects have been demonstrated in terms of backing-up behavior and communication. The benefits of backing-up behavior and communication provide a strong cue of the beneficial effects of observability, as both are important elements in teamwork. The reduction of communication shows the observability display make elements of the remote team accessible thereby lowering the need to communicate explicitly about these elements.

However, no evidence was found that observability displays decreased mental effort[2]. In retro-perspective, it is suggested that the balance between diminished effort due to reduction of communication is compensated by the effort required to process the information that is on the display.

The next chapter (chapter 8) describes and experiment in an urban search and rescue setting that compares working with observability displays to traditional tools. Participants have to search for victims, triage those victims, and assist with landing a helicopter. The aim of chapter 8 is to validate an observability display and whether the observability display leads to actors being at the right place at the right time.

---

[2] The averages of the mental effort seem to be in the opposite direction of the hypothesis but a double sided AONVA test revealed no significant effect.

8

VALIDATING AN OBSERVABILITY DISPLAY IN AN USAR SETTING

Abstract - The aim of this study is to validate whether observability displays lead to improved coordination and performance during distributed activities in a situation that is not always following a predefined plan. Although previous chapters showed coordination benefits when using the observability display, an increase in performance has not yet been demonstrated. In addition, observability displays claim to increase resilience, which has also not yet been demonstrated. In this experiment, participants are either assigned the role of the rescue worker or the medic. They execute USAR related tasks in pairs, both with and without the help of an observability display, in a situation where everything follows a predefined plan and in a situation where deviations occur. The results show a clear preference towards the observability display. Working with the observability display leads to an increase in performance and satisfaction while the frequency of communication and the workload is decreased. More importantly, the results indicate that participants respond effectively to deviations to predefined plans. This provides strong support of the usefulness of observability displays in distributed settings; observability displays are also valuable in terms of situational alterations.

Parts of this chapter are based on:

Greef, T.E. de, Keeris, E. & Brake, G. te. A Game-based Experimentation Environment. In Yuli Porkhovnik (Eds.), *Abstracts and papers of the 37th Annual Conference of the International Simulation & Gaming Association (ISAGA'06),* pp. 36-60, St.Petersburg, 2006.

Game-based Simulation Environment, Guido te Brake, Tjerk de Greef, Jasper Lindenberg, Jouke Rypkema, Nanja Smets, Proceedings of the 3rd International ISCRAM Conference (B. Van de Walle and M. Turoff, eds.), Newark, NJ (USA), May 2006

8.1 Introduction

Chapter 5 of this dissertation explained that observability displays allow actors to detect remote co-actors in the shared environment and comprehend what they are doing and how this impacts the actor's tasks. These displays claim to have a positive effect on the coordination process leading to performance benefits. In addition, observability displays claim to increase resilience which is valuable to many crisis organizations as work is done in a highly chaotic setting with many unstructured and unexpected events. Chapter 6 demonstrated that the complexity of the task has an influence on the display use frequency and chapter 7 showed that the frequency of backing-up behavior increases when an observability display is used.

This chapter describes an evaluation of an observability display in a distributed setting where two actors have to collaborate on an Urban Search and Rescue (USAR) task within a virtual environment. Although chapter 6 and 7 revealed coordination benefits, an increase in performance has not been demonstrated by these studies. The aim of this study is to validate the expected improved performance and coordination effects when coordinating joint activities in a situation that is not always following a predefined plan. Pairs of participants were either assigned the role of the rescue worker or the medic and they executed USAR related tasks a) with and without the help of an observability display and b) in a situation where everything follows the predefined plan and in a situation where deviations to the predefined plan occur. It is expected that working with the observability display leads to a better coordination of joint activities and a better performance. In addition, it is expected that communication is reduced, the trust level increases, and the situation awareness is improved. Moreover, it is expected that actors respond to these deviations in an appropriate way should deviations from the predefined plans occur.

The subsequent section discusses the background of this study. Section 7.3 discusses the method applied in the experiment and section 7.4 discusses the results. The final section draws the conclusions based on the results.

8.2 Background

8.2.1 Communication & Trust

Salas, Sims, and Burke (2005) describe a number of coordination mechanisms that serve as enablers to the coordination process and within this experiment it is expected to find effects on two of these mechanisms, namely communication and mutual trust.

Communication results in distributed information between team members and can easily be distorted or misunderstood. In a situation of mutual trust, team members feel safe to share information and are willing to accept and provide assistance (e.g. backing-up behavior), as team members understood that this is for the good of the team (Nelson & Cooprider, 1996).

Differences in mutual trust and communication are expected to be found. Given that the observability display aims to provide information that benefits joint activities, these informational bits do not have to be communicated. Therefore less explicit communication was expected. Likewise, effects on trust were also expected. Trust may compensate for limited information availability and as such, has a direct effect on how to interpret others' behavior (Jarvenpaa, Shaw, & Staples, 2004). The observability display aims to provide information about remote team-members helping to fill in the information gaps and possibly affecting the trust of the participants. Moreover, trust may be damaged while working traditionally because it is not transparent in regards to deviations from the work plan or how the distributed team responds to unexpected events.

8.2.2 Situation awareness

Situation awareness is defined as perception, comprehension, and projection of elements in the environment (Endsley, 1995; Wickens, 2008) that are relevant for a person's task and situation awareness. In this experiment it was measured for two reasons: first, highly supportive concepts have a tendency to decrease the processing of environmental cues taking thereby taking the operator out of the loop (Endsley & Kiris, 1995). Secondly, given the geographical base of the experiment where victims are to be found and triaged, increased accuracy can be expected when using the observability display because observability displays aim to lower the chance for communication mishaps occurring in distributed settings (Thompson & Coovert, 2006).

8.2.3 Testing in Virtual Environments

It is not feasible to test support technologies in real crisis situations. For the development of innovative technologies and extensive experimentation, other approaches have to be found. A common solution is testing the concepts or tools in exercises that are similar to crisis management. For example, the COORDINATOR tool has been evaluated in an office building where props (representing fire, obstacles, and civilians) were placed in various rooms (Wagner, Phelps, Guralnik, and VanRiper, 2004). Another alternative is simulation, which has the advantage that it is easy to create specific situations in which support can be helpful. For these reasons, we have built a simulation environment for the evaluation of the observability display. This simulation enables control over all factors that may make adaptive interfaces desirable and provides a challenging environment with characteristics similar to real crises.

The interest in game engines as platforms for serious simulation has increased dramatically over the past few years. Game engines have made great advances in user interaction and visualization at low costs, promoting advances within the simulation community (Lewis and Jacobson, 2002). The Unreal Tournament 2004 game engine (Epic Games Inc.) was used in the experiment, easily facilitating the construction of a multi-person simulation in 3D. The Unreal Tournament was used in combination with the Unified System for Automation and Robot Simulation software package (USARSim) that serves to gather data from environment and simulate robots in an Unreal Tournament environment (Lewis, Wang, & Hughes, 2007). The Unreal Tournament game engine in combination with USARSim has been used by others as well (Alnajar, Nijhuis, & Visser, 2010; Balaguer, Balakirsky, Carpin, Lewis, & Scrapper, 2008; Visser, Dijkshoorn, Veen, & Jurriaans, 2011), because it is powerful, has a large support community, and is inexpensive.

8.3 Methods

The goal of the experiment was to validate the claims that, when using an observability display, performance and communication improve both when everything goes according to plan and in a situation that deviates from predefined plans. As such, the display type and the deviation from the predefined plan were manipulated and indirect effects of the coordination process were measured and compared for differences.

The participants worked in teams of two on an Urban Search and Rescue (USAR) related task. Each participant was assigned the role of either the rescue worker or the medic. The rescue worker had to locate and report victims and as such needed to search the area and communicate about the found victim. The medic had to perform triage on the reported victims and approach the reported victims to measure their respiratory rate and systolic blood pressure. Victims could only be triaged after being found and this described a flow dependency between the activities of the two actors. In addition to these individual responsibilities, both were responsible to land a helicopter safely and consequently needed to be at a specific location at a specific time (i.e. a simultaneous constraint dependency).

More specifically, seven directional hypotheses were generated in relation to the research goal, namely:

Hypothesis 1:	The use of the observability display leads to an improved performance in comparison to traditional ways of operating.
Hypothesis 2:	The observability display has a beneficial effect on the coordination process.
Hypothesis 3:	The is a decrease in communication frequency when using the observability display in comparison with traditional operational procedures
Hypothesis 4:	The subjective mental workload is lower when using the observability display
Hypothesis 5:	Situation awareness is higher when using the observability display
Hypothesis 6:	Mutual trust is higher with the observability display
Hypothesis 7:	Individuals are more aware of deviations from a predefined plan and respond to these deviations accordingly.

8.3.1 Experimental Design

A within-subjects experimental design was used with the *display type* and *helicopter arrival time* as independent variables. The display type was manipulated either traditionally (compass display) or by using the observability display. The helicopter arrival time consisted of two conditions: the helicopter could arrive on time (onTime) or arrive off schedule (offTime). The helicopter was thus in 50% of the cases on time and in 50% of the cases off time. In the off-time condition, the helicopter arrived either 2 minutes early or 2 minutes late.

8.3.2 Setup

Participants were situated in the same room but a hardboard screen separated their experimental desks. This prevented participants from observing each other. White background noise was added in order to make communication impossible and participants were instructed to raise their hand when they wanted to communicate. Whenever a participant raised a hand, the experiment leader turned the white noise down. Each experimental desk (Figure 8.1) contained (a) a standard desktop computer running the virtual world, (b) an iPhone running either the observability display or the compass display, (c) a synchronized watch, and (d) a paper-based map of the environment on which participants could draw items on.

The virtual world was hosted from a master computer running the server of Unreal Tournament (Figure 8.1 right) and the two participant computers ran on standard desktop computers that were connected via a standard router. The desktop computers were attached to a sound system with bass boomers simulating the sound of the helicopter. Note that the sound of the helicopter became louder when in the proximity of the participant.

The iPhones running either the observability display or the compass display connected to the master server via a WiFi connection. A special piece of software was adapted from the USARSim project (Lewis et al., 2007) that allowed information extraction from the virtual world: the iPhones synchronized information regarding the tasks that participants had to execute using a wireless socket connection.

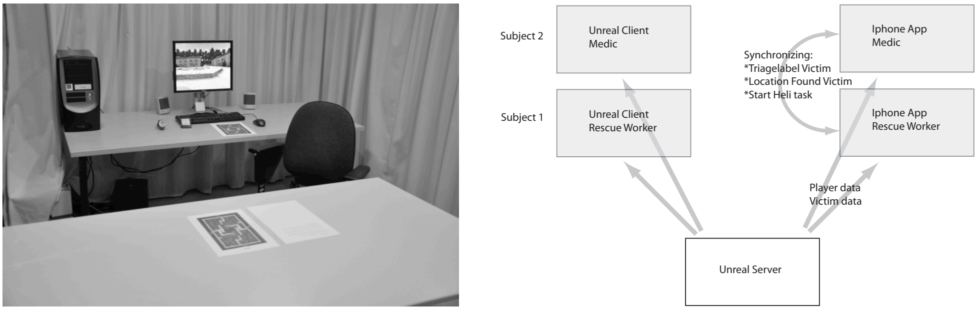


Figure 8.1 – Left: the participant desk. The display shows the virtual environment that is controlled using the mouse and/or keyboard. The iPhone with the observability display lies on top of the keyboard. Participants received a green watch (left of the keyboard) and a geographical map of the environment on which they could make notes. At a specific time, the game froze and participants were asked to turn around and answer situation awareness related questions (paper based). Right: The technical setup of this experiment shows that the two Unreal Tournament clients are connected to the Unreal Tournament server. The iPhones synchronize task-specific data and receive data from the Unreal server using the USARSim software.

8.3.3 Tasks & Roles

The participants executed the tasks in a virtual environment in which they could navigate freely using the arrow buttons, specific key combinations, and/or the mouse to move forward, backward, and sideways. The mouse could also be used to look around and rotate the field of vision. The virtual world was created in unreal tournament 2004 (multiplayer game) and is a replication of a part of a medium sized Dutch village (Smets, te Brake, Buurman, Neerincx, & van Oostendorp, 2011) (see Figure 8.2).

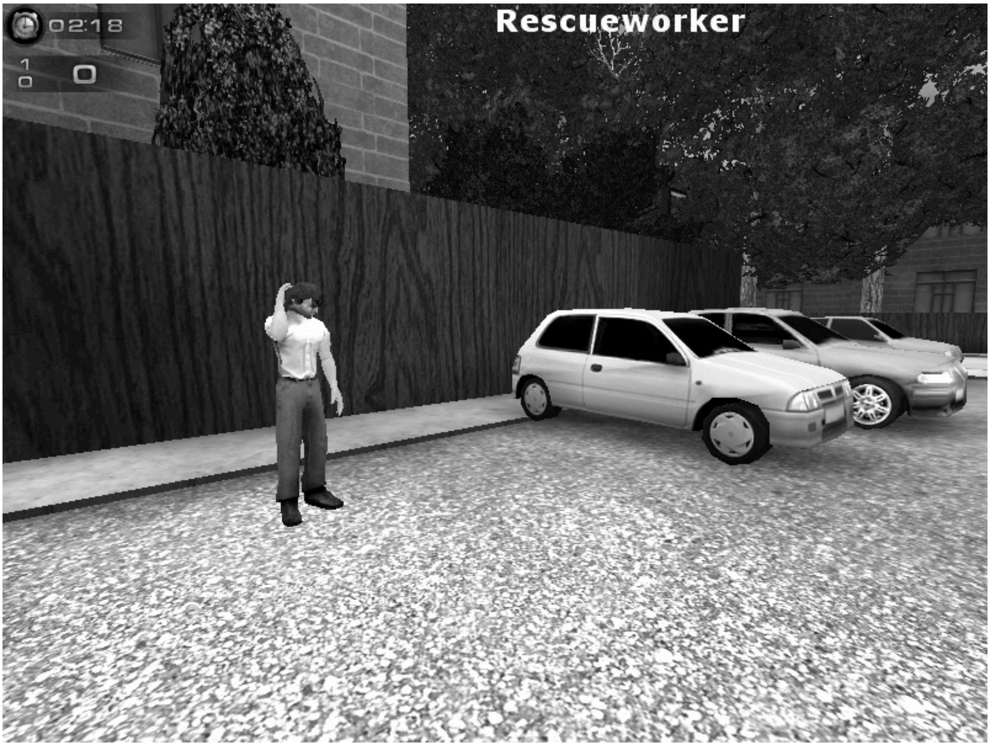


Figure 8.2 – Left: a still from the virtual world.

When a victim was located, the participant needed to report the victim by double tapping on the display (in both conditions) and confirm that a victim was found (Figure 8.3 Left). A victim would than be placed on the map and distributed correspondingly. In the non-observability condition, the participant additionally had to communicate the location of the victim and mark a V on a paper-based map. When communicating, the participants were instructed to use the compass' zone sections (e.g. north, south, south-west) possible extended with statements like "*third street down from the north-west corner*".

After a victim was reported, the medic assessed the victim's health via a triaging process. In both display conditions, the medic could approach the victim and double-tab the display (in both conditions) and confirm to activate the triage task. After double tapping, the medic was offered the systolic blood pressure and respiratory rate of the victim (Figure 8.3 Middle). A '*cheat sheet*' was available that provided a rule-based set to combine the systolic blood pressure and respiratory rate to a triage label.

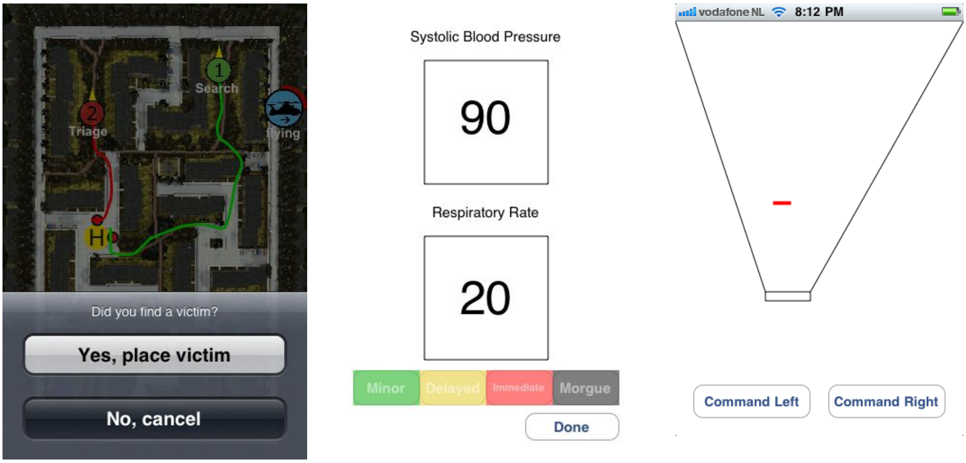
In the last task both participants had to be near the helicopter landing location, perpendicular to one another, to assist landing the helicopter. Both the medic and the rescue worker had to be present before the helicopter could land. The helicopter would hover above the landing spot in the virtual world and as soon as both players had double tapped the display (both conditions) the helicopter-landing task commenced (Figure 8.3 right). The helicopter descended in the virtual world but the participants had to provide hovering instructions to the pilot by keeping a red block within two lines. If that was being done properly, the helicopter landed safely. Note that participants were instructed to make sure that the helicopter was hovering as little as possible because the helicopter was a limited resource and should not be kept waiting.

The participants could communicate in both conditions. Participants were instructed to at least communicate the location of found victims, and were freely able to communicate otherwise. However, white background noise was added in order to simulate that the communication channel was not open. Adding the white background noise resulted in communication difficulties between participants. When participants wanted to communicate, they were instructed to raise their hand: when this occurred, the experiment leader would discontinue the noise.

8.3.4 Observability Display versus Traditional Display

The basis of the observability display was a geographical top-down view of the environment (Figure 8.4 left). Icons were placed on top of the observability display for each participant, each victim, and the helicopter. Moreover, the iconic representation of each participant showed an arrow representing the direction in which the participant was looking. A thin line, whose color matched the participant's color, represented the path the participant had walked. Found victims were represented with a light blue icon with a V and a question mark. When a victim was triaged, the color changed from light blue to a color that matched the triage label as assigned by the medic. The helicopter also had an iconic representation and was at the boundary of the display as long as it was out of range of the resolution of the map. A circular progression bar showed the progression of the helicopter.

The basis of the non-observability condition was a compass (Figure 8.4 right). The compass display showed the iconic representation of the participant in the center of the screen and the arrow corresponded to the player's field of vision.


Did you find a victim?

Systolic Blood Pressure: 90

Respiratory Rate: 20

Figure 8.3 – Left: The victims could be reported as found by double tapping the screen and acknowledging that a victim was found. Middle: The medic was required to triage the victims using the systolic blood pressure and the respiratory rate. Right: both participants were required to land the helicopter safely; this was accomplished when both participants kept the red block within the lines.

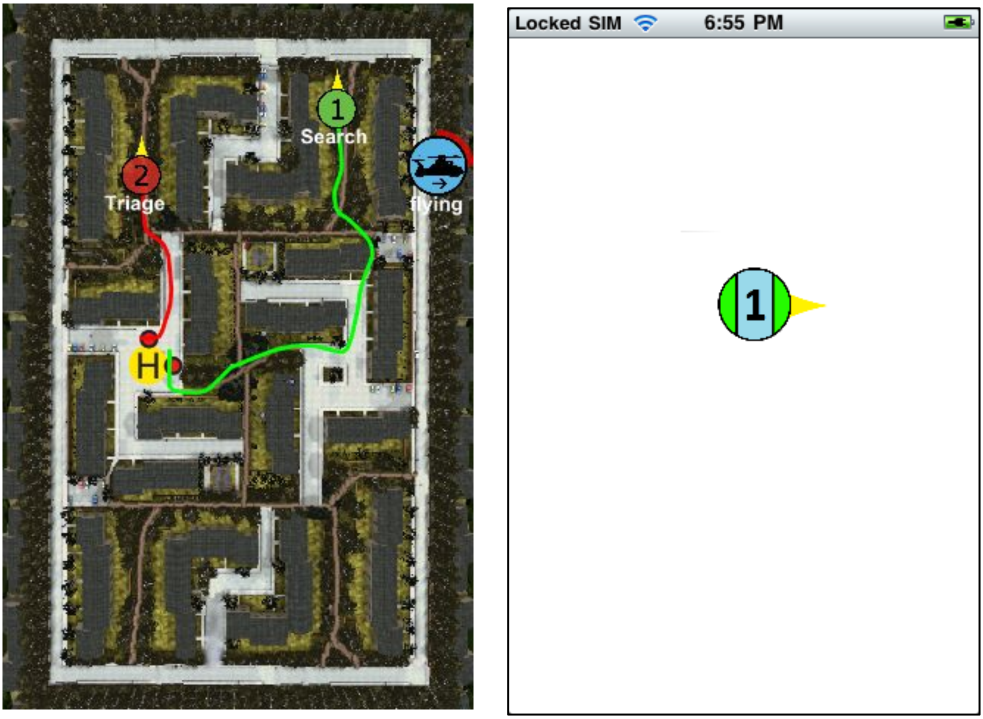


Figure 8.4 Left: the support condition showed the map of the environment and the location. Right: the compass is available when in the non-observability condition and only communicated the field of vision.

8.3.5 Scenarios

A total of four scenarios were developed as each team was required to execute four trials. The virtual world was always the same (i.e., same part of the Dutch village). Each scenario contained 16 victims and the victims were assigned to different locations with a varying health states. The area of the virtual environment was divided into 16 equal squares and the victims were randomly assigned to one of the 16 squares at a random, but remained accessible (e.g. not in an inaccessible house). Each scenario contained thus 16 victims and each scenario contained 8 minor, 4 delayed, and 4 immediate health states.

8.3.6 Participants

Thirty-two volunteers were recruited from the university campus were selected from a population of bachelor and master students with ages ranging from 19 to 34 years (M_{AGE} = 26.31, *SD* = 4.31). The participants were randomly assigned to one of two roles. Teams did not differ significantly on the background variables (education, age, team experience, gaming experience).

8.3.7 Procedure

The human research ethics committee of Delft University of Technology acknowledged that the used experimental protocol had few risks and as such approved the protocol. The experimental procedure consists of three parts. In the first part, participants were randomly assigned to a role, read the documentation, and signed the consent form. Participants were trained in the tasks and needed to complete a trial on both conditions to become familiarized with the tasks, controls, and environment. Participants were also handed over analog watches that were synchronized in their presence. In the second part, the actual four trials took place. At the start of each trial both participants were handed an envelope telling in how many minutes the helicopter was expected to land. Participants needed to use the provided watch to determine the actual landing time of the helicopter (and were allowed to acknowledge and discuss the time). Prior to opening the envelope, the experiment leader provided a clear starting signal and provided the exact starting time, telling the participants to make a note about the starting time. The game froze randomly between four and six minutes (balanced over trials) and the time was randomized to limit the change that participants could anticipate the freeze moment. When the game froze, participants were asked to turn around, sit at the table, and answer paper-based situation awareness questions. After each trial, the participants were presented with a paper-based questionnaire. The third part constituted a free short discussion allowing participants to discuss team experiences and provide remarks. At the conclusion of the test, participants were thanked for their assistance.

8.3.8 Dependent Variables

The dependent variables (Table 6.1) were either recorded at a random moment during the trial, directly after the trial using a questionnaire, or measured during the game using recorded communication or the log file from the iPhone.

First, *self reported observability* was measured directly after each trial using a manipulation check in order to validate whether the observability display actually caused increases in observability levels. In a post experimental questionnaire containing nine statements, the participants were asked to respond to the statements like 'During the task I had a good overview of where my co-worker was' and 'During tasks I was able to determine whether actions of my co-worker affected my task execution'. Their answer could differ from 'strongly agree' to 'strongly disagree' on a 7-point Likert-type scale.

Secondly, the *mental effort* was measured to determine differences while using the observability display and while using traditional methods. Mental effort was measured using the Rating Scale Mental Effort (RSME) (Zijlstra, 1993). The RSME is a one-dimensional scale with ratings from 0 to 150 on which participants had to respond to the question 'how much effort did it cost you to fulfill the task?'. The scale had nine descriptive indicators along its axis (e.g., 12 corresponds to not effortful, 58 to rather effortful, and 113 to extremely effortful).

Third, quality of the *coordination* process was measured using an adapted version of the Inter-team Coordination Questionnaire of Hoegl, Weinkauf, and Gemuenden (2004). Participants had to respond to three items ('The work done on sub-tasks was closely harmonized, 'Our team avoided duplication of error', 'Connected subtasks were well coordinate in our team') on a 7-point Likert type scale labeled from '*strongly disagree*' to '*strongly agree*'.

Fourth, *communication* was measured in two ways. Communication was first measured subjectively by asking participants in a post-trial questionnaire about the frequency ('There was frequent communication within our team') and intensity ('There was intensive communication within our team') of the communication on a 7-point Likert type scale labeled from '*strongly disagree*' to '*strongly agree*'. In addition, the communication was recorded and analyzed providing an objective measure on the frequency of communication between the two participants.

Trust was measured based on a selection of the constructs as used in the 'trust in global virtual teams' questionnaire of Jarvenpaa et al (Jarvenpaa et al., 2004). Trust is a multifaceted and complex construct and Jarvenpaa's questionnaire measures five constructs that are important to trust. Within this experiment, two constructs were measured as these were expected to vary while the other three variables were not measured. A tradeoff was made between the number of questions that participants needed answer and the expectancy of an effect. The expectancy of an effect was low for three of the five constructs while cohesion and satisfaction were expected to vary. It was therefore decided to analyze cohesion and satisfaction separately. *Cohesion* was the fifth variable and the sixth variable was the *satisfaction* construct and both were measured after each trial.

Seventh, performance was measured by analyzing the log files. Dependent on the role, the number of victims found and the number of triaged victims was a performance measure.

Eighth, the waiting time was calculated by analyzing the log files. The waiting time was defined as the waiting time of the helicopter or as the time the rescue workers were waiting. The helicopter waiting time occurred when the helicopter was hovering above the landing location while the two participants were not present. When the participants were present while the helicopter was not, the rescue workers were defined as waiting.

Ninth, a pen-and-paper based version of the situation awareness global assessment task (SAGAT) was used to measure situation awareness at a random time. Within the SAGAT method the scenario freezes and participants were asked to stop the game, turn around, and take a seat behind the table across from the table used by participants to navigate the virtual world (Figure 8.1). Participants were not allowed to look at the virtual work, iPhone device, or the watch. The questions posed were related to the participant's own location, the location of their team-member, of the victims found so far, and the triage labels of the triaged victims (medic only). After 90 seconds, whether finished or not, the participant was asked to return to their workstation and the scenario would continue. These SA questions were analyzed in four dimensions, namely the accuracy of the geographical location of themselves, the accuracy of geographical location of the team-member, the accuracy of the location of the victims, the number of correctly reported triaged victims, and the number of correctly reported victims.

Table 8.1 – Overview of the nine dependent variables and how these were measured

Nr	Dependent Variable	How
1	Observability (*self reported*)	Manipulation Check Questionnaire (5 items)
2	Mental Effort	Rating Scale Mental Effort
3	Coordination	Questionnaire (3 items)
4	Communication	Questionnaire (2 items) Frequency of communication
5	Cohesion (trust)	Questionnaire (4 items)
6	Satisfaction (trust)	Questionnaire (4 items)
7	Performance	Number of victims found Number of victims triaged
8	Waiting Time	Analyzed from iPhone log files
9	Situation Awareness	Geographical reporting of own location Geographical reporting of team-member's location Geographical location of victims Reported triage color (medic only) Correct Number of Victims

8.4 Results

An ANOVA Repeated Measures design was used for data analysis, with *display type* (observability display, compass display) and *helicopter arrival time* (onTime, offTime) as within-subject variables. All dependent variables were analyzed at the team level (individual scores were averaged).

Main effects of display type, helicopter arrival time, and the interaction effect were examined. An alpha level of .05 was adapted for the analyses. Post-hoc multiple comparisons were done using Tukey's test. Table 8.3 and Table 8.4 show the mean values and standard deviation of the dependent variables and a summary of the repeated measures ANOVA, respectively.

One of the 16 teams was excluded from the data analysis because an administrative error occurred leading to the incorrect application of conditions. Therefore, the analysis was conducted with the remaining 15 teams where conditions were applied correctly. Also, a number of missing values in situation awareness measures were detected. A total of five missing values were found for the location of victims, three for the location of oneself, five for the location of the other team member, and four times there were no colors applied to triaged victims. All these values have been replaced by the mean of the individual scores.

Five dependent variables were measured with a questionnaire that was applied directly after each trial and Table 8.2 reports the results of the reliability test per question and per trial (analyzed at the individual level). The reported Crohnbach's Alpha for the coordination (3 items), communication (2 items), cohesion (4 items), and satisfaction (4 items) approximate 0.7, which is a generally accepted reliability value. The reliability of the observability questions (5 items) was low in two occasions. The reliability analysis revealed that Crohnbach's alpha increased to acceptable levels when item 4 was deleted from the questionnaire. Consequently, item 4 of the observability questionnaire was deleted and the observability questionnaire was analyzed based on 4 items.

Table 8.2 - Overview of the values of Crohnbach's Alpha

	Observability Display		Traditional Display	
	OnTime	OffTime	OnTime	OffTime
Observability (5 items)	0.22	0.40	0.79	0.63
If item 4 deleted	0.66	0.74	0.90	0.81
Coordination (3 items)	0.66	0.69	0.82	0.81
Communication (2 items)	0.89	0.80	0.80	0.82
Cohesion (4 items)	0.78	0.86	0.79	0.89
Satisfaction (4 items)	0.87	0.87	0.86	0.93

8.4.1 Self-Reported Observability

Repeated-measures ANOVA revealed a significant effect of display type on self-reported observability (using a post-trial questionnaire) $F(1,14) = 29.26, p < .01$. The self-reported observability was higher ($M = 6.15$, $SD = 0.72$) using the observability display in comparison to the compass display ($M = 4.40$, $SD = 1.48$). Self-reported observability was not found to differ significantly by the helicopter arrival time ($F(1,14) = 1.24, p = .08$) nor did the results reveal a significant interaction effect of display type and helicopter arrival time ($F(1,14) = 1.73, p = .11$).

8.4.2 Mental Effort

Repeated-measures ANOVA showed a significant effect of the display type on the rating scale mental effort (RSME), $F(1,14) = 35.28$, $p < .01$. The mental effort was higher using the compass display ($M = 72.95$, $SD = 18.17$) in comparison to the observability display ($M = 52.18$, $SD = 18.45$). The RSME revealed no effect of the helicopter arrival time independent variable ($F(1,14) = 0.04$, $p = .85$) nor did the results reveal a significant interaction effect of display type and helicopter arrival time ($F(1,14) = 0.37$, $p = .55$).

8.4.3 Coordination

The results revealed a significant effect of display type on the 3-item coordination questionnaire, $F(1,14) = 19.60$, $p < .01$. The coordination score was higher using the observability display ($M = 4.11$, $SD = 0.54$) in comparison to the compass display ($M = 3.34$, $SD = 0.84$). The coordination variable revealed no effect of the helicopter arrival time ($F(1,14) = 1.34$, $p = .27$) nor did the results reveal a significant interaction effect of display type and helicopter arrival time ($F(1,14) = 0.02$, $p = .90$).

8.5 Communication

Communication was measured using a questionnaire and by measuring the frequency of communication. The data from the questionnaire uncovers a significant effect of display type ($F(1,14) = 95{,}43$, $p < .01$) in favor of the observability display ($M = 2.28$, $SD = 1.04$) in comparison to the compass display ($M = 3.75$, $SD = 0.83$). No significant effects were found for the helicopter arrival time on the communication questionnaire ($F(1,14) = 0.14$, $p = .72$) nor for the interaction between the display type and helicopter arrival time ($F(1,14) = 1.44$, $p = .25$). The frequency of communication revealed a significant effect of display type ($F(1,14) = 42.88$, $p < .01$). The average frequency of communication between the medic and the rescue worker was decreased when using the observability display ($M = 6.30$, $SD = 3.72$) in comparison to the compass display ($M = 13.23$, $SD = 4.52$). No significant effects were found for the helicopter arrival time on communication frequency ($F(1,14) = 0.62$, $p = .45$) nor for the interaction between the display type and helicopter arrival time ($F(1,14) = 0.92$, $p = .34$).

8.5.1 Trust

Trust was measured using two constructs, namely cohesion and satisfaction. The data revealed no effect of display type on cohesion ($F(1,14) = 0.40$, $p = .54$), or of helicopter arrival time ($F(1,14) = 1.02$, $p = .33$) nor was there an interaction ($F(1,14) = 1.02$, $p = .33$). The data on satisfaction showed a significant effect of display type ($F(1,14) = 11.96$, $p < .01$). The reported satisfaction was higher for the observability display ($M = 5.98$, $SD = 0.78$) in comparison to the compass display ($M = 4.96$, $SD = 1.20$). The helicopter arrival time revealed no effect on satisfaction ($F(1,14) = 0.89$, $p = .36$) nor on the interaction ($F(1,14) = 0.89$, $p = .36$).

8.5.2 Performance

Performance was measured using two metrics. First, the number of found victims was significantly higher ($F(1,14) = 10.73$, $p < .01$) using the observability display ($M = 8.70$, $SD = 2.55$) in comparison to the compass display ($M = 6.83$, $SD = 2.32$). Moreover, the helicopter arrival time revealed a significant difference ($F(1,14) = 20.32$, $p < .01$) on the number of victims found between the off-time condition ($M = 8.50$, $SD = 2.40$) and the on-time condition ($M = 7.03$, $SD = 2.61$). No interaction effect was found for the number of found victims ($F(1,14) = 0.76$, $p = .40$). Secondly, the number of triaged victims followed a similar pattern, as the data revealed a significant effect of display on the number of found victims ($F(1,14) = 5.37$, $p < .05$) and helicopter arrival time ($F(1,14) = 11.81$, $p < .01$) but no interaction effect ($F(1,14) = 1.08$, $p = .32$). More victims were triaged in the observability condition ($M = 9.67$, $SD = 2.78$) compared with the compass condition ($M = 7.93$, $SD = 3.08$) and less victims were triaged in the on-time ($M = 7.93$, $SD = 2.65$) condition in comparison to the off-time condition ($M = 9.67$, $SD = 3.20$).

8.5.3 Waiting Time

The data revealed a significant effect of display type on the waiting time ($F(1,14) = 6.86$, $p < .05$). With the observability display, the rescue workers waited on average 2.94 seconds ($SD = 55.24$) compared with the compass display where the helicopter waited on average 33.10 seconds ($SD = 64.68$). The data also revealed an interaction effect on the waiting time ($F(1,14) = 7.05$, $p < .05$). Post-hoc analysis showed that waiting time in the compass display off-time condition differs with the three other conditions (all $p < .05$). In the off-time condition, the helicopter waited an average 63.33 seconds ($SD = 62.27$) when participants used the compass display. This off-time compass condition differed significantly with the rescue worker waiting time ($M = 9.93$, $SD = 67.37$) when the observability display was used (Figure 8.5). In the on-time condition, the waiting times did not differ between the observability ($M = 4.07$, $SD = 40.94$) and compass display ($M = 2.87$, $SD = 53.21$) (Figure 8.5). The off-time compass display waiting time differed significantly with both waiting times on the on-time condition (Figure 8.5).

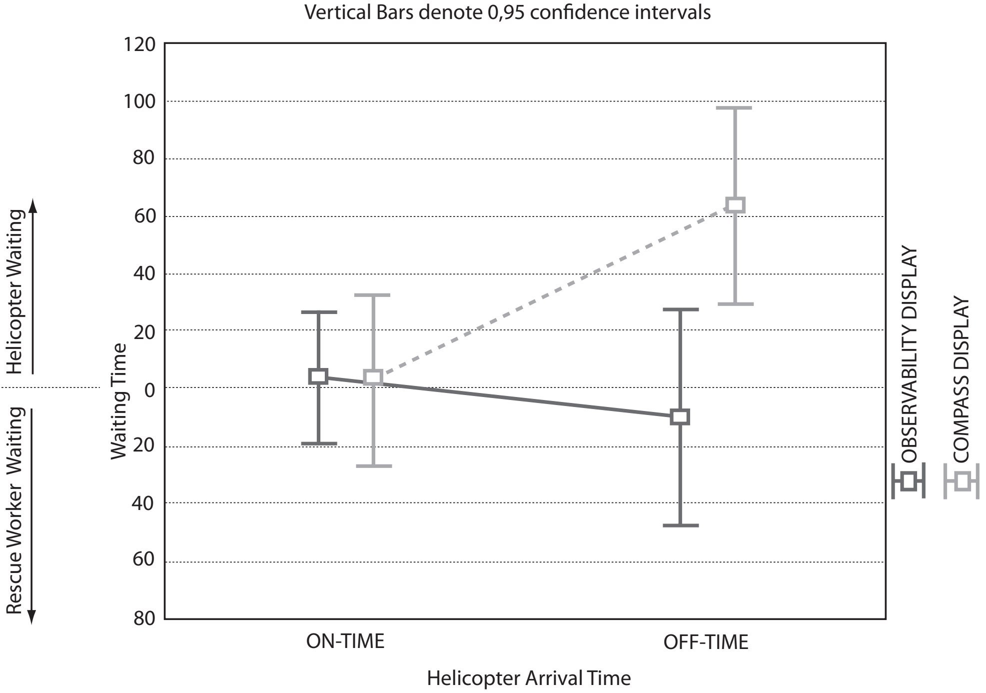


Figure 8.5 – The data revealed an interaction effect on the waiting time. Tukey's post-hoc analysis revealed that the observability display (dark grey line) in the off-time condition differs from all three other points (p < .05). The vertical bars denote the 0.95 confidence intervals.

8.5.4 Situation Awareness

Situation awareness was measured using five different constructs (Table 6.1). The data revealed no effect for display type (*F*(1,14) = 0.01, *p* = .91), helicopter arrival time (*F*(1,14) = 0.97, *p* = .33), or the interaction (*F*(1,14) = .06, *p* =*0.81*) on the geographical reported location of oneself. Likewise, no effects was found of the display type (*F*(1,14) = 0.73, *p* = .40), helicopter arrival time (*F*(1,14) = 0.95, *p* = .34), or the interaction (*F*(1,14) = 0.33, *p* =*.57*) on the of the team member's geographical reported location. The data showed also no effect of the display type (*F*(1,14) = 0.75, *p* = .39), helicopter arrival time (*F*(1,14) = 2.47, *p* = .13), or the interaction (*F*(1,14) = 0.01, *p* =.93) on the reported geographical location of the victims. The medic was asked to report the triage labels of the victims but no effect was revealed of the display type (*F*(1,14) = 0.61, *p* = .45), helicopter arrival time (*F*(1,14) = 4.41, *p* = .054), or the interaction (*F*(1,14) = 1.40, *p* =.26) on the color label of the victims. However, the data did reveal a significant effect of display type on the number of correctly reported victims ($F(1,14) = 4.19$, $p < .05$). The number of correctly reported victims was lower in the observability display ($M = 81.50$, $SD = 27.90$) compared with the compass display ($M = 90.70$, $SD = 22.11$). The data failed to show significant results of the helicopter arrival time ($F(1,14) = 2.86$, $p = .10$) or the interaction ($F(1,14) = 1.04$, $p = .32$).

Table 8.3 - Results from the repeated measures ANOVA

		Display Type			Helicopter Arrival Time			Display Type x Arrival Time		
		F(1,14)	*p*	η_p^2	*F(1,14)*	*p*	η_p^2	*F(1,14)*	*p*	η_p^2
Observability	Questionnaire	29.26	**<0.01 ****	0.68	1.24	0.29	0.08	1.73	0.21	0.11
Mental Effort	RSME	35.28	**<0.01 ****	0.72	0.04	0.85	0.00	0.37	0.55	0.03
Coordination	Questionnaire	19.60	**<0.01 ****	0.58	1.34	0.27	0.09	0.02	0.90	0.00
Communication	Questionnaire	95.43	**<0.01 ****	0.87	0.14	0.72	0.01	1.44	0.25	0.09
	Frequency	42.88	**<0.01** **	0.75	0.62	0.45	0.04	0.92	0.34	0.14
Trust	Cohesion	0.40	0.54	0.03	1.02	0.33	0.07	1.02	0.33	0.68
	Satisfaction	11.96	**<0.01 ****	0.46	0.89	0.36	0.06	0.89	0.36	0.06
Performance	Victims Found	10.73	**0.01 ****	0.43	20.32	**<0.01 ****	0.59	0.76	0.40	0.05
	Victims Triaged	5.37	**0.04 ***	0.28	11.81	**<0.01 ****	0.46	1.08	0.32	0.07
Waiting Time		6.86	**0.02 ***	0.33	3.36	0.09	0.19	7.05	**0.02 ***	0.34
Situation Awareness	Own Location	0.01	0.91	0.00	0.97	0.33	0.03	0.06	0.81	0.00
	Others Location	0.73	0.40	0.03	0.95	0.34	0.03	0.33	0.57	0.01
	Victims Location	0.75	0.39	0.03	2.47	0.13	0.08	0.01	0.93	0.00
	Triage Color	0.61	0.45	0.04	4.41	0.05	0.24	1.40	0.26	0.09
	Number of Victims	4.19	**0.05 ***	0.13	2.86	0.10	0.09	1.04	0.32	0.04

Table 8.4 - Mean and standard deviations of the dependent variables

		Display Type				Arrival Time			
		Observability Display		Compass Display		On Time		Off Time	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Observability	Questionnaire	6.15	0.72	4.40	1.48	5.18	1.62	5.36	1.27
Mental Effort	RSME	52.18	18.45	72.95	18.17	62.43	20.39	62.70	21.78
Coordination	Questionnaire	4.11	0.54	3.34	0.84	3.68	0.84	3.77	0.76
Communication	Questionnaire	2.28	1.04	3.75	0.83	2.99	1.21	3.03	1.63
	Frequency	6.30	3.72	12.23	4.52	9.43	4.93	10.10	5.88
Trust	Cohesion	5.03	0.95	4.92	0.80	4.93	0.87	5.01	36.49
	Satisfaction	5.98	0.78	4.96	1.20	5.42	1.12	5.52	1.53
Performance	Victims Found	8.70	2.55	6.83	2.32	7.03	2.61	8.50	2.40
	Victims Triaged	9.67	2.78	7.93	3.08	7.93	2.65	9.67	3.20
Waiting Time		2.93	55.24	33.10	64.68	3.47	46.65	26.70	73.83
Situation Awareness	Own Location	17.17	19.51	16.85	21.15	18.20	23.10	15.82	12.56
	Others Location	30.64	37.20	36.28	37.86	36.86	42.71	30.07	20.31
	Location of Victims	28.57	78.77	15.42	42.44	27.27	81.04	16.73	39.36
	Triage Color	57.39	30.34	63.46	33.88	83.18	34.40	89.02	28.58
	Number of Victims	81.50	27.90	90.70	22.11	83.18	29.18	89.02	42.13

8.6 Discussion & Conclusions

This chapter described an evaluation of an observability display in a setting where two actors had to collaborate on an Urban Search and Recue (USAR) task within a virtual environment. More precisely, the research objective was *to test whether observability displays lead to improved performance and coordination in situations that do not always follows a predefined plan.*

First, it was tested whether or not the observability display increased awareness about the other actors (including the helicopter) using a questionnaire. The results (self-reported) indeed show that participants had a higher level of observability when the display was used.

Hypothesis 1 was supported because the results indicated a performance increase both in terms of the number of found victims and triaged victims. When participants worked with the observability display, 27% more victims were found and 22% more victims were triaged.

Hypothesis 2 stated that the observability display had a positive effect on coordination. The results on the coordination questionnaire showed a positive effect in the coordination process. Based on these findings hypothesis 2 also seems to be supported.

Hypothesis 3 is supported because the results of the communication questionnaire and the objective communication frequency show a lowered value when using the observability display. This means that the communication had a lower frequency.

Hypothesis 4 claimed a lowered mental effort when using the observability display. The results of the rating scale mental effort (RSME) indeed showed a lower mental effort. Using the observability display, the average response matched the RSME-label 'rather much effort' while RSME-label using the compass display matched the label 'considerable effort'.

Hypothesis 5 claimed that the situation awareness is higher when the observability display is used. The situation awareness was measured using the situation awareness global assessment technique (SAGAT). Participants had to answer a number of questions at a random moment at each trial. The results were analyzed in five dimensions (Table 6.1). The results failed to show effects in four of these dimensions, namely on the correctness of the color of the triage label, the accuracy of the location of oneself, the team-member, or the victims. The results showed an effect of the number of victims that were reported in favor of the compass display. The reported number of victims was 81.50% correct in the observability display condition compared to 90.7% in the compass display. The latter finding could be explained by using the concept of 'knowledge in the head and in the world' (Norman, 1988). Using the observability display, part of the information was in the system (observability display) not requiring the human to commit this information in 'the head', leading to fewer correct answers regarding the number of triaged victims. Therefore, the results lead to reject hypothesis 5.

A note must be made to the situation awareness results. Four of the five measures failed to show either a positive or a negative effect. Because little effect was found, it is stated explicitly that the use of the observability display also did not worsen the situation awareness much neither.

Hypothesis 6 claimed an effect on trust. Trust was measured using two constructs, namely satisfaction and cohesion. The results revealed no effect on cohesion. However, the satisfaction was reported one point higher (21%) on a seven-point scale using the observability display.

The last hypothesis stated that individuals are more aware of deviation to the predefined plan and act accordingly. In this sense it was expected that participants would use the extra time when a delay occurred but still arrive in time for the delayed helicopter. Whenever the helicopter was deviated from the plan, it was expected that participants would notice and continue to find, report, and triage victims but still be on time to assist with landing the helicopter and this did occur based on the data. In general, the data showed that the actors arrived almost three seconds early when they worked with the observability display. In contrast, participants were, on average, 33 seconds late using the compass display meaning that the helicopter needed to wait and loose valuable flying time. More importantly, the data revealed an interaction effect between the display type and the arrival time of the helicopter (the arrival time determines whether the plan is followed or a deviation occurs), in favor of the observability display. The observability display allows participants to be on time (10 seconds early). When working with the compass display and deviations from the plan occurred, participants were 63 seconds late. When the helicopter arrived according to plan, there was no difference between the two displays and participants were, on average, 3.5 seconds late. This means that the observability display allowed participants to better cope with deviations from the plan in comparison to the traditional display. In addition to being aware about deviations to the plan, participants used the additional time to find and triage more victims when the helicopter was late. The performance differed between situations where the helicopter was on time (according to plan) and when the helicopter deviated in relation to the predefined plan (i.e. off-time condition). In the off-time condition, 1.5 (21%) more victims were found on average and 1.74 (22%) more victims were triaged.

This chapter described an evaluation of an observability display in a distributed setting where two actors collaborated on an USAR task within a virtual environment. The aim of this study was to validate the expected improved performance and coordination effects when coordinating joint activities in a situation that is not always following a predefined plan. Seven hypotheses were generated of which five seem to be supported by the results. One hypothesis was only partially supported partially, and one hypothesis was rejected. The results show a clear and predicted preference towards using the observability display. Working with the observability display leads to an increase in performance and satisfaction while the frequency of communication and the workload is decreased. More importantly, the results indicate that participants respond effectively to deviations from predefined plans. This provides strong evidence for the usefulness of observability displays in distributed settings and that observability displays are valuable in terms of changed situations, leading to increased resilience to unexpected events.

9

CONCLUSION

9.1 ePartners for Dynamic Task allocation and Coordination

The present dissertation follows the joint cognitive systems (Hollnagel & Woods, 2005) paradigm shift from *automation extending* human capabilities to *automation partnering with* the human. The computer should be regarded as an electronic partner (ePartner) where the human and the ePartner collaborate in a symbiotic relation to achieve the best performance while operating within safety boundaries. Stated differently, it stresses that the ePartner nor the human may be able to solve problems individually but that both are partners in solving problems effectively and efficiently. The present dissertation focuses on the design and evaluation of such ePartners that support dynamic task allocation and coordination in, possibly distributed, teams in high-risk professional domains. Consequently, the research objective reads:

Design an ePartner in high-risk professional domains that varies its authority on tasks in response to workload dynamics and supports the coordination of joint activities in distributed settings, all leading to improved joint performance.

The results presented in this thesis support the claim that an ePartner aids coordination of joint activities when actors are working in a distributed setting and benefits performance when the ePartner takes initiative to alter work divisions. In order to support these conclusions, a number of key objectives are discussed in the next section.

9.2 Meeting the Key Objectives

The first key objective was:

(1) To develop a framework capable of dividing the work between the ePartner and the human according to predetermined working agreements fitting the whole chain of information processing in a high-risk professional domain.

The first key objective was reached by implementing an object-oriented perspective (chapter 2) on top of the task oriented model currently in practice (Parasuraman et al., 2000). The addition of the object-oriented perspective allows defining object-sets per task and allows assigning a level of automation to each object-set. The addition of the object-set perspective allows making working agreements about work divisions at a fine-grained level. To summarize, the object-oriented task model provides a fine-grained way to organize and agree upon a delegation of work between an ePartner and a human.

The second key objective was:
(2) To identify adaptive automation triggering models that assess the momentary capacity of the human and the task demands upon the human.

The hybrid trigger model of chapter 3 satisfies the second objective. It combines of two models is proposed to increase robustness of the trigger moment. The operator performance model is founded on the differences in the human's worldview and the ePartner's worldview. Depending on the working agreements, differences are brought to the attention of the human using a notification. The operator performance model spots performance variances based on the *number* of differences in the worldviews and on the *average response time* to the notifications that bring the differences in worldviews to the attention of the human. On the other hand, the operator cognition model focuses on the environmental task demands and estimates the workload that is posed on the human using the validated model of cognitive task load (Neerincx, 2003). The validity of the operator cognition model was tested and it has been shown that the model predicts workload correctly.

The third key objective was:
(3) To determine the effect of the adaptive object-oriented task model on the performance and workload of navy professionals.

The third key objective was accomplished by running an experimental study with eight naval officers that used a high-fidelity simulation and workstation (chapter 4). In a within-subject design, they executed an identification task in an adaptive and static automation condition using complex smuggling and traditional scenarios. The results revealed a 60% performance improvement in favor of adaptive automation. The data showed no effect on workload. Moreover, a 65% positive performance effect in favor of adaptive automation was found for the more complex smuggling scenarios. This latter finding shows that adaptive automation provides a viable candidate to assist warfare officers in future operational situations that are predicted to become increasingly complex (Homan, 2008).

The fourth key objective was:
(4) To define which elements are important to present on an observability display that increase awareness of remote actors.

This objective was achieved with a literature review identifying a number of problems related to working in distributed settings (e.g. Powell, Piccoli, & Ives, 2004; Thompson & Coovert, 2006). In addition, an observational study at two training sessions of the Netherlands Urban Search and Rescue team revealed coordination issues while working in distributed settings (chapter 5; de Greef, Oomes, & Neerincx, 2009). The common denominator in distributed settings is the failure to directly observe actions or responses, and sense states of remote actors. The direct observations help to anticipate information processing needs and create an awareness of the weak spots in the team. Observability displays are therefore proposed displaying performance, behavior, intention, task progression, and condition information of remote actors (chapter 5).

The fifth key objective was:
(5) To understand whether coordination and the frequency of use of the observability display changes when a team gains experience and when the task is less or more complicated.

This objective was accomplished by conducting an experiment that systematically controlled team experience and task complexity for a dispersed team of three persons and measured the use of the observability display, performance outcome, and the coordination process. The results (chapter 6) revealed no effect of team experience on the dependent variables. However, the data revealed significant effects of task complexity in terms of decreased coordination and decreased performance during the complex task conditions. The decreased values serve as an extension to earlier studies that showed that more complex tasks (a) can lower the effort (Veltman & Jansen, 2004) and (b) increase the focus on task-related activities (e.g. Xiao, Hunter, Mackenzie, Jefferies, & Horst, 1996) at the cost of team-related activities such as coordination. More importantly, the use frequency of the display is significantly lower during complex tasks, which potentially indicates a reduced effect of observability displays on the coordination of joint activities in complex environments. For this reason observability displays should adapt different notification styles based on the complexity of the environment (cf. Streefkerk, Esch-

Bussemakers, & Neerincx, 2007). To summarize the results, the display use frequency depends on task complexity and no evidence was found that the display use changes with experience.

The sixth key objective was:
(6) To determine the backing-up behavior effects of using an observability display to coordinate joint activities when being separated geographically.

The sixth objective was accomplished with an empirical study that compared backing-up behavior, performance, communication, and workload when working with and without an observability display (chapter 7). A self reported measure of observability showed that participants were more aware of each other when using the display. The observability display was not found to affect performance but the study showed beneficial effects on backing-up behavior and communication. The positive results on backing-up behavior and communication support the hypothesis that observability displays have a positive effect on teamwork and coordination. However, no evidence was found that observability displays help decrease mental effort. In retro-perspective, it might well be that the diminished effort thanks to the reduced communication and the improved observability is compensated by the effort required to process the information that is on the observability display.

The seventh key objective was:
(7) To test whether observability displays lead to improved performance and coordination in situations that not always follow the predefined plan.

This objective was accomplished by an experiment in which participants were assigned either the role of a rescue worker or a medic (chapter 8). In pairs the participants executed urban search and rescue (USAR) related tasks with and without the help of an observability display in situations that developed according to a predefined plan or deviated from the plan. The results showed a clear improvement when the participants were using the observability display. Working with the observability display led to an increase in performance and satisfaction while the frequency of communication and the workload was lower. More importantly, the results indicated that participants reacted effectively to deviations from the predefined plan demonstrating the usefulness of observability displays in distributed settings in terms of awareness and decision-making in a changing situation.

9.3 Scientific Contribution

The goal of the present dissertation was to design an ePartner that supports both dynamic task allocation and the coordination of joint activities, processes that are regarded as essential in effective and efficient human machine collaboration. In order to design and evaluate the ePartner, seven key objectives were defined and the previous section described how these were achieved. The present section discusses scientific contributions of the dissertation.

Significance of Object-oriented Task Model & Working Agreements

The object-oriented framework (chapter 2) extends current task division models (Endsley, 1987; Fitts, 1951; Parasuraman et al., 2000; Sheridan & Verplank, 1978) by overcoming three shortcomings that surface when bringing the paradigm of adaptive automation to real world settings. The first shortcoming is that current adaptive automation task models are ambiguous how to divide work within a single task. Such a task division preferably matches how humans divide work (cf. Parasuraman & Miller, 2004). The object-oriented task framework therefore allows a fine-grained division of work that links closely with how humans delegate work to a colleague and matches the jargon used in the domain. Humans' work divisions follow statements like "*you take the contacts in the north*" in the military domain or "*you take all other approaching aircraft while I focus on the two non-separated aircraft*" in the air traffic control domain. Secondly, the object-oriented framework allows allocating parts of the work that are less critical in terms of severity or responsibility. It allows a gradual delegation of responsibility to the ePartner while keeping the human in firm control of those tasks and objects that are regarded important. And third, the object-oriented framework allows end-users to have the final say with respect to what levels of automation the ePartner is authorized to reach. This decision is best left to the end-user because he or she is the domain expert and can weight different aspects better than the system designer who has less experience with the domain.

To summarize, the object-oriented task model provides a fine-grained way to organize and agree upon a division of work between an ePartner and a human that matches how humans in a team tend to delegate work. The object-oriented task model is a extension to the task division models currently in use (e.g. Parasuraman et al., 2000).

Partially Validated Hybrid Triggering Model for Operational Settings

Whether or not adaptive automation is successful is to a large extend dependent on the correct timing of the adaptation, which in turn is dependent on the quality of the triggering model that is embedded in the ePartner. A hybrid trigger model (cf. Parasuraman, Mouloua, & Molloy, 1996; Wilson & Russell, 2007; de Greef & Arciszewski, 2007) is proposed to increase robustness of the trigger moment (chapter 3). The proposed hybrid trigger model can be applied to real-world operational settings and is linked to the object-oriented task model. The hybrid model consists of an operator cognition model and an operator performance model. Both are based on the object-oriented task framework, that allows creating work division that are based on the objects in the domain. The operator performance model measures performance based on differences in the world-views of the ePartner and the human. The operator cognition model focuses on the environmental task demands and assesses the workload of the human.

The performance is based on the number of and the average response time to notifications. Notifications are a result of differences between the world-views of the human and the ePartner. Task demands can be determined by counting the number of objects and by the ratio of complex objects and easy objects. The proposed hybrid approach is expected to be more robust in comparison to singular triggering models: if there are more independent indicators pointing to a high workload, the human is more likely to experience a high workload. It is expected that the hybrid approach alleviate artifacts such as a bias in the performance of the tasks.

Validation of an Adaptive ePartner in a Real-World Setting

At present, most adaptive-automation research (e.g. Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006; Hilburn, Jorna, Byrne, & Parasuraman, 1997; Kaber & Endsley, 2004; Kaber, Perry, Segall, McClernon, & Prinzel, 2006; Moray, Inagaki, & Itoh, 2000; Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2003) focused on laboratory experiments and only limited research (Hilburn et al., 1997) aimed to investigate adaptive automation in real-world settings. The validation of ePartners in such a real-world setting extends the scientific knowledge base of adaptive systems with an evaluation of a real-world adaptive task. The extended model was evaluated with eight naval officers using a high-fidelity naval command and control environment. The results showed an overall efficiency effect of 60%. Furthermore, no negative side effects of adaptive automation were found. In addition, the positive efficiency effects appeared most strongly in the more complicated asymmetrical scenarios (65%). This latter finding shows that adaptive automation can be a valuable contribution to future naval command and control systems.

Testing with Domain experts is important

Naturalistic decision-making advocates (Zsambok & Klein, 1999) that professional domain experts, such as naval officers, and naïve laboratory subjects utilize different cognitive mechanisms to cope in demanding situations. The results discussed in chapter four provide some support for the claim advanced in naturalistic decision-making. During the adaptive automation experiment (chapter 4), it was observed that the professional naval operators had a restricted focus of attention on the more important objects, in combination with an acceptance of a larger risk due to the diminished attention to the other objects. It is believed that the experts managed their subjective workload at a 'comfortable' level (mostly around level three on a five-point scale) while accepting an increased risk due to not finishing or delaying tasks. The evaluation of the adaptive ePartner provided additional support that professional naval operators can manage workload by reducing task goals, which has been demonstrated by Sperandio (1971) and discussed by Veltman & Jansen (2004). This shows that it is important to study how experts behave in their natural habitat and that validations using experts are critical to fully predict the impact of novel systems in high-risk professional domains. It must be noted that the ePartner's coordination support role was tested with naïve subjects only. Future studies are required to determine the impact of the ePartner's coordination support role on the behavior of professionals. This limitation is discussed in more detail in section 9.5.

Observability as a Design Concept

The aim of observability displays is to monitor remote co-workers in the shared environment. The perception allows comprehension by all involved actors what remote actors are doing and how this impacts the joint tasks. This requires an observability display to present the performance, behavior, intention, task progression, and state of remote actors. The use of an observability display leads to benefits in the coordination process (chapter 7, chapter 8) and an increased resilience to deviations to a plan (chapter 8). The validity of the concept of observability on coordination and resilience is demonstrated in these chapters. The positive effect on resilience shows that actors are better able to respond to deviations from a plan, a trait that is very desirable in organizations such as crisis management that frequently encounter situations that are highly dynamic and unstructured. However, it must be noted that the experiments described in chapter 7 and 8 were conducted using abstract tasks in a controlled setting to provide answers to a number of fundamental questions. This is a limitation of the study that is addressed in more detail in section 9.5. It thus remains an open question how these effects generalize to the crisis management domain. Nevertheless, the positive effects of observability displays on coordination and resilience promise to have a positive effect on activities during the management of real crises.

Also important to note is that observability takes a broader perspective to shared situation awareness in the sense that it adds to the information that is required by participants (cf. the organization-tailored situation awareness perspective, Neerincx et al, 2011). Observability displays show information that is not primary to the (shared) operational goal (cf. a goal-directed task analysis), but secondary. The observability data allows the actors in a distributed team to cope with a variety of changed conditions and unexpected events in a better way because they have additional information about their team members. A goal-directed task analysis helps to define situation awareness requirements and the analysis starts with goals that need to be satisfied. However, a goal-directed task analysis only leads to very limited information requirements related to coordination because coordination usually is a hidden process and therefore not an explicit goal of its own. This is an important notion for designers working on distributed systems: the coordination of joint activities must be included as an explicit goal when designing information distribution systems.

Workload is a delicate balance

Workload is the relationship between the cognitive available resources and the effort required to deal with the tasks (Kantowitz, 1988). Introducing support concepts is beneficial from a workload perspective as long as the effort related to using the observability display do not outweigh the benefits to the task. In comparison to traditional tools (e.g. phone calls, situation reports), it was expected that workload would be reduced using the observability display. However, the workload results in chapter 7 failed to show such an effect. Chapter 8 on the other hand did show that observability displays lead to a reduction of workload. In the latter case, the reduction of communication and costs of coordination outweighed the costs of interpreting the display. In the case of chapter 7, the advantages to the task were limited. It is thus important to system designers to analyze the advantages to coordination in relation to the costs of interpreting the display within the work practice. Designers should be aware that observability displays impose additional attention costs on the human given that the display makes the coordination process explicitly visible while the process is considered a background task when observability displays are not used.

Guidelines to structure ePartnership

Allowing the machine to partner with the human requires a number of guidelines on the partnership between the human and the ePartners from an architectural perspective. First, the notion that the ePartner uses algorithms to process information of the world leads to a notion that the machine has a view or belief in the world (similar to the beliefs in beliefs-desires-intention framework used in intelligent agents). The human also processes information leading to a human view of the world. Chapter 2 suggests storing these opinions in separate worldviews. This allows both the human and the machine to inspect and compare the two world-views. Equally important is the guideline that the ePartner should reason about the activities and state of the human, requiring the ePartner to build a model of human performance in relation to environmental conditions (chapter 3).

Additional Requirements for adaptive behavior

The results from the adaptive ePartner evaluation (chapter 4) revealed two new requirements. The first requirement states that the ePartner should show evidence when the human operator asks the ePartner for its opinion (cf. Harbers, van den Bosch, & Meyer, 2011). This means that the ePartners should explain the rationale of its decision, allowing the human operator to understand why the ePartner comes to a specific conclusion (something that increases trust); seeing a different opinion most likely increases critical thinking of the human. The second requirement expresses the fact that the ePartner must explain what it has done with objects under its supervision when it transfers objects back to the human.

9.4 Reflection for Designers & Policy Makers

Adaptive Automation Support Smaller Crew Sizes

It is often assumed that advanced technologies allow smaller crews in command centers aboard naval ships (Laurent et al., 2003), which is only partially true. Embedding advanced technological systems nearly always changes operational procedures. The role of the operator changes from monitoring and controlling a machine to supervising a machine (in case critical errors occur and the machine can no longer cope). On of the main problems with this supervisory role is that the human is no longer ‘in the loop’ and loses awareness of the situation (Endsley & Kiris, 1995). The paradigm of adaptive automation attempts to avoid this problem by letting the operator at times of low workload engage in some of the activities of the machine, thereby letting the human staying in the loop (Rouse, 1988; Scerbo, 2001). At times of overload, the human can still cope with the situation because the machine initiates a new work division that lowers the workload of the human. Empirical studies (e.g. Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006; Hilburn, Jorna, Byrne, & Parasuraman, 1997; Kaber & Endsley, 2004; Kaber, Perry, Segall, McClernon, & Prinzel, 2006; Moray, Inagaki, & Itoh, 2000; Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2003) support these claims of adaptive automation. However, these studies are mainly conducted in a laboratory setting. The results described in chapter four show that adaptive automation in real world settings is possible and matches the results of the earlier laboratory studies. An adaptive ePartner compensates for reduced manning initiatives: the human is kept in the loop at times of low workload and can cope in high workload situations because the automation takes over parts of the work.

Transferability of adaptive automation using an object-orientated task model

Chapter 2 to 4 discussed the application of adaptive automation to a real world domain and took the navy as an example. However, the object-oriented approach is transferable to other domains as well. This section discusses the possibilities to transfer the framework of adaptive automation in combination with the object-oriented task model to other domains.

The air traffic control domain seems a viable candidate to apply the concept, which is a good thing given the expected growth of the flying movements (Eurocontrol, 2010). This increased growth in flying movements will also vary the workload on air traffic controllers more and a system that helps air traffic controllers to keep their workload within a bandwidth of workload seems therefore useful. The obvious objects of interest would be the aircraft. Working agreements could be based upon the separation distance between the aircrafts and the deviation in flight plans.

The chemical plant industry and cyber security domain can serve as two other examples to demonstrate the transferability of the object-oriented task framework. Within the chemical industry, operators monitor the chemical plant in order to ensure that the chemical processes stay within safety boundaries. A plant typically consists of several chemical transfer units to transform and separate materials. The transfer units could be seen as the central objects and working agreements could be based on the attributes of these transfer units, such as temperature, pressure, viscosity, or valves settings. As another example, cyber security aims to minimize successful attacks by hackers. For this purpose, network activity is monitored to recognize patterns and determine potential attacks (cf. Bradshaw et al., 2012). The monitoring of a data network can be supported using an adaptive ePartner with an object-oriented task model. Heavy fluctuations in the network traffic lead to equally heavy operator workload fluctuations and increased support might be opportune at high workload moments. When the network traffic is low or steady, the operator can retake the work from the machine and stay in the loop. The central objects would be the network traffic packages and the activities on ports could be used as a starting point for working agreements.

A final example is virtual-reality expose therapy. During such therapies, a patient is confronted with fear-related stimuli that are presented in a virtual environment allowing the patient to slowly deal with its fear. A therapist monitors the patient and the therapist can adapt the stimuli in the virtual environment. However, this type of therapy allows a therapist to treat only one patient at a time. Ideally, a therapist could monitor and treat several patients at the same time by letting multiple patients experience their own virtual-reality exposure therapy. This requires, however, a system that is capable to manage multiple expose therapy sessions at the same time. It is important that this system has capabilities to take over certain patients in case the therapist is confronted with a situation where one or more patients require immediate assistance. The objects would be the patients and working agreements are made per patient such that the computer role is managed per patient and per session. Such an setup was studied in an exploratory study (Paping, Brinkman, & Mast, 2010).

Observability displays for USAR, OCHA, and Defense

Mission reports and field observations of the Urban Search and Rescue (USAR) domain have indicated coordination issues leading to less effective usage of human resources (chapter 5). The coordination problems are due to the distributed nature of USAR operations. Within the USAR domain, the coordination problems do not necessarily lead to hazardous situations per se, but an accumulation of events in combination with coordination problems could lead to a slow down or even a collapse of the rescue mission.

An observability display shows the (progress of) activities, behavior, intention, and states of remote actors; the visualization allows actors to determine the impact on joint activities. Empirical validation of observability displays revealed improved coordination of activities (chapter 7 and 8), better backing-up behavior (chapter 7), and an increased awareness of deviations (chapter 8). It is therefore recommended to take observability and its effects into account when designing for collaboration in distributed settings.

Within the USAR rescue organization, knowing what everybody is doing and which deviations occur are important for mission effectiveness. The Office for Coordination of Humanitarian Affairs (OCHA), which is hosted by the United Nations (UN), coordinates large-scale international rescue operations. Embedding observability displays in big scale rescue organizations allows the involved parties to better understand what everybody is doing, what their intentions are, and how well they are doing. This could improve the coordination process, especially given the level of autonomy that many of these rescue parties have. Using observability displays lead to a more efficient use of human resources because, for example, the observability display shows which teams are currently searching which buildings diminishing the chance that a rescue team from another nation will search a building that has already been searched. This approach can be extended by a collaborative system where the actors check the accuracy of the map and make corrections when needed (Gunawan, Alers, Brinkman, & Neerincx, 2011).

There is another interesting example where observability could aid distributed operations. Common operational pictures (COP) are increasingly being used in many defense organizations. A COP aims to help the commanding officers grasp the situation better which, in turn, improves the decision making process. Historically, these COP's are oriented around tracks. However, current COP's only tell part of the story. They are limited to a near real-time representation of the (aggregate of) tracks and so far lack ways of integrating these observations into a larger whole. Adding observability would significantly improve the quality of COP's. Increased awareness on the progress, state, intention, and behavior of participants in the battle space allow commanders to better determine the state of the battlefield. A battlefield is a highly dynamic and uncertain system, and knowing what the troops do, how this impacts joint activities, and how far troops are in achieving their goal is very important. At the moment of writing, initial steps are taken in this direction by crafting an information model that represents units, units' behavior, capabilities and tasks, overarching plans and possibly deviations thereof (Arciszewski & de Greef, 2011).

9.5 Limitations

It is important to provide reflections and discuss limitations that need to be taken into account while interpreting the conclusions. In relation to applied research, Schraagen explored "…*the human factors criteria that can be relaxed without compromising scientific quality* *and methodological rigor."* (Schraagen, 2010, p. 9). Given that it is hard to define exact criteria, an important lesson of his study was to properly argument constraints and choices in experimental studies. This section describes some of the limitations of this dissertation.

First, the validation of observability displays (chapter 6, 7, and 8) were conducted using naïve university students that had to do rather simple and abstract tasks. This was done because we had a number of fundamental questions at this early stage of the observability display concept. The experimental studies with abstract tasks and naïve participants lead to fundamental knowledge applicable to a wide arena of users. Although a translation of the results to the high-risk professional domains is possible and seems intuitive, a direct translation of the results to the USAR domain requires more study. Most likely, increased awareness on activities, (progress on) tasks, behavior, and states is valuable given the problems these distributed teams face, benefitting performance. However, experts show significant differences in cognitive and coping strategies. Therefore future research is required to study the effects of observability displays with USAR experts on backing-up behavior, coordination, and workload. The daily routine might influence the use frequency of the display because experts are trained to monitor the environment instead of looking at the world by means of a display.

Secondly, the adaptive automation study had a rather small group size. It is always difficult to free human resources to engage in an evaluation. This shows the other side of the medal of using domain experts: domain experts are hard to get in large numbers. A bottleneck of using domain experts in experiments is their scarcity and limited availability. The scarcity of domain experts highlights that there is often a limited number of experts available for specific tasks (e.g. astronauts, warfare offices). Due to their scarcity these domain experts usually have a very busy work schedule and therefore have limited time to participate in research activities. These reasons constrained our group to eight naval officers participating in the evaluation. And using a small number of participants can lead to results that are not representative of the domain.

Third, the study described in chapter 4 was specifically targeted at demonstrating the positive effects of adaptive automation. The validation did not specifically look at the negative effects of adaptive systems. Two potential hazards of adaptive automation are: (1) unexpected automation behavior confusing the human about the automation's operation and (2) mode awareness problems causing humans to miss a change in automation level resulting in conflicts with the automation. In an early paper, Billings & Woods (1994) condemn adaptive automation because the potential beneficial effect could be offset by unpredictable behavior leading to confusion.

9.6 Future research

The results entail a number of recommendations for future research. The object-oriented task model provides an additional object dimension to task allocation models as currently proposed (cf. Parasuraman et al., 2000). Future studies should look at the robustness of the object-oriented framework in other domains. Equally important is how users are able to make working agreements in other domains and how transparent these working agreements are in various situations. Naval experts seemed having little problem with the proposed working agreements but additional research in this direction is valuable.

In terms of triggering dynamic task allocation, it is important to continue to look for models that predict human workload challenges. The use of eye tracking devices, for example, looks like a promising way to link the visual fixation of the human eye with the objects in the object-oriented task model. Integration over time would allow us to make statements about the visual attention directed to specific objects that could improve the ePartners understanding of what the user has observed. Some exploratory work has been conducted in this area (Bosse, Lambalgen, Maanen, & Treur, 2009; Imants & de Greef, 2011).

Dynamic task allocation models risk inducing confusion and mode awareness problems. In an early paper, Billings & Woods (1994) condemn such systems because the negative effects outweigh the positive ones. It would be valuable to investigate whether the object-oriented task model does indeed suffer from these negative effects. In addition, it would be interesting to look at underload effects in combination with the object-oriented task model.

Observability displays show, among other factors, intention and task (progression). It would be interesting to look for models that automatically determine which tasks are being performed and how far they have processed in relation to a goal. Automatic inference of intention is also valuable to participating actors as this is another important factor in joint activities. In addition, it would be interesting to study expert rescue workers using observability display in a setting that combines both human actors and computerized actors. Finally, further explorations seem worthwhile to use observability displays to display the emotional state of actors that work in isolation.

9.7 Take away message

In accordance with other human-computer interaction studies, this dissertation promotes a focus shift from ***automation extending*** the human to ***automation partnering with*** the human using the ePartner analogy. This dissertation has therefore discussed two essential roles that are of importance for effective human machine cooperation, namely dynamic task allocation and coordination. First, dynamic task allocation refers to a dynamic allocation of tasks initiated by the machine in order to keep the workload of the human within a certain bandwidth. The second role deals with supporting the coordination process while working at a distance.

Successful implementation of dynamic task allocation (chapter 2 and 3) and of coordination support when working distributed (chapter 5) can be reported. Empirical results provide support to claim that the developed ePartner shows beneficial performance effects for dynamic allocation of tasks (chapter 4) and for support of the coordination of joint activities (chapter 8). In addition, indirect effects of coordination support were found in terms of backing-up behavior (chapter 7). On the downside, the observability display is used less with complex tasks (chapter 6).

SUMMARY

Highly automated systems tend to be incompatible with humans, leading to degradations in performance that may lead to serious problems in high-risk professional domains. High-risk professional domains are characterized on one side by highly varying task demands and a high density of (potentially ambiguous) information and on the other side constrained by resources. The two sides of high-risk professional domains provide an opportunity to dynamically tailor activities by the human and the automation. This dissertation focuses on ePartners that aid in dynamic task allocation and coordination of activities within high-risk processional domains.

Adaptive Automation

Adaptive automation concerns the dynamic allocation of tasks between the human and the automation, based on ePartner's proposal or decision. The overall idea of adaptive automation is that the ePartner manages the human's workload in such a way that the risks of overload and under-load are minimalized.

The literature catalogs a number of task allocation models that distinguish several levels of automation. However, these models describe task allocation strategies at a rather abstract level and fail to provide cues how to *implement* task allocation in real world settings. Chapter two therefore proposes the object-oriented task model. When the human and the machine share a task, it is much easier to *divide the objects* the task is dealing with compared to further subdividing the task itself. Using the object-oriented approach, it is possible to assign the easy objects to the automation and assign the difficult objects to the human. A navy officer, for example, makes a working agreement with the ePartner that (s)he is in charge of the objects that fly low and slow (i.e., which show ambiguous, potentially "dangerous", difficult behaviors) and that the ePartner has the authority to delegate the fast and high flying objects (i.e., which show consistent "safe" and easy behaviors) to the automation. Working agreements allow the human to define object sets using object attribute values such as speed, height, and distance. Each object set is assigned to one of five levels of automation that describe the processing and signaling behavior of the objects. Adaptive automation is managed by adjusting or adding boundary values of the relevant object attributes, e.g. the automation gets in addition to the fast and high flying objects also the authority of the low and slow flying object that are classified as large body airplanes whereas the human stays in charge of the low and slow flying objects that have a different classification.

An important element in adaptive automation concerns the question *when* to trigger adaptation. A hybrid approach combines multiple trigger methods and is believed to be more robust to the dynamic environmental conditions of future missions. Chapter three therefore proposes a hybrid trigger mechanism that combines 1) a performance model and 2) a model of the complexity of the environment. The performance model is based on earlier studies and defines performance in terms the number of signals and average response time to these signals. The model of the complexity of the environment is new, combining the number of objects with the theoretical complexity of the object. In an experiment with 18 naïve participants, the number of objects and associated complexity of those objects proved to predict the workload. The experiment thus supports the complexity of the environment model.

Chapter four investigated the effects of adaptive automation using the object-oriented task model and the hybrid triggering mechanism. Eight navy officers worked behind a prototyped combat management system using a high-fidelity naval warfare scenario. The results revealed a large performance improvement, especially in the more complex scenarios. This latter finding makes adaptive automation a likely candidate to incorporate in future combat management systems as military scenarios are expected to become complex (asymmetrical) while manning reduction initiatives stress the military system. Surprisingly, no reduction of workload was found, which is explained by the adaptive-operator effect that states that experts can adjust their task goals in order to cope with variable and highly demanding situations.

Observability

Chapter five proposes an observability display that supports the coordination process while working in a distributed setting. Actors that work in close proximity benefit from observing each other. These observations help to anticipate information processing needs and generate awareness of the weak spots in the team. The lack to directly observe actions, responses, and states of actors is a key problem in dispersed teams. In other words, observability allows actors to detect remote co-workers in the shared environment leading to comprehend what they are doing and how this impacts the joint activities.

Chapter six studied whether coordination and the frequency of use of the observability display changed when a distributed team gained experience when the task was complex or easy. Both task complexity and team experience are two factors that might influence the use of observability displays. An experiment was conducted that systematically controlled team experience and task complexity for a dispersed team of three persons and measured the use of the observability display, performance, and the coordination process. The results revealed that the display use frequency was dependent on the task complexity and no evidence was found that the frequency of display use changed when a team gains experience.

Chapter seven examined the effects of an observability display in a distributed setting on backing-up behavior and team performance using a complex information-processing task. Backing-up behavior concerns providing assistance whenever the capacity of one team member is being surpassed and is a key component of adaptive teams (making a team resilient). In an experimental between-subject design, 20 teams processed and distributed information and needed to solve Sudoku puzzles. Even though the observability display did not affect performance, beneficial effects appeared in terms of backing-up behavior and communication. Increased backing-up behavior occurred when using the observability display. The observability display presented information of the remote team, thereby lowering the need to communicate explicitly (as measured in the experiment). However, no evidence was found that observability displays decreased mental effort. Possibly, the balance between diminished effort due to reduction of communication was compensated by the required effort to process information of the observability display.

The aim of chapter eight was to test whether observability displays led to improved coordination and performance during distributed activities in a situation that is not always followed the predefined plan. A total of 17 pairs executed USAR related tasks with and without the help of an observability display in a situation where everything followed the predefined plan and in a situation where deviations to the predefined plan occurred. Working with the observability display led to an improvement of performance and satisfaction while the frequency of communication and the workload was lower. More importantly, the participants proved to respond effectively to deviations to predefined plans.

Take Away Message

In correspondence with other human-computer interaction studies, this dissertation promotes a focus shift from ***automation extending*** the human to ***automation partnering with*** the human. This dissertation shows how an electronic partner (ePartner) can act as task allocator and coordination facilitator to substantially improve the human-automation team work in high-risk professional domains (e.g. for military object identification and urban search and rescue).

SAMENVATTING

In het algemeen kan gesteld worden dat zeer geautomatiseerde systemen niet goed samenwerken met de mens. De hiermee samenhangende verlaging van de prestatie kan leiden tot serieuze problemen in hoog-risico beroepsdomeinen. Deze hoog-risico domeinen worden enerzijds gekenmerkt door uiteenlopende taakeisen en een hoge dichtheid van (potentieel ambigue) informatie en anderzijds door beperkte middelen. Deze twee kanten bieden uitstekende mogelijkheden om de activiteiten door zowel de mens als de automatisering dynamisch af te stemmen. Dit proefschrift richt zich op elektronische partners (ePartners) die ondersteuning bieden bij een dynamische taak verdeling en de coördinatie van activiteiten binnen de hoog-risico domeinen.

Adaptive Automation

Adaptieve automatisering is de dynamische toekenning van taken aan mensen en computer, geïnitieerd door de ePartner. Het idee achter adaptieve automatisering is dat de ePartner de werkbelasting van de mens zo regelt dat het risico op overbelasting en onderbelasting wordt geminimaliseerd.

De literatuur beschrijft meerdere taakallocatiemodellen waarbij verschillende niveaus van automatisering besproken worden. Deze modellen gaan echter uit van taakallocatiestrategieën op een zeer abstract niveau en bieden geen oplossing aan hoe taakallocatie te implementeren in operationele settingen. Hoofdstuk 2 beschrijft daarom het object-georiënteerde taakmodel. Wanneer de mens en de machine samenwerken is het vele malen makkelijker om de objecten binnen een taak te splitsen dan de taak verder op te delen. Op basis van het beschreven model is het mogelijk simpele objecten te delegeren aan de machine en de lastige taken objecten juist aan de mens. Een voorbeeld hiervan is een marineofficier die met de ePartner overeenkomt dat eerstgenoemde verantwoordelijk is voor objecten die laag en langzaam vliegen (die potentieel gevaarlijk zijn), terwijl de ePartner beslissingen over objecten die hoog en snel vliegen (die derhalve veilig en makkelijk gedrag vertonen) delegeert aan de machine. Werkafspraken bieden de mens de mogelijkheid om sets van objecten te definiëren via attribuutwaarden zoals snelheid, hoogte en afstand. Elke set kan toegekend worden aan één van vijf automatiseringsniveaus, die zich onderscheiden op het verwerkings- en signaleringsmechanisme. Adaptieve automatisering wordt geregeld door aanpassing of toevoeging van grenswaardes van de set van objecten. Zo kan in het genoemde voorbeeld de machine extra verantwoordelijkheid krijgen over langzame, laagvliegende objecten die worden geclassificeerd als grote vliegtuigen, terwijl de mens verantwoordelijk blijft voor de overige objecten.

Een belangrijk vraagstuk binnen de adaptieve automatisering betreft de vraag *wanneer* adaptatie moet plaatsvinden. Een hybride aanpak combineert meerdere methoden; deze aanpak wordt geacht beter om te kunnen gaan met veranderende omgevingscondities van toekomstige missies. In hoofdstuk drie wordt daarom een hybride trigger-mechanisme beschreven dat 1) een prestatiemodel en 2) een model van de complexiteit van de omgeving combineert. Dit prestatiemodel is gebaseerd op eerdere studies en beschrijft uitvoering op basis van de hoeveelheid signalen en de gemiddelde respons-tijd op deze signalen. Het model van de omgevingscomplexiteit is nieuw, waarbij het aantal objecten gecombineerd wordt met de theoretische complexiteit van het object. In een experiment met 18 naïeve deelnemers bleek dat het aantal objecten en de daarmee samenhangende complexiteit van deze objecten de werkbelasting correct voorspelden. Dit experiment bewijst de validiteit van het complexiteitsmodel van de omgeving.

In hoofdstuk vier zijn de effecten bestudeerd van adaptieve automatisering waarbij gebruik wordt gemaakt van het object-georiënteerde taakmodel en de hybride trigger methode. Acht marineofficieren werkten op een prototype waarbij gebruik werd gemaakt van waarheidsgetrouwe maritieme scenario's. De resultaten demonstreerden een verhoogde prestatie, met name in de complexere scenario's. Deze laatste bevinding laat zien dat adaptieve automatisering een serieuze kandidaat is voor de toekomstige 'command & control management systemen', aangezien verwacht wordt dat militaire scenario's steeds complexer (asymmetrisch) worden terwijl het militaire systeem onder druk komt te staan door personeelsbezuinigingen. Verrassend genoeg werd geen vermindering van de werkbelasting gevonden, wat verklaard kan worden door het adaptieve operator effect dat stelt dat experts hun doelen kunnen bijstellen om om te kunnen gaan met variabele en veeleisende situaties.

Observability

In hoofdstuk vijf wordt een 'observability display' beschreven dat het coördinatieproces ondersteunt in een gedistribueerde setting. Actoren die in elkaars directe nabijheid werken kunnen elkaar observeren. Deze observaties helpen de actoren te anticiperen op elkaars informatiebehoefte en de zwakke plekken in het team bloot te leggen. In een gedistribueerd team is één van de problemen dat actoren elkaars acties, reacties en toestanden niet direct kunnen observeren. 'Observability' zorgt er voor dat gedistribueerde actoren kunnen overzien wat ieder aan het doen is en hoe deze acties de gezamenlijke activiteiten beïnvloeden.

In hoofdstuk zes wordt onderzocht of de coördinatie en de gebruiksfrequentie van het 'observability display' veranderen op het moment dat een gedistribueerd team ervaring opdoet met het uitvoeren van een complexe taak. Zowel taakcomplexiteit als team-ervaring kunnen van invloed zijn op het gebruik van het 'observability display'. In een experiment met een gedistribueerd team van drie personen werden team-ervaring en taak complexiteit gemanipuleerd waarbij het gebruik van het 'observability display', de prestatie en het coördinatieproces werden gemeten. Resultaten laten zien dat het gebruik van het 'observability display' afhankelijk is van de taakcomplexiteit, terwijl de frequentie van het gebruik hiervan niet veranderde naarmate het team meer ervaring opdeed.

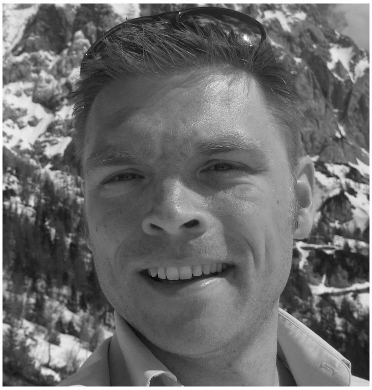
Hoofdstuk zeven bestudeert de effecten van het 'observability display' op 'backing-up' gedrag tijdens een complexe informatieverwerkingstaak. 'Backing-up' gedrag is het gedrag waarbij assistentie geboden wordt zodra de capaciteit van een teamlid te kort schiet en is een belangrijke spil in adaptieve teams; het maakt dat teams veerkrachtig worden. In een experimentele setting werd aan 20 teams gevraagd informatie te verwerken en te delen terwijl ze ondertussen Sudoku puzzels moesten oplossen. Hoewel het 'observability display' niet van invloed was op de prestatie van de taak, bleek het wel van positieve invloed op het 'backing-up' gedrag en de communicatie. Tijdens gebruik van het 'observability display' werd er meer 'backing-up' gedrag vertoont. Het toonde informatie over het andere team, waardoor een afname in de behoefte aan communicatie werd waargenomen. Er werd geen bewijs gevonden dat het 'observability display' de mentale inspanning verminderde. Mogelijk is de winst als gevolg van de verminderde inspanning door de reductie van de communicatie gecompenseerd door de vereiste mentale inspanning om informatie van het 'observability display' te verwerken.

Het doel van hoofdstuk acht is om te testen of het 'observability display' leidt tot een verbeterde coördinatie en uitvoering gedurende verschillende activiteiten in een situatie die niet altijd een voorop gezet plan volgt. In totaal moesten 17 paren USAR gerelateerde taken uitvoeren met en zonder de hulp van een 'observability display' in een situatie die volgens plan verliep en in een situatie die afweek van het vooropgezette plan. Het werken met het 'observability display' leidde tot een verbetering in de prestatie en 'satisfactie', terwijl er minder werd gecommuniceerd met een lagere werkbelasting. Het belangrijkste was dat de deelnemers zeer effectief konden omgaan met wijzigingen in het werkplan.

Take Away Message

Zoals ook blijkt uit andere mens-machine interactie studies, promoot dit proefschrift een verschuiving van *automatisering als een verlengstuk* naar *automatisering als een samenwerkingspartner*. Dit proefschrift laat ziet op welke wijze een elektronische partner (ePartner) kan fungeren als een taakallocator en een coördinatiefacalitator die het mens-machine teamwerk in hoog-risico domeinen (bijv. Militare object identificatie en Urban Search & Rescue) substanteel verbetert.

CURRICULUM VITAE



Tjerk de Greef (born on the 4th of August 1975 in Rheden) received his secondary education (V.W.O.) at the Rhedens Lyceum in Rozendaal and graduated in 1995. Afterwards, he graduated at HAN University of Applied Sciences (Hogeschool van Arnhem en Nijmegen) in 1999 and obtained his Master's degree in Information and Computer Sciences at Utrecht University in 2004. After receiving his master's degree, he was offered a position at the Netherlands organization for Applied Research (TNO), Defense, Safety and Security where he was mainly involved in human factors related research for the Royal Netherlands Navy. In 2008 he started as a PhD student at Delft University of Technology on the topic of designing ePartners for dynamic task allocation and coordination in high-risk professional domains such as Urban Search & Rescue and the Navy. From 2012, he holds a post-doc position at Delft University of Technology, in close collaboration with the Oxford Institute of Ethics, Law and Armed Conflict (University of Oxford), on the topic of *ethics and automation*. In his capacity as a researcher he has published in number international journals and conferences and managed several projects. He has also been involved in the organization of various conferences and workshops. He frequently serves as an external reviewer for international journals such as *IEEE Transactions on Systems, Man, and Cybernetics* and *Ethics and Information Technology*. Since 2010 he is the treasurer of the European Association of Cognitive Ergonomics (EACE).

ACKNOWLEDGEMENTS

The last four and a half years were an interesting journey and I am very proud to present you my dissertation. I met so many interesting people and visited so many interesting working sites that I like to pay gratitude to the Royal Netherlands Navy and the Netherlands Urban Search and Rescue (USAR) team to grant me access to their 'world'. Sailing to Spain on a Navy Vessel and exercising with the USAR team in the Dubai desert are experiences I will never forget and which are crucial to understand the work domain (which is a requirement to reason about technology).

I also would like to thank Delft University of Technology for providing me with the necessary tools and means to do research. I owe a special thanks to Mark Neericx, Stijn Oomes, Willem-Paul Brinkman, and Catholijn Jonker. Mark Neerincx and Stijn Oomes have brought me to the intelligent interaction group (formally known as the man-machine interaction group) and helped to initiate my research. Mark, thank you for supporting me through these years. I also like to thank Willem-Paul Brinkman to initiate supervising me on a daily bases after Stijn left the group. While not directly involved in the studies, the enthusiasm and support of Catholijn Jonker gave me on many occasions the mental push I needed. In addition, working in the intelligence interaction group has been fun and a privilege to work in such a multidisciplinary setting with so many different people.

A significant part of the work was funded by the 'Collaboration at a distance' IOP-MMI project (MMI-06001A/B) that was run by NL Agency (formally known as SenterNovem), an agency of the Dutch Ministry of Economic Affairs. Other parts were funded by the Netherlands Defense Research and Development organization under the 'Smart Operations' program V929 that was hosted by TNO.

I also like to thank many people, family, and friends for supporting me while I was pursuing my studies. I start by acknowledging my fellow PhD's to spend time to discus research elements helping me to shape many sentences and create models, in addition to providing moral support in difficult times: Nike Gunawan (which I collaborated closely on the Collaboration at a distance project), Jan-Willem Streefkerk, Nanja Smets, Olivier Blanson Henkemans, Marc Grootjen, and Peter-Paul van Maanen. I continued to visit TNO in Soesterberg and I would like to express gratitude to a number of TNO colleagues for providing support: Guido te Brake, Jouke Rypkema, Bert Bierman, Kees Houttuin, Jasper Lindenberg, Jurriaan van Diggelen, Karel van der Bosch, and Jan-Maarten Schraagen: thanks for you interest in my research.

Research cannot be conducted alone and therefore I specifically would like to thank Henryk Arciszewski and Rick van der Kleij. Rick, we have conducted a number of studies together and I believe that our education and experiences complement nicely. Henryk, we have collaborated since 2005 and I value our collaboration and our shared work very much. I truly believe that we serve as a nice example of a symbiosis in research as our backgrounds and attitude complement perfectly. It is such an honor that you stand by me while I defend parts the work that emerged through our collaboration. This brings me to Sander Oorthuys: Sander, it is such an honor that you are my other paranymph and are such a good friend. I am happy that we once again share a public performance after our first shared performance at the 'Dreigroschenoper' during secondary school.

Crucial to the academic work presented in this dissertation were Master students: Harmen Lafeber, Puck Imants, Lisanne Brons, Minh Po Sam, Tim Nederhoff, Maarten van Zomeren, Els Koster, Ferdinand van Oostrom, and Arvind Mohabir. Thank you for helping my research endeavors. I know some of you are still struggling and finishing, but hold on and you will get there: it's rewarding!

An important element in research is taking distance from your work; friends and family have helped taking that distance, fortunately. My close group of 'pikkemannen' have provided incredibly much support. Remco, Erik, Joop, Wouter, Jasper, Megan, Els, Lineke, Marije, and Sandy: Thank you so much for being there. I hope to continue to meet individually and as a group at a campsite, at the slopes, at a bar discussing the potential, or at a dinner party. Let's continue to make fun of each other because it keeps life understandable and keeps us on the floor!!!! And Megan, thanks for reviewing two chapters and I am glad that one of my writing slips made you smile: I will always think about a *hovering* helicopter while *hoovering* the floor. But there are more friends that supported me in my endeavor: Marc (van Rijnsoever), Martine, Marlies & Gilad, and Nienke & Jurgen: it is so nice to meet, watch children grow up, and cherish many memories.

I also like to thank the 'de Snoo' family. Henk's humor and Madeleine's food gave me the necessary fuel to continue. Femke, thanks for introducing me to kite-surfing and Eva for opening my eyes to the economic value of hospitality that reminded me that we are doing research for a purpose.

Last, I like to mention my own family for supporting me. Gerben, my older brother: I have so many good memories from playing with Lego on Sunday mornings, surfing together, visiting a beer festival in Amsterdam. These are important memories to cherish, remove negative energy and continue on my PhD journey. My father and my mother are obviously very important to me. Willie, I know that you would have been very proud seeing me defend my thesis and it is such a tragedy that you cannot see me do this. My father is very important to me and requires special attention. Ton, as years pass by I increasingly understand life and the lessons you taught me, the sacrifices you took; your the willingness to continue is sensationally inspiring. Two years ago, we took, accompanied by Felix, a trip to Slovenia/Croatia and I have so much good memories about this trip. Let's soon do this again.

Maartje, I have never understood a human-relationship any better as with you. David Sedaris writes in one of his latest books about the odd thing in a relation that it is actually quite nice to sit and be silent while still communicating and laughing. At times, it is just so nice to sit and not talk while thinking the same. At other times it is good to talk, to challenge each other, and to laugh. I find it incredible to share the passion for music, mountains, food, the sea, and kite surfing. Let's continue doing and live our lifes!

Felix, you are the sunshine in my life. You bring so much joy and happiness that life has become so much brighter. I could not have expected that and it goes without saying that I find you the most incredible creature in the world. Unfortunately for you, you will have to share this special position soon with a brother or sister.

APPENDICES

Activity Awareness Questionnaire (chapter 6)

Each trial, participants had to answer five questions concerning the team's activities. Participants never had to answer questions marked with a * about their selves, only about the two other team members.

'What is the color of the puzzle person A is solving now?' *
red / blue / yellow / this person is not solving a puzzle at this moment
'What is the color of the puzzle person A is going to solve after the current puzzle?' *

red / blue / yellow / not indicated
'How many puzzles has person A solved this block until now?' *
….
'How many yellow puzzles has person A solved this block until now?' *
….
'How many puzzles have person A, B and C solved together until now this block?'
….
'How many yellow puzzles have person A, B and C solved together until now this block?'….
'Who has solved the most puzzles this block until now?'
A / B / C
'Who will solve the most puzzles this block?'
A / B / C
'How long has person A been working on this present puzzle?' *
0-2 minutes / 2-4 minutes / 4-6 minutes / longer than 6 minutes

Factoids as used in the experiment in chapter 7

Solution The Silver group plans to attack a secular school in Omicronland on January 1 at 9:00 AM

1 The Gray and Teal groups do not employ suicide bombers

2 There will be a suicide bomber attack at a school

3 The Silver group does not work in Piland

4 The Silver group only attacks during the day

5 The Sienna and Rose groups only target the military

6 The target is either a secular school, religious school or army base in a coalition country

7 No attacks are being planned on religious organizations in Sigmaland

8 The target is in a coalition country (Muland, Xiland, Omicronland, Piland or Sigmaland)

9 An attack is being planned for the first of the month

10 The Silver group has a history of attacking domestic assets

11 The Blue, Silver, Turquoise, Gray, or Teal groups may be planning an attack

12 Bloggers are discussing the role of schools in misleading the youth in Omicronland, Piland and Sigmaland

13 The Blue and Silver groups prefer unprotected targets

14 Reports from Omicronland, Piland and Sigmaland indicate surveillance ongoing at public gathering places

15 The Blue and Silver group operatives have entered Omicronland

16 The Blue, Silver, Turquoise, and Gray groups have the capacity to operate in Muland, Xiland, Omicronland, Piland and Sigmaland

17 Attacking buildings when there are many people present increases casualties

18 There are fewer attacks in the heat of Summer (June and July)

19 The Blue group is comprised of former teachers and does not target schools

20 No attacks are being planned on religious organizations in Xiland, Omicronland and Piland

21 Muland and Xiland have only religious schools

22 An attack is being planned for the first month of the year

23 The Blue group is not involved

24 Members oft the Silver, Turquoise, Teal, Sienna and Rose groups have experience with explosives

25 Silver and Turquoise group members have entered Xiland and Omicronland

26 The Blue, Silver, Turquoise and Gray groups prefer to attack in daylight

27 The Turquoise group focuses on destroying energy infrastructure

28 No attacks are being planned on religious organizations in Muland and

Omicronland

29 Sigmaland has closed all its schools

30 There will be a suicide bomb attack at 9:00

31 Reports from the Teal group indicate standard levels of activity

32 Caches of explosives have recently been found in Muland, Xiland and Omicronland

33 The schools in the coalition area are not well defended

34 Daring day attacks make the attacker seem unstoppable

REFERENCES

Alnajar, F., Nijhuis, H., & Visser, A. (2010). Coordinated Action in a Heterogeneous Rescue Team. *Simulation*, 1-10.

Archer, R. D., & Lockett, J. F. (1997). WinCrew - a tool for analyzing performance, mental workload and functional allocation among operators. In L. Bannon & J. McCarthy (Eds.), *Proceedings of the First International Conference on Allocation and Functions (ALLFN'97)* (Vol. II, pp. 157-165). Galway, Ireland: IEA Press, Louisville.

Arciszewski, H., & de Greef, T. (2011). A Smarter Common Operational Picture: The Application of Abstraction Hierarchies to Naval Command and Control. *16th ICCRTS: Collective C2 in Multinational Civil-Military Operations.*

Arciszewski, H. F. R., de Greef, T. E., & van Delft, J. H. (2009). Adaptive Automation in a Naval Combat Management System. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, *39*(6), 1188-1199.

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors*, *48*(4), 693-709.

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

Balaguer, B., Balakirsky, S., Carpin, S., Lewis, M., & Scrapper, C. (2008). USARSim: a validated simulator for research in robotics and automation. *Search.* IEEE/RSJ.

Billings, C. E. (1997). *Aviation Automation: The search for a human-centered approach.* Mahwah, NJ: Erlbaum.

Billings, C. E., & Woods, D. D. (1994). Concerns about adaptive automation in aviation systems. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 24-29). Hillsdale, NJ: Erlbaum.

Blanson Henkemans, O. A. (2009). *ePartner for Self-Care / How to Enhance eHealth with Personal Computer Assistants.* Delft University of Technology, Delft.

Bolderheij, F. (2007). *Mission-Driven sensor management analysis, design, implementation and simulation (PhD thesis).* Delft University of Technology, Delft.

Bolstad, C. A., & Endsley, M. R. (1999). Shared mental models and shared displays: An empirical evaluation of team performance. *Human Factors*, *43*(3), 213–217. Human Factors and Ergonomics Society.

Bosse, T., Lambalgen, R. V., Maanen, P.-P v., & Treur, J. (2009). Attention Manipulation for Naval Tactical Picture Compilation. *Mind.*

Bowers, C. A., Jentsch, F., Salas, E., & Braun, C. C. (1998). Analyzing Communication Sequences for Team Training Needs Assessment. *Human Factors*, *40*(4), 672-679.

Bradshaw, J. M., Carvalho, M., Bunch, L., Eskridge, T., Feltovich, P., Johnson, M., & Kidwell, D. (2012). Sol: An Agent-Based Framework for Cyber Situation Awareness. *Kuenstliche Intelligenz*.

Burke, C. S., Stagl, K. C., Salas, E., & Pierce, L. (2006). Understanding Team Adaptation: A Conceptual Analysis and Model. *Journal of Applied Psychology*, *91*(6), 1189-1207.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology*, *42*(3), 249-268.

Cannon-Bowers, J A, & Salas, E. (1998). *Making decisions under stress: implications for individual and team training*. Washington, D.C.: American Psychological Association.

Cannon-Bowers, J A, Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan Jr. (Ed.), *Individual and group decision making: Current issues*. Hillsdale, NJ: Lawrence Erlbaum Assiciates.

Carroll, J. M., Rosson, M. B., Convertino, G., & Ganoe, C. H. (2006). Awareness and teamwork in computer-supported collaborations. *Interacting with Computers*, *18*(1), 21-46.

Clamann, M. P., Wright, M. C., & Kaber, D. B. (2002). Comparison of performance effects of adaptive automation applied to various stages of human-machine system information processing. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 342-346). Baltimore: Human Factors and Ergonomics Society, Santa Monica.

Clamann, M., & Kaber, D. (2003). Authority in adaptive automation applied to various stages of human-machine system information processing. *Proceedings of the human factors and ergonomics society 47th annual meeting* (pp. 543-547). Santa Monica, CA.

Clark, H. H. (1996). *Using language*. Cambridge, U.K.: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 115-159.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring Team Knowledge. *Human Factors*, *42*(1), 151-173.

Coram, R. (2002). *Boyd: The Fighter Pilot Who Changed the Art of War*. New York: Little, Brown & Company.

Dabbish, L., & Kraut, R. E. (2008). Awareness Displays and Social Motivation for Coordinating Communication. *Information Systems Research*, *19*(2), 221-238.

de Greef, T. E., & Arciszewski, H. F. R. (2007). A Closed-Loop Adaptive System for Command and Control. In D. D. Schmorrow & L. M. Reeves (Eds.), *Foundations of Augmented Cognition - Third International Conference, FAC 2007, Held as Part of HCI International 2007* (Vol. 4565, pp. 276-285). Beijing, China: Springer Berlin, Heidelberg.

de Greef, T. E., Arciszeski, H. F. R., & Neerincx, M. A. (2010). Adaptive Automation Based on an Object-Oriented Task Model: Implementation and Evaluation in a Realistic C2 Environment . *Journal of Cognitive Engineering and Decision Making*, *31*, 152-182.

de Greef, T. E., & Lafeber, H. (2007). Utilizing an Eye-Tracker Device for Operator Support. *4th Augmented Cognition International (ACI) Conference being held in conjunction with the HFES 51st Annual Meeting* (pp. 67-72). Baltimore, Maryland: Strategic Analysis, Inc, Arlington.

de Greef, T. E., van der Kleij, R., Brons, L., Brinkman, W. P., & Neerincx, M. A. (2011). Observability Displays in Multi-Teams. (S. M. Fiore & M. Harper-Sciarini, Eds.)*10th International Conference on Naturalistic Decision Making (NDM 2011).* Orlando, FL, U.S.A.: University of Central Florida.

de Greef, T. E., Oomes, A. H. J., & Neerincx, M. A. (2009). Distilling Support Opportunities to Improve Urban Search and Rescue Missions. In J. A. Jacko (Ed.), *13th International Conference on Human Computer Interaction, Interacting in Various Application Domains* (Vol. 4, pp. 703-712). San Diego: Springer.

de Greef, T. E., van Dongen, K., Grootjen, M., & Lindenberg, J. (2007). Augmenting Cognition: Reviewing the Symbiotic Relation Between Man and Machine. In D. D. Schmorrow & L. M. Reeves (Eds.), *Proceedings of the 3rd International Conference of Foundation of Augmented Cognition held as part of HCI International 2007* (Vol. 4565, pp. 439-448). Berlin, Heidelberg: Springer Verlag.

Endsley, M. (1987). The application of human factors to the development of expert systems for advanced cockpits. *Proceedings of the Human Factors Society 31st Annual Meeting*, 1388-1392.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Fadictors*, *37*(1), 32-64.

Endsley, M. R., Bolte, B., & Jones, D. G. (2003). *Designing for Situation Awareness.* . New York: Taylor & Francis.

Endsley, M., & Kaber, D. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462-492. Taylor & Francis Ltd.

Endsley, M., & Kiris, E. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors*, 381-394.

Eurocontrol. (2010). *Long-Term Forecast: IFR Flight Movements 2010-2030*. Brussels.

Feltovich, P. J., Bradshaw, J. M., Clancey, W. J., Johnson, M., & Bunch, L. (2008). Progress appraisal as a challenging element of coordination in human and machine joint activity. In G. M. P. O'Hare, A. Ricci, M. J. O'Grady, & O. Dikenelli (Eds.), *Lecture Notes in Computer Science; Engineering Societies in the Agents World VII*. Berlin/Heidelberg: Springer.

Fitts, P. M. (1951). *Human engineering for an effective air navigation and traffic control system.*

Flight of the drones. (2011).*Economist.*

Geddes, N D. (1985). Intent inferencing using scripts and plans. *Proceedings of the First Annual Aerospace Applications of Artificial Intelligence Conference* (pp. 160-172). Dayton, Ohio: U.S. Air Force, Washington.

Gopher, D., & Donchin, E. (1986). Workload - An examination of the concept. In K. Boff & J. Kaufmann (Eds.), *Handbook of Perception and Human Performance* (Vol. 2, pp. 4141-4949). New York: Wiley-Interscience.

Grabowski, M., & Sanborn, S. D. (2003). Human Performance and embedded intelligent technology in safety-critical systems. *International Journal of Human-Computer Studies*, *58*, 637-670.

Grootjen, M., Neerincx, M. A., & Weert, J. C. M. (2006). Task based interpretation of operator state information for adaptive support. *Proceedings of Human Factors and Ergonomic society 50th Annual Meeting* (pp. 236-242). Santa Monica, CA, USA: HFES.

Gunawan, L. T., Alers, H., Brinkman, W.-P., & Neerincx, M. A. (2011). Distributed collaborative situation-map making for disaster response. *Interacting with Computers*, *23*(4), 308-316. Elsevier B.V.

Hancock, P., Chignell, M., & Lowenthal, A. (1985). An Adaptive Human-Machine System. *Proceedings of the IEEE Conference on Systems, Man and Cybernetics* (pp. 627-629).

Harbers, M. (2011). *Explaining Agent Behavior in Virtual Training.* Utrecht University.

Harbers, M., van den Bosch, K., & Meyer, J. J. (2011). A Theoretical Framework for Explaining Agent Behavior. *SIMULTECH 2011.*

Heath, C., & Luff, P. (1992). Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms. *Computer Supportive Cooperative Work (CSCW)*, *1*, 69-94.

Hilburn, B., Jorna, P. G., Byrne, E. A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua, J. Koonce, & V. D. Hopkin (Eds.), *Human-Automation Interaction: Research and Practice* (pp. 84-91). Mahwah: Lawrence Erlbaum Associates.

Hilburn, B., Molloy, R., Wong, D., & Parasuraman, R. (1993). Operator versus computer control of adaptive automation. In R. S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 161-166). Columbus, OH: The Ohio State University.

Hoegl, M., Weinkauf, K., & Gemuenden, H. G. (2004). Interteam coordination, project commitment, and teamwork in multiteam R&D projects: A longitudinal study. *Organization Science*, *15*(1), 38-55.

Hollnagel, E., & Woods, D. D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton: Taylor and Francis.

Homan, K. (2008). Veelzijdigheid van militaire macht in veranderende veiligheidsomgeving. *Armex*, *92*(1), 10-13.

Imants, P., & de Greef, T. (2011). Using eye tracker data in air traffic control. *ECCE 2011: Proceedings of the 29th Annual European Conference on Cognitive Ergonomics* (pp. 259-260). ACM.

Inagaki, T. (2000a). Situation-adaptive autonomy for time-critical takeoff decisions. *International Journal of Modelling and Simulation*, *20*(2), 175-180. IASTED.

Inagaki, T. (2000b). Situation-adaptive autonomy: Dynamic trading of authority between human and automation. *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, "Ergonomics for the New Millennium"* (pp. 13-16). San Diego, CA: Human Factors and Ergonomics Society, Santa Monica.

Jarvenpaa, S. L., Shaw, T. R., & Staples, D. S. (2004). Toward Contextualized Theories of Trust: The Role of Trust in Global Virtual Teams. *Information Systems Research*, *15*(3), 250-267. INFORMS: Institute for Operations Research.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness. Cognitive science series.* Harvard University Press.

Kaber, D. B., & Riley, J. M. (1999). Adaptive Automation of a Dynamic Control Task Based on Secondary Task Workload Measurement. *International Journal of Cognitive Ergonomics*, *3*(3), 169-187. Lawrence Erlbaum Associates.

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human Factors*, *47*(4), 730-741.

Kaber, D., & Endsley, M. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, *5*(2), 113-153.

Kaber, D., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics*, *36*(5), 447-462.

Kamphuis, W., Essens, P. J. M. D., Houttuin, K., & Gaillard, A. W. K. (2010). PLATT: a flexible platform for experimental research on team performance in complex environments. *Behavior Research Methods*, *42*(3), 739-753.

Kanki, B. G., Folk, V. G., & Irwin, C. M. (1991). Communication variations and aircrew performance. *The International Journal of Aviation Psychology*, *1*(2), 149-1662.

Kantowitz, B. H. (1988). Defining and Measuring pilot Mental Workload. In J. Comstock (Ed.), *Mental-State Estimation* (pp. 179-188). Hampton, VA: National Aeronautics and Apace Administration, Scientific and Technical Information Division.

Klein, G. A. (2001). *Sources of power: how people make decisions*. London: the MIT Press.

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten Challenges for Marking Automation a “Team Player” in Joint Human-Agent Activity. *IEEE Intelligent Systems*, *19*(6), 91-95.

Laurent, J., Rogers, R., Randolph, J., Wren, S., M, L., Predue, C., & Hunt, J. (2003). *Navy actions needed to optimize ship crew size and reduce total ownership costs*. Washington DC, USA.

Lee, J. D. (2006). Human Factors and Ergonomics in Automation Design. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1571-1596). Hoboken, NJ: Wiley.

Lee, J. D., & Sanquist, T. F. (2000). Augmenting the Operator Function Model with Cognitive Operations: Assessing the Cognitive Demands of Technological Innovation in Ship Navigation. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, *30*(3), 273-285.

Lewis, Michael, Wang, J., & Hughes, S. (2007). USARSim: Simulation for the study of human-robot interaction. *Journal of Cognitive Engineering and Decision Making*, *1*(1), 98-120. SAGE Publications.

Lospinoso, J. A. (2007). The ELICIT experiment: Eliciting organizational effectiveness and efficiency under shared belief. *12th international command and control research and technology symposium.*

Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination . *ACM Computing Surveys*, *26*(1), 87-119.

Mathieu, J. E., Goodwin, G. F., Heffner, T. S., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273-283.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: lessons from complex environments. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 149-203). San Francisco, CA: Jossey-Bass.

Miller, C A, & Parasuraman, R. (2007). Designing a Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors*, *49*(1), 57-75.

Miller, Christopher A, & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. *Human Factors*, *49*(1), 57-75. Human Factors and Ergonomics Society.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive Automation, Trust, and Self-Confidence in Fault Management of Time-Critical Tasks. *Journal of experimental psychology*, 44-57.

Neerincx, M A. (2003). Cognitive task load design: model, methods and examples. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 283-305). Mahwah, NJ: Lawrence Erlbaum Associates.

Neerincx, M A, & Grant, T. (2010). Evolution of Electronics Partners: Human-Automation Operations and ePartners During Planetary Missions. *Journal of Cosmology*.

Neerincx, M A, & Lindenberg, J. (2008). Situated Cognitive Engineering for complex task environments. In J. M. Schraagen, J. M. C. Militello, L. Ormerod, & R. Lipshitz (Eds.), *Naturalistic Decision Making and Macrocognition* (pp. 373-390). Aldershot, UK: Ashgate Publishing Limited.

Nelson, K. M., & Cooprider, J. G. (1996). The Contribution of Shared Knowledge to IS Group Performance. *MIS Quarterly*, *20*(4), 409. doi:10.2307/249562

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Opperman, R. (1994). *Adaptive user support*. Hillsdale, NY: Erlbaum.

Paping, C., Brinkman, W.-paul, & Mast, C. V. D. (2010). An Explorative Study into a Tele-delivered Multi Patient Virtual Reality Exposure Therapy System. *Wounds A Compendium Of Clinical Research And Practice*, *203*.

Parasuraman, R, & Miller, C. A. (2004). Trust and Etiquette in High-Criticality Automated Systems. *Communications of the ACM*, 51-55.

Parasuraman, R, & Riley, V. (1997). Humans and Automation: Use, Misuse, disuse, Abuse. *Human Factors*, *39*(2).

Parasuraman, R, Bahri, T., Deaton, J., Morrison, J., & Barnes, M. (1992). Theory and Design of Adaptive Automation in Aviation Systems. Warminster, PA: Naval Air Warfare Center.

Parasuraman, R, Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, *38*(4), 665-679. Human Factors and Ergonomics Society, Inc.

Parasuraman, R, Mouloua, M., Molloy, R., & Hilburn, B. (1993). Adaptive function allocation reduces performance costs of static automation . In R. S. Jensen & D.

Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 178-185). Columbus, OH: The Ohio State University.

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

Powell, A., Piccoli, G., & Ives, B. (2004). Virtual Teams: A Review of Current Literature and Directions for Future Research. *The DATA BASE for Advances in Information Systems*, *35*(1), 6-36.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *International Journal of Aviation Psychology*, *10*(4), 393-410.

Prinzel, L. J., Freeman, F., Scerbo, M., Mikulka, P., & Pope, A. (2003). Effects of a psychophysiogical system for adaptive automation on performance, workload and event-related potential P300 component. *Human Factors*, *45*, 601-614.

Rasker, P. C., & Willeboordse, E. W. (2001). Werkafspraken in de commandocentrale [Work arrangements in the Combat Information Centre]. Soesterberg: TNO: Human Factors.

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. Amsterdam: North-Holland.

Rocker, C. (2009). Requirements and guidelines for the design of team awareness systems. *Proceeding of the international conference on information and communication technologies 2009* (pp. 164-172).

Rouse, W. (1988). Adaptive Aiding for Human/Computer Control. *Human Factors*, *30*(4), 431-443.

Rouse, W. (1994). Twenty years of adaptive aiding: Origins fo the concept and lessons learned. In M. Mouloua & R. Parasuraman (Eds.), *Human Performance in automated systems: Current research and trends* (pp. 28-32). Hillsdale, NJ: Erlbaum.

Rouse, W. B., Geddes, N. D., & Curry, R. E. (1987). Architecture for Interface: Outline of an Approach to Supporting Operators of Complex Systems. *Human-Computer Interaction*, *3*(2), 87-122.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., & Lorensen, W. (1991). *Object-oriented modeling and design*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Salas, E, Sims, D. A., & Burke, C. S. (2005). Is there a “Big Five” in Teamwork. *Small Group Research*, *36*, 555-599.

Sarter, N. B., & Woods, D. D. (2000). Team play with a Powerful and Independent Agent: A Full-Mission Simulation Study. *Human Factors*, *42*(3), 390-402.

Scerbo, M. (1996). Theoretical perspectives on adaptive automation. In R Parasuraman & M. Mouloua (Eds.), *Automation and human performance: theory and* *applications* (pp. pp. 37-63). Mahwah, New Jersey: Lawrence Erlbaum Associated Publishers.

Scerbo, M. (2001). Adaptive Automation. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 1077-1079). London: Taylor & Francis.

Schraagen, J. M. (2010). *Methodology: A model and guidelines for human factors research under constraints (TNO-DV 2010 A449).* Soesterberg, the Netherlands.

Sergiu, F., & Luchian, P. (1999). *Task Force for Techincal Intestigation of the 24 March Fire in the Mont Blanc Vehicuar Tunnel.*

Sheridan, T. B., & Verplank, W. L. (1978). Human and Computer Control of Undersea Teleoperators. DTIC Research .

Shlaer, S., & Mellor, S. J. (1992). *Object lifecycles: modeling the world in states*. Yourdon Press Upper Saddle River, NJ, USA.

Smets, N., te Brake, G. M., Buurman, T., Neerincx, M. A., & van Oostendorp, H. (2011). Effects of mobile support on situation awareness and navigation in a field and game environment. *Entertainment Computing*, *2*, 39-46.

Smith-Jentsch, K. A., Kraiger, L., Collings, F., Cannon-Bowers, J. A., & Salas, E. (2009). Do familiar teammates request and accept more backup? Transactive memory in air traffic control. *Human Factors*, *51*(2), 181-192.

Sperandio, J. C. (1971). Variations of Operator's Strategies and Regulating Effects on Workload. *Ergonomics*, *14*(5), 571-577.

Stout, Renée 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 The Journal of the Human Factors and Ergonomics Society*, *41*(1), 61-71. Human Factors Vol 41(1) Mar 1999, 61-71.

Streefkerk, J., Esch-Bussemakers, M., & Neerincx, M. A. (2007). Context -aware notifications for mobile police officers. *Engineering Psychology and Cognitive Ergonomics: 7Th International Conference* (pp. 436-445). Berlin: Springer-Verlag.

Thompson, L. F., & Coovert, M. D. (2006). Understanding and developing virtual computer-supported cooperative work teams. In C. Bowers, E. Salas, & F. Jentsch (Eds.), *Creating high-tech teams: Practical guidance on work performance and technology*. Washington, DC: Amercian Psychological Association.

USAR.nl. (2005). Mission Report Pakistan.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, *41*(5), 656-669. Taylor & Francis Ltd.

Veltman, J. A., & Jansen, C. (2004). The Adaptive Operator. *Human Performance, Situation Awareness and Automation Technology Conference*. Daytona Beach FL.

Vicente, K. J. (1999). *Cognitive Work Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates .

Visser, A., Dijkshoorn, N., Veen, M. V. D., & Jurriaans, R. (2011). Closing the gap between simulation and reality in the sensor and motion models of an autonomous AR . Drone. *Micro*, 40-47.

Voshell, M. G., Woods, D. D., Prue, B., & Fern, L. C. (2007). Coordination Loops: a new unit of analysis for distrubuted work. In K. Mosier & U. Fischer (Eds.), *Proceedings of the Eighth International NDM Conference*. Pacific Grove, CA.

Wickens, C. D. (1984). Processing resources in attention. In R Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63-101). Orland, FL: Academic Press.

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

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance.* NJ: Prentice Hall Upper Saddle River.

Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Human Factors*, *49*(6), 1005-1018.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 1-16). Mahwah, NJ: Erlbaum.

Woods, D. D., Tittle, J., Feil, M., & Roesler, A. (2004). Envisioning Human-Robot Coordination in Future Operations. *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, *34*(2), 210-218.

Xiao, Y., Hunter, W. A., Mackenzie, C. F., Jefferies, N. J., & Horst, R. L. (1996). Task Complexity in Emergency Medical Care and Its Implications for Team Coordination. *Human Factors*, *38*(4), 636-645.

Zaccaro, S., Rittman, A., & Marks, M. (2001). Team leadership. *The Leadership Quarterly*, *12*(4), 451-483. Elsevier.

Zijlstra, F. R. H. (1993). Efficiency in work behaviour. *Delft University Press.*

Zsambok, C. E., & Klein, G. (1999). *Naturalistic decision making.* Mahwah, NJ: Lawrence Erlbaum Associates.

van Delft, J. H., & Arciszewski, H. F. R. (2004). *Eindevaluatie automatiserings- en ondersteuningsconcepten Studie Commandovoering.* Soesterberg, The Netherlands: TNO Human Factors.

9 789088 914317