MASTER

Urban search and rescue robots
the influence of team membership of robots on team performance

Horsch, C.H.G.

Award date:
2012

Urban Search and Rescue Robots

The influence of team membership of robots on team performance.

Corine Horsch

April 2012

identity number 0553003

in partial fulfilment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

Supervisors:

Dr. Ir. Raymond Cuijpers	Eindhoven University of Technology
Dr. Jaap Ham	Eindhoven University of Technology
Nanja Smets MSc	TNO

Abstract

Urban search and rescue (USAR) is the first response to disasters. In these difficult conditions robots can play a crucial role by mapping the disaster area and searching for victims. Making those robots more like team members could possibly increase performance (Fong, Thorpe, & Baur, 2003).

The general idea behind forming teams is that a team performs better than separate individuals. Klein, Woods, Bradshaw, Hoffman, & Feltovich (2004) defined ten challenges for making robots team players. However, it is not yet investigated that when a robot is made a team player according to these challenges team performance actually increases. Therefore the first part of the research question is: When a robot is made a team player does team performance increase?

By making a robot more like a team member other concepts might be influenced as well. The implementation of the ten challenges will probably increase shared situation awareness (SSA) and team identification (TI) and both concepts have been related to performance in previous studies. Therefore, the second part of the research question is: Can the expected increase in performance be explained by increased shared situational awareness or increased team identification?

The task for the 38 teams of two persons and two robots was to find six victims in a virtual world as fast as possible and make a useful map for the firemen who would enter the building later on to really save the victims. Participants were randomly assigned to one of the three conditions, ranging from robots as tools to robots as fully functional team members. For this virtual experiment each participant had one screen that displayed the robot's camera and one screen with the virtual map on it. SSA, TI and performance were measured.

The results showed that the level of team membership of the robots had a negative effect on the time needed to find three victims. Possible explanations could be that the use of the advanced system took more time, or that it induced higher workload, or the need to communicate disappeared because of the advanced robots. Furthermore, no evidence was found that SSA and TI could explain performance, however age and game experience did.

The main lesson learned from this experiment is that it was really hard to achieve true human-robot collaboration and it could be argued that only human-robot interaction took place during the task. Furthermore, it is more sensible to let only people who do know each other participate in this kind of team research. And lastly, we advise to include workload in follow-up experiments.

List of Tables

List of Figures

1 Introduction

Urban search and rescue (USAR) is the first response when man-made structures collapse (Murphy, 2004). For example, when an earthquake destroys a flat or a big car accident obstructs a narrow tunnel. A USAR situation is a safety-critical situation, which means that an error or failure could result in death, injury, loss of property, or environmental harm (Leveson, 1986, as cited in Drury, Scholtz, & Yanco, 2003). Besides being safety-critical, the conditions of a USAR site are poor. USAR teams operate in a chaotic and unstructured environment, where access to facilities like hospitals, power supply, and other resources is limited (Smets & Neerincx, 2011, p.7).

The people working in a USAR team are highly trained and the majority has a daily job as a fireman, police officer, in the military, or as a caretaker. Besides the people in the field, there is often staff present and a support team consisting of a physician, a construction specialist, an ICT-technician, a logistic worker, a cook etc. The workload of USAR teams is often very high due to little preparation time, unknown territory, physical challenges, and emotional demands like dealing with life and death (De Greef, Van der Kleij, Brons, Brinkman, & Neerincx, 2011). In these difficult conditions robots can play a crucial role by mapping the disaster area and searching for victims. Especially when it is not (yet) safe for humans to go into the hot zone. A USAR team consisting of people and robots that work well together would probably save more victims at a higher speed, than USAR teams consisting of only humans. A prerequisite is that the team has to function well and therefore, a robot should function more like a team player (Fong et al., 2003). Klein et al. (2004) have identified ten challenges (see section 2.5) that have to be met to accomplish team membership. However, to our best knowledge, it is not yet investigated that when a robot is made a team player according to these challenges team performance actually increases. Therefore the main research question is: When a robot is made a team player does team performance increase?

Moreover, making a robot a team player might also affect shared situation awareness (SSA) and team identification (TI). SSA and TI are both related to performance, so these concepts may be able to explain part of the difference in performance that is provoked by the level of team membership. Therefore, the secondary research question in this thesis is: Can the expected increase in performance be explained by increased shared situational awareness or increased team identification?

In the following section we will elaborate on concepts that play an important role in human-robot cooperation and could possibly improve the team performance. First performance, automation and teamwork will be reviewed.

Next, it will be described how several authors suggest to make robots team players, and an example will be given. Lastly, the effects of making a robot a team player on shared situation awareness and team identification will be discussed.

2 Related work

2.1 Performance

First we need to establish what team performance is. In general performance consists of efficiency and effectiveness. Efficiency can be measured by assessing the resources, like time needed to complete a task, whereas effectiveness measures how well the task is completed. In teams the performance of team members can be measured in isolation or together, the latter denoted as team performance (Steinfeld et al., 2006). In a USAR situation the goal is to find and save as many victims as possible, as fast as possible. In a USAR team efficiency can be measured by the time needed to find those victims, while the number of victims saved can serve as an effectiveness metric.

2.2 Automation

In robot development the autonomy level of robots has been increasing over the years. However, recent research suggested that an increase in autonomy did not result in better overall performance (Kaber & Endsley, 2004; Scholtz, Antonishek, & Young, 2003; Kaber, Onal, & Endsley, 2000). For example, Bruemmer, Dudenhoeffer, & Marble (2002) conducted experiments with four different autonomy levels: tele-operation, safe mode, shared control, and full autonomy. In the tele-operation mode, the user had full continuous control over the robot. In the safe mode the operator directed the movements, but the robot would only execute them if they did not harm the robot. So the robot stopped in front of obstacles. In the shared control mode the robot determined its own path, based on a given way points. The robot avoided obstacles itself but asked for help when needed. During the full autonomous mode the robot could execute tasks such as "follow that target" or "search this area". It turned out that the shared control mode was the most successful during a USAR competition. Probably because the robot was better in making judgements about its environment than the distant operators. At the same time, the operators were given the means to override the robots safe behavior, if for instance the robot had to move a chair (Bruemmer et al., 2002). This experiment showed that automating all tasks that are possible to automate, does not lead to optimal performance. Whether automation is a good idea depends on the context (Wickens, Lee, Liu, & Gordon Becker, 2004, p. 432).

2.3 Team work and performance

Instead of developing robots as autonomous tools, making them more like team members could possibly increase performance (Fong et al., 2003). The general idea behind forming teams (whoever or whatever is part of the team) is that a team performs better collectively than the sum of the individual performances. A task consists of different aspects and people have different skills. A specialist is always more effective or efficient than a generalist (Kenrick, Neuberg, & Cialdini, 2010, ch. 12). So, a good functioning team leads to an increase in performance compared to the summed performance of individuals on the condition that the team functions well. A good functioning team needs at least properly functioning individual team members, a right allocation of tasks, and coordination. One underlying mechanism of making teamwork effective is backing-up behavior (≈ supporting each other) (McIntire & Salas, 1995, as cited in De Greef et al., 2011). McIntyre and Salas note that backing-up behavior is critical in making a team perform better than the sum of individual performances. Studies have shown that backing-up behavior is positively related to team performance when workload is high (Porter, 2005; Porter et al., 2003 as cited in De Greef et al., 2011).

2.4 Definition of a team

In the previous paragraphs we employed the term team, but a definition was not yet discussed. Over the years researchers have given several definitions of teamwork, collaboration, and cooperation (Baker & Salas, 1992; Nieva, 1985; Bradshaw et al., 2009). In this research the team definition of Nieva (1985) will serve as a guideline; "*two or more interdependent individuals performing coordinated tasks toward the achievement of specific task goals*".This definition consists of two components. Firstly, all team members are task oriented; they aim for the same goal. Secondly, the members depend on each other. For example, because they have different skills, which means coordination is needed. Nieva notes himself that by the definition a distinction is made between real team situations and multi-individual situations (also made by Wagner et al., 1977) when the team goal is divisible into subtasks that different team members can take care of (Nieva, 1985).

2.5 Making robots team players

The definition of a team already gave a hint that a team player should be capable of coordination. A few researchers, however, specifically concretised how robots should be made team players and described the concepts that are important for human-robot collaboration.

Ten challenges
For example, Klein et al. (2004) defined ten challenges for making automation a team player (Klein et al., 2004). These challenges are: Basic compact, Collaboration, Cost Control, Adequate models, Predictability, Interpreting signals, Directability, Goal negotiation, Revealing status and intentions, and Attention management. With a 'basic compact' Klein et al. (2004) mean

the (implicit) agreement to work together. Linked to the 'basic compact' are the 'collaboration' and 'cost control' concept. The first one includes for example the give-and-take principle. 'Cost control' means that team members try to keep costs for coordination, like communication, at a minimum. With 'adequate models' is meant that team members have correct mental models of the other team members on which behavior can be 'predicted' and 'signals can be interpreted'. 'Directable' means that the team members can control each other's behavior (to a certain extent). Furthermore, team players should be able to 'negotiate about goals' and their importance when the situation changes. Lastly, team players should 'reveal their own status and intentions' and 'manage the attention' of the others to improve common ground. In table 1 an overview of the ten challenges and their brief description is given.

Information sharing

Besides ideas from a theoretical point of view as described by Klein et al. (2004), guidelines for human-robot teamwork also arose from field studies. After the terrorist attacks of 9/11 a few robots were employed during the ten days response period. In addition to interviews with rescuers, post-hoc analyses of video data and logs were done resulting in seventeen findings (Casper & Murphy, 2003). One of the conclusions was that information gathered by the robot should not only be available to the operator, but also to the right authority. It appeared that it took often twelve hours before information about victims reached the right authority. One of the authorities suggested a device that displays maps of who is working where and provides updated event information. Moreover, information should not be just broadcasted to all team members. The analysis showed that information should be shared according to the needs of each team member (Casper & Murphy, 2003).

Coordination and communication

Furthermore, also coordination and communication is seen as important for a human-robot teamwork. For an experiment in which an intruder must be detected and arrested on a cluttered Navy Pier with two humans and five robots several coordination policies were developed (Bradshaw et al., 2009). In terms of coordination, requests for an action are accepted when they come from a higher authority or team members can initiate an action themselves. In addition, communication is needed for coordination. Therefore, the team members had the obligation to send an acknowledgement to the requester when they accepted a request, they had to notify the requester when the action was finished (completed, aborted, failure). Furthermore, the team leader must be informed about all requests and outcomes. The authors concluded that if these messages regarding coordination are part of a robotic system, robots will be seen less like a tool and more like a team member (Bradshaw et al., 2009).

Table 1: The 10 challenges of Klein et al. (2004)

10 challenges	**Description**
A basic compact	To be a team player, an intelligent agent must fulfill the requirements of a basic compact (agreement to facilitate coordination, work toward shared goals, prevent breakdowns in team coordination) to engage in common-grounding activities
Adequate models	To be an effective team player, intelligent agents must be able to adequately model the other participants' intentions and actions vis-à-vis the joint activity's state and evolution - *for example, are they having trouble? Are they on a standard path proceeding smoothly? What impasses have arisen? How have others adapted to disruptions to the plan?*
Predictability	Human-agent team members must be mutually predictable
Directability	Agents must be directable
Revealing status and intentions	Agents must be able to make pertinent aspects of their status and intentions obvious to their teammates
Interpreting signals	Agents must be able to observe and interpret pertinent signals of status and intentions
Goal Negotiation	Agents must be able to engage in goal negotiation
Collaboration	Support technologies for planning and autonomy must enable a collaborative approach (facilitate give-and-take)
Attention management	Agents must be able to participate in managing attention
Cost control	All team members must help control the costs of coordinated activity

Making robots team players in short
In summary, besides the ten challenges from Klein et al. (2004), information sharing (Casper & Murphy, 2003), coordination, and communication (Bradshaw et al., 2009) are mentioned to make robots more like team players. Most ideas about human-robot collaboration do overlap, at least partly. Information sharing as mentioned by Casper and Murphy (2003) is also covered by Klein et al. (2004) in the 'revealing status and intentions' challenge. Furthermore, coordination and communication from Bradshaw et al. (2009) are closely related to the ten challenges as well. Coordination is captured in 'directability', 'collaboration', and 'revealing status and intentions', whereas communication is captured in 'interpreting signals', 'attention management', 'cost control' and 'revealing status and intentions'. In general we can state that the ten challenges are the most comprehensive description of how to turn a robot into a team member.

2.6 An example of a robot as a team player

Breazeal et al. (2004) conducted an experiment with a robot that functioned to a certain extent as a team player. They made a robot 'Leonardo' that was able to learn through social skills. So in a scenario Leonardo followed instructions from a human to associate names with objects. Subsequently, Leonardo was asked to perform actions on the labeled objects. For example, press the left button. Breazeal et al. (2004) used learning which has similar characteristics as collaborative interaction. Leonardo was sensitive to social cues and was able to show non-verbal cues as well. The robot was able to understand limited speech, interpret visual gestures, share attention, and took appropriate turns. The underlying learning mechanism benefited from using social skills, or as Breazeal et al. (2004, p.342) stated: "*... the teacher becomes a more effective teacher and the learner a more effective learner - each simplifies the task for the other.*"

Leonardo and the teacher monitored the learning process and each other continually. That means the robot communicated its internal state to maintain mutual believes. This was done through demonstrations of the learned task, expressive gestures, and eye gaze. Through expressive gestures and eye gaze, Leonardo also prompted the teacher for feedback and extensive explanation, when he was not sure about the task representation. For example, when Leonardo did not understand the teacher he widened his eyes, lent forward and rose an ear towards the person. If Leonardo did not know a part of an assignment, for example the green button in 'point to the green button', he showed confusion by tilting his head and shrugging. By leading the robot through the task feedback can be given immediately at appropriate times (Breazeal et al., 2004).

The ten challenges applied to Leonardo

If we map Klein's challenges to Leonardo we assess that Leonardo includes: 'directability', 'revealing status and intentions', 'interpreting signals', and 'collaboration'. Furthermore, he is probably engaged in a 'basic compact' and has 'adequate models' in his belief system. The task manager model within Leonardo took care of 'goal negotiation'. However, 'predictability' and 'cost control' were not explicitly present in Leonardo. Though it can be argued that natural communication already kept the costs low compared to text or other interface options. The goal of the authors was to build a socially intelligent, cooperative humanoid robot and unknowingly they adhere to almost all challenges of Klein et al. (2004). The authors, however, did not test the social intelligence or cooperative behavior of the robot. They also did not test if performance (e.g. learning time) improved by building Leonardo in this way.

2.7 Consequences of making a robot a team player

By making a robot more like a team member, a robot will not only behave more as a team player, other performance related concepts might be influenced as well. By 'revealing the status and intentions' of the robot and making sure team members have 'adequate models' of each other, shared situation awareness (SSA) will probably increase. Furthermore, the 'basic compact' and 'collaboration' challenge will probably effect the extent to which members feel a part of the team; team identification. Both the relation between SSA and performance, and team identification and performance have been studied before. In the next sections a brief overview of SSA, team identification and their relation with performance will be given.

2.8 Shared situation awareness

Situation awareness

Shared situation awareness is derived from situation awareness (SA). A general accepted definition of situation awareness is given by Endsley: "*The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and projection of their status in the near future.*" (Endsley, 1988, p. 97). The definition implies three levels of situation awareness; perception; the sensing of the environment (level 1), understanding the environment (level 2), and projection; predicting consequences and the near future state of the environment (level 3). Note that, not all three levels have to occur at the same time in the same amount. For example, a driver can have seen (SA level 1) and understood (SA level 2) the meaning of a red traffic light, but does not realize the impact of it on travel time (SA level 3) (Wickens et al., 2004).

Shared situation awareness
In a (distributed) team not only the SA of an individual is important. To be effective as a team the members need to develop a high degree of shared situation awareness (SSA) (Bolstad, Cuevas, Gonzalez, & Schneider, 2005). In short SSA is the reflection of how similarly team members view a given situation. More elaborately SSA means that individuals perceive, comprehend, and interpret the situation's information in a similar manner (in line with Endsley's definition) (Bolstad et al., 2005).

Human-robot interaction awareness
Some authors think it is too simplistic to just combine the SA scores of individuals into SSA. They state that SSA requires extra activities such as coordination and information sharing, next to the cognitive processes that underlie SA (Salas, Prince, Baker, & Shrestha, 1995). That (S)SA is not an easy concept is demonstrated by Drury et al. (2003). They give 16 different definitions of several awareness concepts ranging from group-structural awareness to conversational awareness and also include Endsley's definition of situational awareness. The authors themselves focus on human-robot teams and define a new awareness concept: human-robot interaction (HRI) awareness. With HRI awareness they mean the understanding the human has of the location, activities, status, and surroundings of the robot *and* the knowledge the robot has of the human's commands. If a team exists of multiple humans and robots then HRI awareness consists of five components: human-robot, human-human, robot-human, robot-robot, and humans' overall mission awareness.

2.9 Situation awareness and performance

"*Like workload, SA is thought to correlate with performance.*" (Uhlarik, 2002, p.1) As the subtlety in the citation indicates, performance and SA are often seen as related. However, there is an important distinction between SA and performance. Somebody can have a high situation awareness, while no performance is displayed. The other way around can also occur. A very high performance can be delivered, while SA is not so high (Wickens et al., 2004). For example, when you are absorbed in doing a task so you are not aware of the environment. Or when you drive to your work in the morning, you perform well by reaching your work, but you may not have a clue about the buildings nor billboards you past. Your SA is low, but your performance is high. The last case could be explained by automaticity. Nevertheless, the relation between SA and performance is less interesting in routine tasks, like driving to your work, than in unexpected events (Wickens, 2000). When a human or USAR team have to respond to an unexpected event, SA, performance and their relationship become more important. And the gut feeling is that they do correlate. Some studies confirm this relationship between SSA and performance (Bolstad et al., 2005), while other experiments do not find a relationship (De Greef et al., 2011). Again other researchers only find a relationship between SA and a specific performance measure and no relationship between SA and another performance measure (Riley & Strater, 2006).

2.10 Team identification

Team identification (TI) is a closely linked concept to group cohesiveness, which is often seen as an important factor for effective team performance (Guzzo & Shea, 1992, p. 284). "*Identification is the perception of oneness with or belongingness to a group, involving direct or vicarious experience of its successes and failures.*" (Ashforth & Mael, 1989). People who identify highly with their team, view themselves primarily as team member. Whereas, low team identification indicates that people see themselves primarily as an individual. The level of team identification and according self-image affect perception and behavior (Van Vugt & Hart, 2004). According to Ashforth & Mael (1989), however, identification only comprises cognition and not behavior or affect. Behavior and affect may arise from cognition, though (Ashforth & Mael, 1989).

2.11 Team identification and performance

Nevertheless, research has shown that people work hard on collective tasks when the group is valued highly (Barreto & Ellemers, 2000). This result suggests that people show pro-group behavior out of an intrinsic motivation due to team identification (Barreto & Ellemers, 2000). Furthermore, a study about transnational teams has found team identity as a mediator in their model to predict team performance (Shapiro, Furst, Spreitzer, & Von Glinow, 2002). Team players with a higher team identification feel more intertwined with the team's performance; the needs of the team are seen as a part of the person's individual needs. When team identity is high, effort-withholding behavior conflicts with the team members' self-definition. Self-definitions are chosen by people themselves and are typically satisfying. Therefore, acting against those self definitions (by showing effort-withholding behavior in high team identification situations) is experienced as self punishment (Mael & Ashforth, 1995, p.310; Ashforth & Mael, 1989, p.21, as cited in Shapiro et al., 2002). In summary, a high team identification will lead to less effort-withholding behavior and therefore to a higher performance.

3 Research question

In a USAR situation it is critical to save as many victims as possible, as fast as possible. In order to accomplish this goal USAR teams consist of people with different skills that work together. Robots could be a useful addition to these teams, since they have different capabilities than humans. The added value of a team compared to separate individuals is exactly that team members have different skills that can be utilized. The performance of human-robot teams, however, only increases when a team functions well. Therefore, a robot should be made more like a team player. Klein et al. (2004) have identified ten challenges (discussed in section 2.5) that have to be met to accomplish team membership. However, to our best knowledge, it is not yet investigated that when a robot is made a team player according to these challenges team performance actually increases. Therefore the *first part of our research question* is: When a robot is made a team player does team performance increase?

Moreover, making a robot a team player might also affect shared situation awareness (SSA) and team identification (TI). SSA and TI are both related to performance, so these concepts may be able to explain part of the difference in performance that is provoked by the level of team membership. Therefore, the *second part of our research question* is: Can the expected increase in performance be explained by increased shared situational awareness or increased team identification?

First, we expect team performance to increase over the conditions in which the level of team membership increases (*hypothesis 1*). Teams are thought to perform better compared to the sum of the performance of individuals. Well functioning teams make optimal use of their different capabilities through coordination and communication, and keeping each other up to date will probably result in less out-of-the-loop related problems.

In condition one where only human-human collaboration takes place, performance is expected to be the worst. So, those teams will probably find the least victims (effectiveness) and it will probably take the most time (efficiency). In condition three, where the robot fully acts as a team player, performance is expected to be the best. So, the most victims will probably be found (effectiveness) and finding the victims will probably take the least time (efficiency). Team performance in condition two is expected to lie between the performance in condition one and three. In condition two the situation awareness will probably be higher than in condition one, whereas team identification will probably be lower than in condition three.

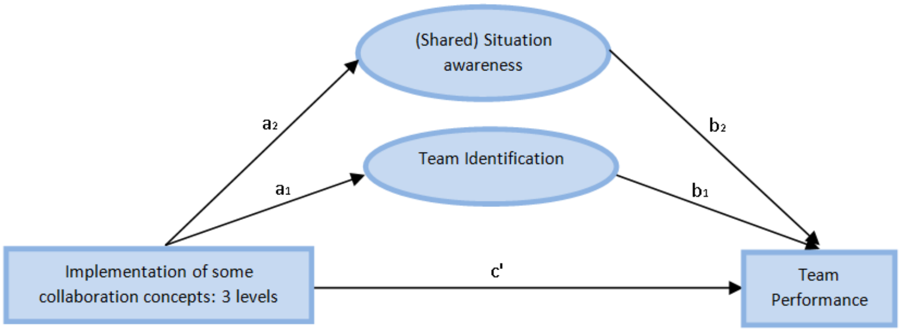


Figure 1: Mediation Model

Secondly, we expect that part of the difference in team performance can be explained by the level of shared situation awareness and team identification (*hypothesis 2*). In figure 1 this mediation is depicted, and we expect a_1, a_2, b_1, and b_2 to be different from zero. Not in every experiment a relationship between SSA and performance is found, though it is often assumed that these concepts are positively related (Kaber et al., 2000; Endsley & Kiris, 1995). Therefore, we expect that an increase in performance can be partly explained by SSA. Furthermore, we expect that an increase in team performance can also be partly explained by team identification. When team members identify themselves highly with the team, their goals correspond to the team goals and the members will probably display less effort-withholding behavior, which results in better performance (Shapiro et al., 2002; Barreto & Ellemers, 2000).

4 Method

4.1 Task

The task for the team was to find six victims in a virtual world as fast as possible and make a useful map for the firemen who would enter the building later on to really save the victims. The description of the task as handed out to the participants can be found in appendix A.

4.2 Design

This experiment had a between subject design. Participants were randomly assigned to one of the three conditions. The participants worked in teams of two. Between the conditions the level of team membership of the robotic system was varied. The level of team membership was based on the ten challenges of Klein et al. (2004). In condition one no collaboration between the participants and the robots was supported. In condition three all challenges were supported, because all challenges are important for effective teamwork as argued by Klein et al. (2004). In condition two only the challenge 'revealing status and intention' was implemented. 'Revealing status and intention' was supposed to support shared situation awareness, which could be an important explanatory factor for performance. Therefore, this challenges was investigated on itself, next to all challenges together.

Table 2: The implementation of the 10 challenges of Klein et al. (2004)

10 challenges	**Implementation in condition three**
A basic compact	Explicit basic compact (association)
Adequate models	Description of robot
Revealing status and intentions	Overview map
Predictability	= adequate model + status and intention
Directability	Lack of time after 12 minutes (robot speech)
Goal Negotiation	Narrow passage, robot advises to take another route
Collaboration	Victim found, help to determine condition of victim
Attention management	level of importance + alarm with high importance message
Cost control	WoZ (no irrelevant info) and filters
Interpreting signals	Not implemented

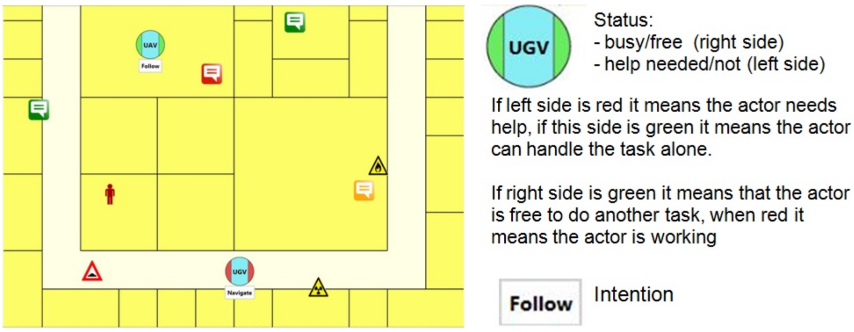
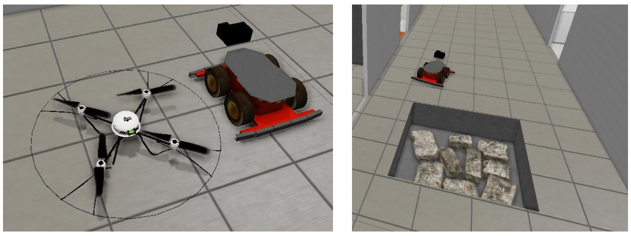


Figure 2: How intentions and status are shown in the map

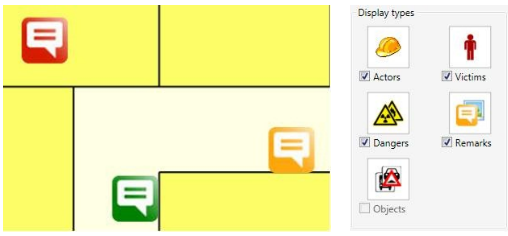
In condition one, the robot was like a tool and only human-human collaboration took place. When humans collaborate they perform most of the ten challenges automatically. Nevertheless, some challenges were made explicit. The 'basic compact'; an agreement to cooperate, was explicitly achieved by asking the participants "*What do you mean by collaboration?*". After that question participants were probed with "*What are things you do and do not do when collaborating?*". The questions were followed by the remark that the aim of the experiment was to collaborate with each other. The other challenges that were explicitly mentioned were 'revealing status and intentions', 'directability', 'goal negotiation', and 'collaboration' in the form of helping each other out. In appendix A the task description in which the challenges were made explicit can be found. The task description also included a picture and a little description of the robots.

Condition two differed from condition one in that the robotic system 'revealed its status and intentions' through an interactive map. So the robot was already a little bit more than just a tool. In figure 2 can be seen how the 'status and intentions' were indicated on the overview map. The map showed where the robots were, what they were doing and if the robot was capable of doing its task. The robots were represented as circles and the task they were performing was indicated below the robot in a rectangle. The sides of the circles that represented the robots could change colors. The sides could be green or red to indicate if the robot needed help or not and if the robot was busy or not.

In condition three the robotic system is also capable to perform most of the other challenges of Klein et al. (2004). Table 2 provides an overview of how the challenges were practically implemented and figure 3 gives some examples. In order to establish a 'basic compact' with the robot the question "*What do you mean by collaboration with a robot?*" was added to the other questions during the introduction. Furthermore the description of the



(a) Adequate models (b) Goal negotiation



(c) Attention management (d) Cost control

Figure 3: Some examples of implementations of a few challenges

mission deviated in a way that if the team was mentioned the robots were included in the team. Furthermore, the robotic system was able to 'direct' the participants, 'negotiate about goals', and ask for help ('collaboration') via speech. In appendix D the exact sentences can be found. To provide an 'adequate model' of the robot a picture of the robots along with a description of the capabilities of the robot was given (See appendix B). Furthermore, participants were able to indicate how important a message they placed on the map was in condition three. When a highly important message was placed on the map the other participant heard a beep to 'attract attention' to it. Lastly, 'costs of information and communication' could be controlled by the participants by using filters in the overview map. Furthermore, the robots did not share irrelevant information about all kinds of processes, like driving speed, toxic gas values, etc. The last challenge 'interpretation' was not supported by the robotic system.

4.3 Participants

In this experiment 76 persons participated, 52 were male and 24 were female. The participants formed 38 teams of two persons, 34 teams consisted of two males, only six teams were totally female, and 34 teams were mixed. 20 teams did not know the other participant and in 18 teams the participants did know each other. The participants were on average 26 years old and 65 of them possessed a driver license. 32 participants indicated that they never play first person shooter games, 22 participants play less than once a month,

Figure 4: Screenshots of the virtual environment in Unreal

From left to right: an example of a destroyed office, an example of an victim lying under a desk, the UAV (flying robot)

eight participants play once a month, eight participants play a few times a month, and six participants play from once a week to daily. In general can be concluded that the participants are not or little experienced in playing games.

4.4 Materials

For this virtual experiment six laptops were connected through through UTP-cables. One of the laptops functioned as a server on which Unreal Development Kit and UsarSim were installed. Figure 4 depicts some screenshots of the destroyed office environment that was created with the Unreal Editor. The *P3AT* robot and the *Air* robot were used which are standard robots from UsarSim. UsarSim is a virtual simulation environment for urban search and rescue robots. It uses Epic Games' Unreal Engine 3 to provide a high fidelity simulator at low costs. The robots were controlled with a standard game controller (Logitech Dual Action).

Figure 5 depicts the Organizational Awareness Display developed by TNO which was used for the participants to create a map. This program allowed people to add all kind of messages, like a victim report, an obstacle, a fire etc. on a map of the destroyed office. The Organizational Awareness Displays off the participants were linked, so they could see from each other what they were doing. Moreover, this program was connected to Unreal so that the actual positions of the robots and people were displayed. Furthermore, walkie-talkies were used for communication between the participants. In order to keep the way of communication between people and robots more equal, only verbal communication between the participants was allowed and non-verbal cues were eliminated. Therefore, the participants were located in a different cubicles and walkie-talkies supported communication.

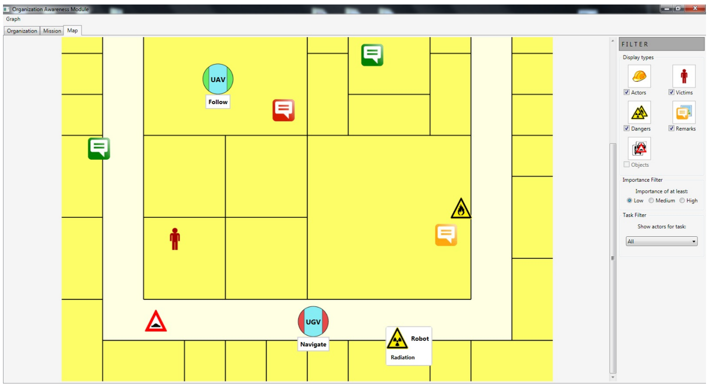


Figure 5: Screenshot of the organization awareness display

4.5 Procedure

Participants were provided with an informed consent form describing the experiment. All the question participants had were answered by the experiment leader after which the form was signed. Next, the questions about collaboration (mentioned in section 4.2) were asked in the form of an association exercise to establish the 'basic compact'. After that an elaborated description of the mission and the robots was provided to the participants. The participants were encouraged to ask questions whenever something was unclear to them. After reading the descriptions the participants had to decide together who was going to operate which robot. Then the controls of the robots were explained and a brief explanation of the walkie-talkies was given. In condition two and three the participants also got a little manual which explained the overview map. Then the experiment began with a training session in a small room, separate from the office (see Figure 6 for two screenshots of the training area). The participants were asked to get familiar with the controls of the robot for a few minutes. Furthermore, they could see how a victim looked like. The teams also tested the walkie-talkies and if applicable the overview map. After the training the participants were ready for the real task and the mission started. During the mission the Wizard of Oz provided the robots' 'status and intentions' on the overview map in condition two as well as the speech in condition three. After the mission the laptops were closed and the participants filled out the questionnaires. Starting with the demographical questions, followed by the (S)SA measures, team identification questions, the manipulation check and the collaboration questions. At the end there was some space to evaluate the robot and other comments could be written down. The participants were debriefed and thanked. In total the experiment lasted around 60 minutes.

Figure 6: Screenshots of the training area

4.6 Measurements

First, a demographic questionnaire asked about the participants gender, age, relationship to the other participant and driving and gaming experience. The questions about driving experience asked if the participant had a driver license (yes/no), if the answer was yes the next question was how long they had the license (years), and how many kilometers they drive per year. The gaming questions asked about the kind of game controllers the participants were familiar with, what kind of game console they had at home, if any, and how often the participant played (first person shooter) games on a seven point scale ranging from never to daily.

Later on team variables were calculated from the demographics to describe the **team characteristics**[1]. For example, average team age was calculated by summing the age of the two participants and divide it by two. Likewise, average time possessing a driver license and average driven kilometers per year were calculated. If one of the team members did not possess a driver license, the answer of the other member was divided by two. Team gender was defined as mixed, males or females and a dichotomous variable was created to indicate if the participant knew each other or not. Lastly, average team game experience was calculated by taking the average score of the question *"How often do you play first person shooter games?"*

[1] Team characteristics were already partly described in the participant section.

$$\text{Integrated team performance} = P_{victims} + P_{objects} + P_{areascovered}$$

$$P_{victims} = \frac{1}{vr_1} + \frac{1}{vr_2} + \frac{1}{vr_3} + \frac{1}{vr_4} + \frac{1}{vr_5} + \frac{1}{vr_6}$$

$$P_{objects} = \frac{1}{or_1} + \frac{1}{or_2} + \frac{1}{or_3} + \frac{1}{or_4}$$

$$P_{areascovered} = \frac{1}{ar_1} + \frac{1}{ar_2} + \frac{1}{ar_3} + \frac{1}{ar_4} + \frac{1}{ar_5} \cdots + \frac{1}{ar_{17}}$$

vr_i means the finding rate[2] of victim i
or_i indicates the ratio[3] of teams that have drawn object i in their map
ar_i indicates how many teams from the total number of teams[4]
have put some information on the map about area i

Figure 7: Formulas used to calculate the integrated performance score

Performance
It was intended to measure performance with the time a team needed to find the six victims. After conducting the experiment, it turned out that only two teams succeeded in finding all six victims within 22 minutes: the maximum given time. Therefore team performance was measured by taking the time it took the teams to find three victims. This performance measure will be called *time performance* from now on. Six of the 38 teams did not find three victims within the maximum given time, so they were excluded from analyses that concern time performance. Furthermore, a performance score was calculated instead of just counting the number of victims a team found. Figure 7 lists the equations used for calculating the score. This performance score included the difficulty of finding a certain victim; the number of the four main obstacles indicated on the map by the teams; and the areas that the teams indicated on the map they had searched. Figure 8 shows the map with the victims, obstacles and designated areas within the office. This performance score compared the teams with the other teams that

[2] e.g. victim 3 was only found 11 times, so the finding rate is $\frac{11}{38}$
[3] e.g. object 4 was only drawn 15 times on the map, so the object drawn rate is $\frac{15}{38}$
[4] e.g. information about area 12 was only drawn 20 times on the map, so the area information rate is $\frac{20}{38}$

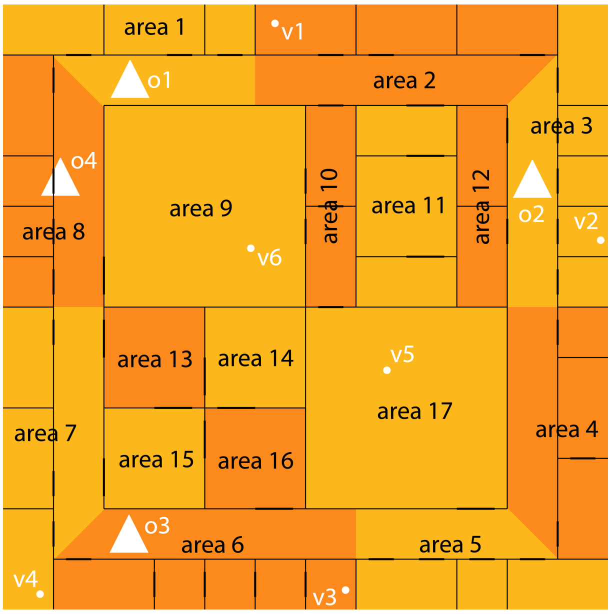


Figure 8: Map with victims, objects and areas

participated in the experiment and included both mission goals; finding six victims and draw a good map. By calculating a score more variance was obtained than by just counting the number of found victims. From now on this measure will be called the *integrated performance*.

Situation Awareness

Situation awareness was measured by two different questionnaires which can be found in appendices E.2 and E.3. The first SA questionnaire consisted of a big part of the SPASA questions proposed by Gatsoulis, Virk, and Dehghani-Sanij (2010). The two items that were deleted to increase *Cronbach's Alpha* from 0.67 to 0.71 can also be found in appendix E.2. The second SA questionnaire was based on a questionnaire used by Mioch, Smets, and Neerincx (2012). This scale turned out not to be reliable in this experiment (*Cronbach's Alpha* = 0.18). Table 3 lists the questions that were used to obtain a shared situation awareness score. The answers of both participants in a team were compared to each other. For example, the participants were asked *which color the floor of the office had*. The teams got a zero if both participants gave the same answer, otherwise a one. For questions like *how much time did you have for the mission* the answers of the participants in a team were subtracted. So this means how closer the score is to zero how better the team's SSA was. First the scores of the questions regarding the same aspects were summed (table 3 also shows the exact clusters) and then the scores per cluster were turned into Z-scores. Per team the average of all Z-scores was taken, resulting in one overall SSA score.

Table 3: Shared situation awareness questions. Per level and specific metric the sum was calculated, then the sums were transformed into Z-scores, and lastly the average of all questions was taken to serve as SSA measure.

SSA level 1

Questions	Team score for question
• What color is the floor of the building? • What was located to the right of the UGV at the end of the mission? • What color is the ceiling of the building? • What color are the most seats in the office? • What color are most walls in the office?	When both participant in a team answered the same they got a 0, otherwise a 1.

SSA level 2 (number of rooms)

Questions	Team score for question
• How many rooms have you searched together? • How many rooms did the driving robot (UGV) search? • How many rooms did the flying robot (UAV) search?	Difference between the answers of both participants within a team.

SSA level 2 (percentage)

Questions	Team score for question
• What percentage of the building have you searched together?	Difference between answers.

SSA level 3 (meters)

Questions	Team score for question
• How wide is the office?	Difference between answers.

SSA level 3 (minutes)

Questions	Team score for question
• How much time do you have for the mission? • How much time would you need to search the entire building?	Difference between answers.

Table 4: Team identification questions

- I feel emotionally attached to my team
- I feel emotionally attached to the other person in my team
- I feel emotionally attached to the robots in my team
- I feel a strong sense of belonging to my team
- I feel a strong sense of attachment to the robot in my team
- I feel as if the team's problems are my own problems
- I feel as if the problems of the other person in my team are my own problems
- I feel as if the robot's in my team problems are my own problems
- I feel like part of the family in my team

Team identification
Team identification was measured by ten items based on four items chosen by Van Der Vegt and Bunderson (2005) from an affective commitment scale from Allen and Meyer (1990). In table 4 the used questions can be found. The items were measured on a 7-point scale ranging from totally disagree to totally agree (Van Der Vegt & Bunderson, 2005). The average of all questions per team was calculated to serve as team identification measure with a Cronbach's Alpha of 0.89. The data was negatively skewed and therefore a log transformation[5] was done.

Manipulation check questions
Furthermore, a manipulation check was done. From the manipulation check questions one question was deleted to increase Cronbach's Alpha from 0.74 to 0.76 (See appendix E.6 for the questions and deleted questions). The data was negatively skewed and therefore a log transformation[6] was done.

Coordination and communication
Furthermore, questions about collaboration covered coordination and communication, and trust and satisfaction. In appendix E.7 an overview of the coordination and communication questions is given. Coordination and communication is seen as important for effective collaboration as mentioned earlier in section 2. Therefore, the participants were asked about their experience regarding coordination and communication during the mission. In addition, audio of the vocal communication between the team members was

[5] $TIlog = lg10(7-(TI1+TI2+TI3+TI4+TI5+TI6+TI7+TI8+TI9+TI10)/10)$
[6] $MClog = lg10(7-(MC1+MC2+MC3+MC4+MC6+MC7)/6)$

recorded. Later on the messages were categorized into the following categories: providing information, asking a question, giving an assignment, positive answer, negative answer, confirm message, ask for repetition of former message, and unknown. This categorization is in line with earlier research about an UAV team that showed four types of messages: informational, confirmation, warning, and encouragement/reinforcement (Yagoda & Coovert, 2011). An example of a warning message "come forward to avoid poles" would be categorized as assignment in this experiment. The current categorization deviates on the encouragement/reinforcement messages like "that's a good position" which are included in the information messages in this experiment.

Trust and satisfaction

In general, trust is seen as important for effective teamwork. Especially in complex systems that require coordination trust has been shown to be a determining factor (Granovetter, 1985; McAllister, 1995; Seabright, Levinthal, & Fichman, 1992, as cited in Aubert & Kelsey, 2003). In appendices E.8 and E.9 the trust and satisfaction questions are listed. A Principal Component Analysis (PCA) was conducted to reveal the underlying factors within the trust questions. For example, trust in the team, trust in the robots, and trust in the task. In table 5 the deduced factors and their Cronbach's alpha can be found. Satisfaction, could possibly explain unexpected results and was measured for that reason.

Table 5: Factors distracted from the trust questionnaire

Trust in team *Cronbach's Alpha = .93*
I would be able to count on my team
I would be able to count on the other person in my team
I can trust my team
I can trust the other person in my team
Trust in robots *Cronbach's Alpha = .85*
I would be able to count on the robots in my team
As a firefighter / firefighter, I would use the robots
I can trust the robots in my team
I can assume that the robots in my team will work properly
I would entrust the robots in my team to find victims
Trust in task *Cronbach's Alpha = .82*
I can assume my team will work properly
I can assume the other person in my team will work properly
I would entrust my team to find victims
I would entrust the other person to find victims
Careful with team and team members *Cronbach's Alpha = .74*
I am wary of my team
I am wary of the robots in my team
I am wary of the other peron in my team

5 Results

To check our manipulation between the three condition a one-way independent ANOVA (condition: three levels) was conducted. There was an effect of condition on the manipulation check questions, which indicates that the level of team membership in condition one was lowest (M = .31, SD = .23)[7], in condition two team membership was highest (M = .15, SD = .19), and condition three fell in between (M = .23, SD = .20, $F(2, 73) = 3.79$, p = .027, r = .31). A planned contrast revealed the manipulation check score differed for having a robotic system compared to a paper map (condition 1 vs. condition 2 and 3), $t(73)$ = -2.41, p = .019, r = .27.

5.1 Team performance per condition

To investigate the relationship between the level of team membership and team performance (*hypothesis 1*), we analysed the integrated and time performance scores for the three conditions. In figure 9 the integrated team performance is showed for each condition. Results revealed no effect of the condition on the integrated performance (the measure where the number of victims, objects, and areas were included in the performance score). In table 6 is shown that the different performance aspects (victims, areas, objects) did not differ across conditions.

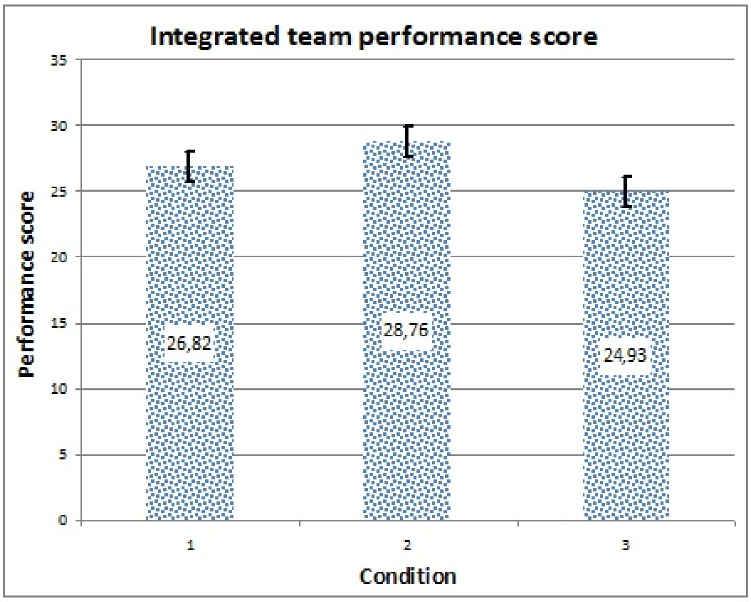


Figure 9: Integrated performance (error bars indicate the Standard Error)

[7] Note that the data was negatively skewed, and therefore a the scores were subtracted from the maximum score (7) and after that a log transformation was done

Table 6: Integrated team performance split up into the separate aspects: victim performance, area performance, and object performance. No differences within these aspects were found between the conditions

		Victims		Areas		Obstacles	
	N	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Con 1	13	5.76	2.71	17.45	3.45	3.61	1.74
Con 2	12	6.81	2.68	18.10	4.36	4.53	1.58
Con 3	13	5.49	1.77	15.54	4.06	3.90	1.76
	38	ns, r = 0.23		ns, r = 0.27		ns, r = 0.23	

Moreover, when time performance was assessed (the time the teams needed to find three victims), there was a trend that how higher the level of team membership of the robots how longer it took to find the victims, $F(2, 29) = 2.45$, $p = .104$, $r = .38$. In other words, the direction of the found effect is opposite to the one expected. A planned contrast revealed that having a robotic system increased the time needed to find three victims compared to only having a paper map, $t(29) = 1.99$, $p = .056$, $r = .35$. Which is also against expectations.

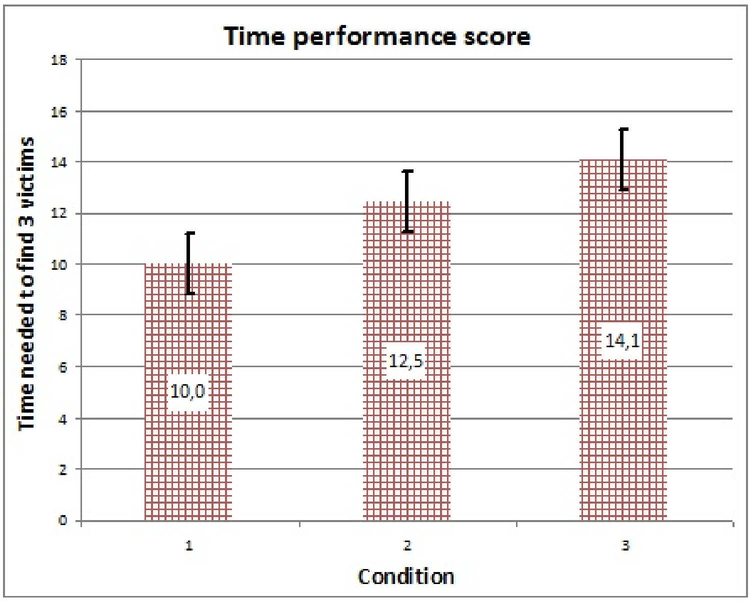
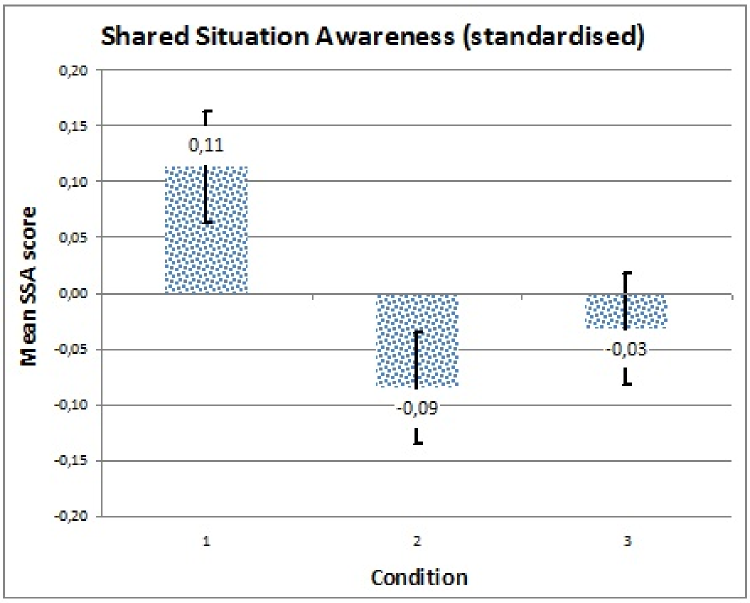


Figure 10: Time performance (error bars indicate the Standard Error)



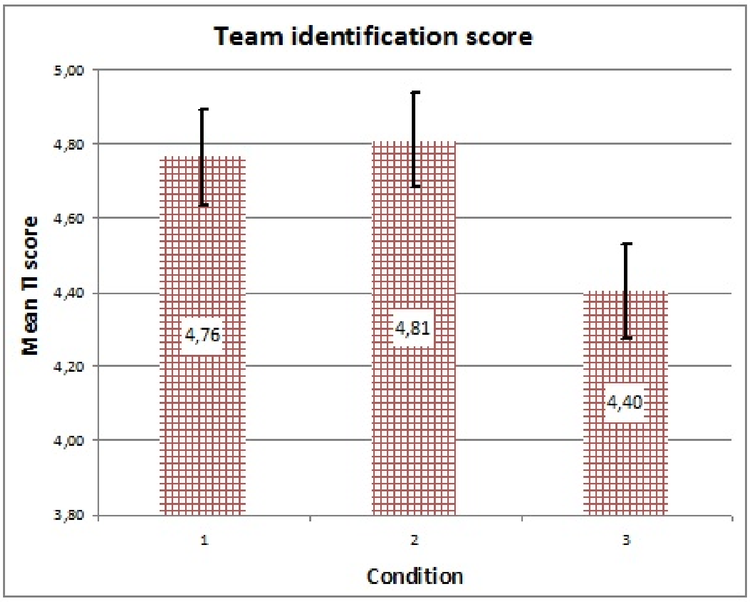


Figure 11: Average shared situation awareness and team identification scores per condition. In condition two SSA and TI are the best*. The error bars indicate the Standard Error.

*Note that SSA indicates 'how similar participants view the situation' and that SSA was measured by taking the difference between the answers of both participants, which means that the scores closer to zero indicates a better SSA. Next the scores were transposed into Z-scores, meaning how lower the score how higher the SSA. SSA is worst in condition one, and team identification is worst in condition three.

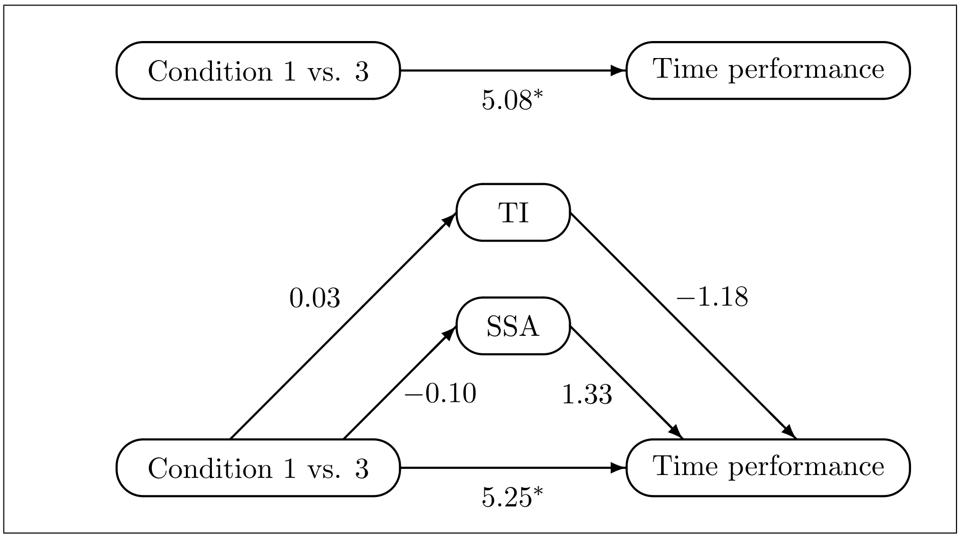


Figure 12: Multiple mediation graph with the belonging coefficients (*denotes p <.05). Time performance can be partly explained by the conditions. However, it was not found that SSA and TI could explain time performance.

5.2 The explanatory power of SSA and TI for team performance

In order to investigate if SSA and TI could explain a difference in performance (*hypothesis 2*) a mediation analysis as put forward by Baron and Kenny (1986) was conducted through the PROCESS script[8]. The conditions were used as independent variable, performance as dependent variable, and SSA and team identification served as mediators. To get a feeling for the values of the mediators per condition see figure 11.

First, a mediation analysis was conducted for integrated performance. This analysis did not show that SSA and TI could explain a difference in integrated performance nor that integrated performance was related to the conditions. So, no total effect (c), nor indirect paths (ab), nor direct effect (c') were uncovered for integrated team performance. This is in line with the previous one-way independent ANOVA that did not show an effect of condition either on integrated performance. As figure 11 already illustrates no differences were found in SSA or TI between the conditions (a paths) with two one-way independent ANOVAs either.

[8] Hayes, A. F. (2012). An analytical primer and computational tool for observed variable moderation, mediation, and conditional process modelling. *Manuscript submitted for publication*

Table 7: The coefficients of the paths and their 95% confidence intervals of the mediation analysis on time performance are shown in this table. Only the intervals of the c paths do not include zero and are thus significant (*denotes p <.05).

Path	Coefficient	Lower bound	Upper bound
a_{TI}	.03	-.16	.22
a_{SSA}	-.10	-.64	.44
b_{TI}	-1.18	-14.42	12.06
b_{SSA}	1.33	-1.79	4.44
$c_{total}{}^*$	5.08	.274	9.43
$c'_{direct}{}^*$	5.25	.63	9.88

Secondly, it was investigated if time performance could be partly explained by SSA and TI. A mediation analysis which excluded condition two and only compared condition one and three on time performance as outcome variable was conducted. Condition one and three were compared because theoretically the expected difference between these conditions is the biggest and time performance differed the most between these conditions. The coefficients of this mediation analysis can be found in figure 12 and the bootstrapping confidence intervals in table 7. This mediation analysis only revealed a direct (c') and a total effect (c). So, no explanatory power of SSA and TI for time performance was found. Nonetheless, the level of team membership of the robots can explain time performance (total effect), b = 5.08, $t(20)$ = 2.44 p = .024.

5.3 Performance explained by team characteristics and teamwork concepts

In order to build a model that might be able to predict performance an exploratory linear regression was conducted that included several team characteristics. Performance was the outcome variable and average team age, driving experience (years and kilometers), game experience, team gender, and acquaintance of team members were inserted as predictor variables. In figure 13 team age and game experience are plotted against integrated team performance. The regression analysis showed that a model that included average team age and how often the team members play games could predict integrated team performance, $beta_{\text{age}}$ = -.48, $t(35)$ = -3.66, p = .001, $beta_{\text{game experience}}$ = .35, $t(35)$ = 2.70, p = .011. What can be seen is that if the team members are older performance decreases and when the team members have more game experience performance increases. These two team characteristics could explain 43.1% of the variance of the integrated team

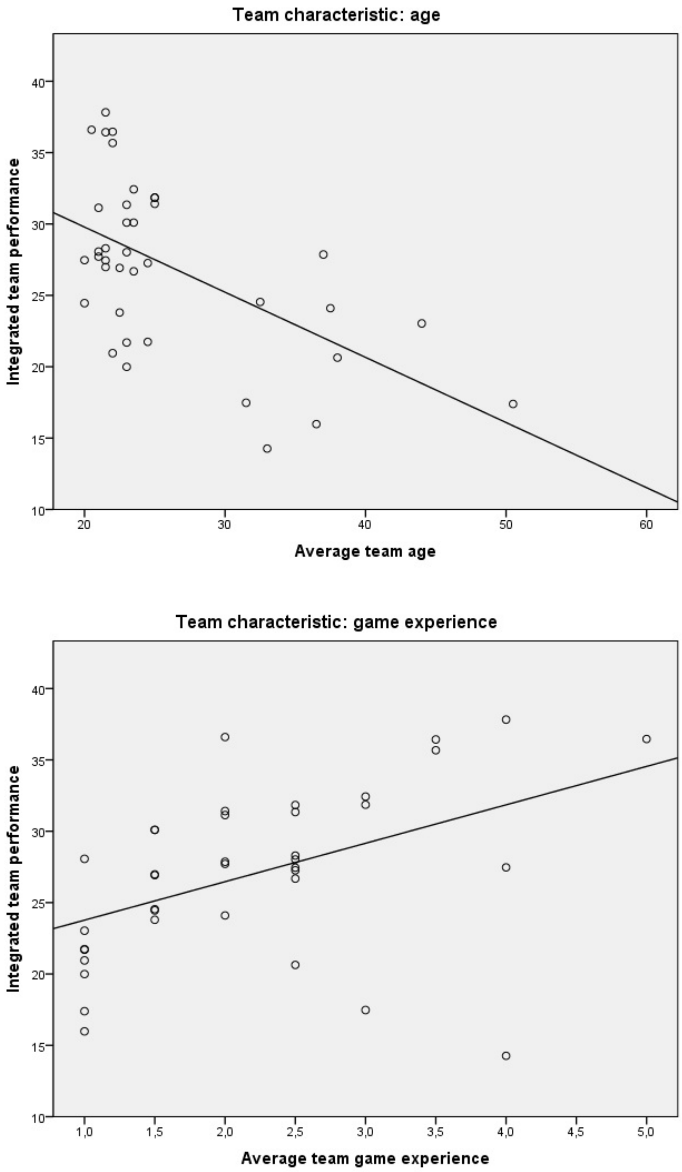


Figure 13: The team characteristics that explain integrated performance are average team age and average game experience. The formula obtained by the regression analysis was: *Integrated performance* = 32.68 + 2.06 × *Average game experience* - 0.39 × *Average team age*, with R^2 = .43

performance score, R^2 = .43, $F(2,35)$ = 13.28, p = .000. Other exploratory regressions showed that time performance can not be explained by team characteristics and none of the teamwork concepts (team trust, team coordination, team satisfaction, and team communication) were able to predict integrated team performance nor time performance. Coordination, communication, trust, and satisfaction did not even correlate with integrated performance or time performance.

5.4 Differences in teamwork between conditions

To assess teamwork the scores for coordination, communication, trust, and satisfaction were analysed per condition. The teamwork concept scores were not different between the conditions except for vocal communication. In figure 14 the vocal communication per condition is presented. There was an effect of condition on the number of total messages, $F(2,35) = 3.42$, p = .044, r = .40. Furthermore, one of the subcategories of communication was different. The number of information messages decreased over the conditions (condition 1: M = 37, SE = 3.9, condition 2: M = 32, SE = 5.6, condition 3: M = 20, SE = 2.7, $F(2,35) = 4.10$, p = .025 r = .40).

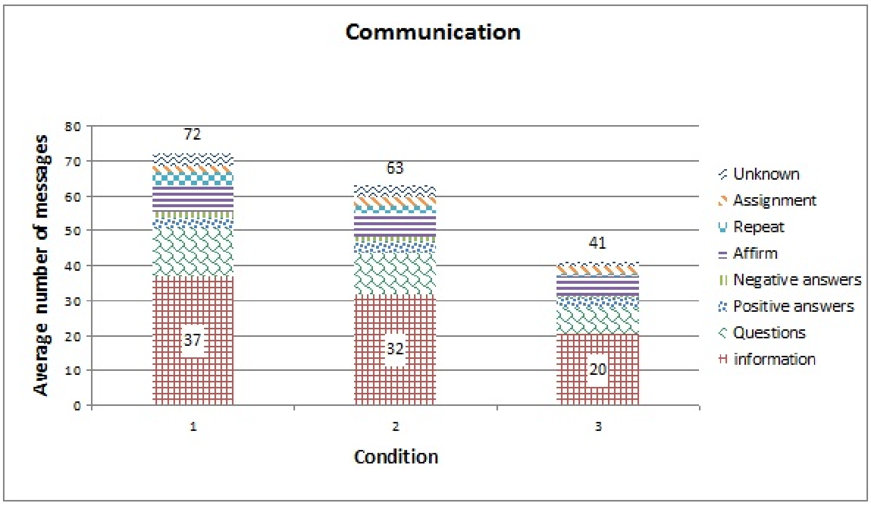


Figure 14: Vocal communication between the participants within a team. The total number of messages sent by participants differed between the conditions. In condition one it were 37 messages, in condition two it were on average 32 messages, and in condition three it were only 20 messages.

5.5 Individual performance and situation awareness

Until now, only shared situation awareness has been analysed. A disadvantage of SSA, however, is that the chance of getting a good SSA score is higher for couples who both have good situation awareness. When both participants are well aware of the environment the chance of giving the same answer to a question like "*Which color did the floor of the office had?*" is higher. When one of the team members' SA is not that good, the team's SSA score will probably be not that good either. Therefore, the difference in individual situation awareness, measured by the SPASA questionnaire, between the conditions was also investigated. In table 8 the SA scores per condition are provided, as well as the individual performance, which is a performance score that only includes the number of victims found by one person. Two one-way independent ANOVAs showed no difference in SA or individual performance between the conditions. Furthermore, a correlation analysis between individual performance and SA was done to asses the controversial relationship between these two concepts. The results showed a medium effect between individual performance and situation awareness, $r = .39$, $p = .000$.

Table 8: Individual performance measured by a score based on the number of victims found, and situation awareness based on the SPASA questionnaire. No difference were found, however, individual performance and SA do correlate $r = .39$, $p = .000$.

		Individual performance (victims)		Situation Awareness (SPASA)	
	N	*Mean*	*SD*	*Mean*	*SD*
Condition 1	26	1.60	.57	2.82	.40
Condition 2	24	1.73	.65	2.94	.30
Condition 3	26	1.52	.68	2.85	.42
	76	*ns*, $r = 0.14$		*ns*, $r = 0.14$	

6 Discussion

In order to investigate whether team performance increases when a robot is made a team player we performed a study in which the level of team membership of the robots was varied between three conditions. Results showed that time performance indeed did differ across conditions, whereas the integrated performance score did not. We expected that performance would increase when team membership of the robots increased (*hypothesis 1*). This was not what we found, we found that time performance decreased when the robots behaved more like team players. Shared situation awareness and team identification were measured in order to investigate if the difference in performance could be explained by these concepts. Results did not provide evidence for this expectation (*hypothesis 2*).

Extra analyses to explore possible models to explain performance revealed that game experience and age could predict integrated team performance. In addition the team work concepts were assessed per condition. Results showed that participants communicated less, when the level of team membership of the robots was higher. Lastly, the controversial relationship between SA and performance was investigated. The results showed a positive correlation between SA and individual performance.

In the following we will reflect in detail on a possible explanation for the decreased performance when the robots' team membership increased. After that an interpretation of the results of the exploratory models will be given. Followed by a brief discussion about the differences between the time performance and integrate performance score. Lastly, the limitations of this experiment will be reviewed and some suggestions for future work will be given.

6.1 Team membership and team performance

Team performance decreased instead of increased when the robots behaved more like a team member. A possible explanation might be that mental effort increased due to the 'extra system'. The Cognitive Task Load (CTL) model (Neerincx, 2003) distinguishes three general factors that have an effect on mental effort and performance (Grootjen, Neerincx, & van Weert, 2006). These factors are time occupied, task switches, and level of information processing. In this experiment mental effort nor the three factors were measured, so we cannot confirm that workload increased by the robotic system. However, in the following paragraphs we will reason about a difference in time occupied, task switches and level of information processing between the conditions to find a plausible explanation for the found reverse effect.

Time occupied
The time the participants were occupied during the mission was probably equal in all three conditions. The goal was to find six victims as fast as possible, so it was a continuous task not depending on events making the task more or less demanding. Participants in every condition were engaged in the mission all the time.

Task switches
On first sight the introduction of the robotic system (condition two and three) does not change the task switches; participants operate their robot, search for victims, and draw victims and interesting objects on the map. However, because more functionality, like the robot's position, is included in the interactive map, the users probably use that map more often. In condition one participants operate the robot, while searching for victims and keep track of where they are, which are kind of simultaneous processes. Whereas participants who have the interactive map available drive and fly around while searching and probably pause every now and then to check on the map where they are, resulting in more task switches.

Level of information processing
The level of information processing is based on the Skill-Rule-Knowledge framework by Rasmussen (2003, as cited in Neerincx, 2003). The Skill-Rule-Knowledge framework identifies three levels of information processing. The lowest level consists of skills that do not need a lot of information processing. Next, there are situations that trigger rule based information processing, which demand some cognitive effort to resolve an if-then rule. Lastly, the framework distinguishes a knowledge-based level for information processing. This level is highly demanding, since it involves problem analysis, knowledge, and planning a solution to deal with new situations. Controlling the robot and searching for victims is the same for all participants and requires mostly rule- and knowledge-based information processing. Drawing a victim or obstacle on the map, however, demands rule-based information processing for the interactive map, whereas only skill-based information processing is required to draw the object on the paper map. To indicate a victim on the interactive map participants have to think something like; I first have to click at the right place in the map, then I have to choose if it is a victim, an obstacle or something else, and then I can type some extra information. On the contrary, physically drawing on paper is a skill which participants did not have to think about.

The CTL-model in this experiment in short
In conclusion, the robotic system especially the interactive map probably

induced more task switches and required a higher level of information processing. According to the CTL model this results in a higher mental workload and a decrease in performance. This suggests that interaction with the robotic system was not natural enough in the sense that if interaction feels natural to humans it does not require a higher level of information processing. Furthermore, it could be argued that using the interactive map takes more time than the paper map and therefore a difference in time performance is found between the conditions. On top of that, the difference in time performance between condition two and three could be explained by the fact that the robot asked for help with determining the medical status of the victim, the first time the participants found a victim. Not only the speech of the robot took extra time, but participants gave a more elaborated description of the status of the victim, which is not included in any performance measure.

6.2 Explanatory models of team performance

Through a mediation analysis and some exploratory regressions we tried to reveal the underlying concepts of team performance. First, the mechanisms SSA and TI were investigated, followed by some team characteristics and teamwork concepts. In the following paragraph the results of those analyses will be interpreted.

SSA and TI
The second hypothesis stated that SSA and team identification would explain team performance at least to a certain extent. However, team performance did not increase when shared situation awareness or team identification increased or decreased. The mediation analysis showed that only condition (the level of team membership) had an influence on performance. The differences in performance could not be explained by SSA or team identification. Maybe other concepts that possibly differ between conditions, for instance participants' confidence in successful completion of the mission, might be able to explain team performance. Though this is a fairly wild guess.

Age and game experience
The effect of age and game experience on performance is in line with expectations, with age performance decreased and game experience increased performance. The robot task took place with a virtual environment using a game engine. In addition the robots were also controlled through typical game controllers. So, the positive effect of game experience was expected. The negative influence of age on performance was already found in an earlier study that was very similar but with a real robot (Mioch et al., 2012). That experiment showed that the age of the operator influenced the performance

and SA in a pre-set-up disaster scenario. So, the fact that we found age to negatively effect performance is in line with previous research.

Vocal communication
Lastly, the results suggested that the difference in vocal communication between the conditions could possibly explain why teams in condition three perform worse than teams in condition one. The results showed that in condition three only 20 messages were sent on average, whereas 37 messages were sent on average in condition one. Because of the robotic system, especially the interactive map, there was less need to communicate for the participants. For example, in condition three it is not necessary for participants to say through the walkie talkie that they found a victim, since they also indicate this on the shared map. So, the tendency to not share information may arise. For example, it seems less important to vocally communicate what your next step is, while the interactive map does not provide this kind of information. So a possible explanation for the decreased performance over the conditions may be this lack of communication.

6.3 Different performance measures

Time performance measured efficiency whereas the integrated performance measured effectiveness. So, it may be concluded that the level of team membership of the robots did degrade efficiency but not effectiveness. In terms of the number of team members this is a plausible result; more team members require more coordination at the expense of efficiency. Furthermore, results suggested that integrated performance could be explained by age and game experience, whereas time performance could not be explained by any team characteristics. In other words, the younger participants and more experienced gamers found the victims that were hardest to locate and received a higher integrated performance score, but were not faster in finding the first three victims. So, team characteristics influenced effectiveness, but not efficiency. That only effectiveness is influenced by team characteristics could possibly explained by the fact that experienced gamers started searching systematically. A systematic approach is likely to result in finding more difficult victims, thus a better overall performance (effectiveness). In contrast, the systematic approach does not automatically lead to quickly finding three victims; all participants had about the same chance of 'accidentally' finding the first victims. Nevertheless, above conclusions should be accepted with care, because time performance only measured the performance during the beginning of the mission, whereas integrated performance comprised the whole mission and all goals. In short, interpreting the results in terms of efficiency and effectiveness should be done with care because the metrics had a different scope.

6.4 Limitations

In section 2.4 we established the definition of a team[9], from this definition follows that collaboration is different from interaction. Interaction means at least performing an action on someone or something else (Breazeal et al., 2004). For an interaction to be labelled as collaboration, however, more is needed. For example communication, coordination, and task division. A lot of research is done in the human-robot-interaction community that is labelled as collaboration research, which actually only consists of interactions. These interaction can take place through advanced interfaces as speech recognition or gesture input. These progressive interfaces, however, do not make an interaction also collaboration (Breazeal et al., 2004).

By following the guidelines of Klein et al. (2004) we tried to make sure collaboration took place between humans and robots. However, during the experiment it became clear that it is really hard to accomplish real collaboration. For example, it is not clear if and how the advice of the robot is experienced as realistic team member behavior. The same holds for asking for help and giving assignments. The manipulation check questions mainly focused on the SA aspects and not on team member behavior, so we do not know how participants experienced the robot's behavior. In the end questionnaire, participants were asked about the robot's positive and negative aspects, only one comment was made about the speech of the robot. The rest of the comments all considered the features of the robot instead of the behavior. Apparently, the robot features struck the participants more than the behavior, which suggests that the robots' team behavior was not that obvious to the participants. Furthermore, the 'interpretation' challenge could not be implemented and the 'revealing status and intention' challenge occurred only in the beginning of the experiment, when the robots changed their 'status and tasks'. In a nutshell, it is questionable if or to what extent real collaboration between the robots and the participants occurred.

In this study, every challenge occurred only once (or in some cases two times). The single occurrence was an explicit consideration, firstly to ensure that everyone had approximately the same experience. And secondly, and maybe even more important, to prevent the participants from getting annoyed. For example, if the UAV would have said at each door that it is too wide to pass, this could have caused irritation, because participants had learned that by themselves after trying to pass a door for a few times. Furthermore, if the robot would continuously direct the participants this might have led to annoyance as well, because people dislike being bossed around by a 'stupid robot'. We wanted to avoid irritation at all costs, because irritation

[9] A team consists of "*two or more interdependent individuals performing coordinated tasks toward the achievement of specific task goals*" (Nieva, 1985)

would lead to disengagement and a decrease in performance. Therefore, we designed the experiment in a way that every challenge occurred only once during the mission. After conducting the experiment, we may conclude that this choice was not the best option, since we did not find a difference in team identification between the conditions.

6.5 Future research

Team identification has not been studied much. Shapiro et al. (2002) argue that this might be because it is irrelevant in short-term, laboratory-based groups. However, they also hope that future studies in their research field (transnational teams) will include team identification, because team identification could explain some unexpected achievement differences in certain situations. (Shapiro et al., 2002) The current experiment included team identification. However, the experiment did not show a difference in team identification between the conditions. A possible explanation could be that some teams did not know each other, whereas others did. Thus some teams were short-term, laboratory-based teams as Shapiro et al. (2002) call them. The results of this experiment showed a trend in team identification between short-term and long-term teams. Teams that knew each other scored higher on team identification (M = .28, SD = .15)[10], than teams that did not know each other (M = .38, SD = .21, $F(1{,}37) = 2.88$, p = .100, r = .27). Therefore, it is advised to include only short-term or long-term teams in future research when team identification is the subject of research, because acquaintance affects team identification.

As argued before extra workload due to the extra system might explain the decrease in performance. Therefore, it is highly recommended to include a workload measure when different robot configurations are evaluated. For example a secondary task (Gawron, 2000, section 3.1.9) or the NASA-TLX could be used. Both measurement techniques are widely used to measure workload. The NASA-TLX is a questionnaire concerning 6 underlying concepts: mental demands, physical demands, temporal demands, frustration, effort, and performance (Hart, 2006). In addition to these validated measures, a measurement dedicated to measure workload tied to a human-robot system, like the HRI-WM can be used (Yagoda, 2010). The HRI-WM includes five workload attributes: task, system, team process, team configuration, and context. So, an advantage of the HRI-WM is that this measure also includes team dynamics. However, as stated before the HRI-WM should not be used instead of previously validated subjective workload measures, such as the NASA-TLX, but in conjunction with it.

[10] note that these statistics concern the transformed team identification score (inverse, log transformation)

7 Conclusion

Robots are and will be deployed in USAR situations, because they can explore dangerous environments were people cannot go. The participants in this experiment confirmed this as valuable and useful application (end questionnaire, appendix E.10). Furthermore, USAR situations give researchers a dynamic and unknown environment to test their robots in which human and robot behavior can reveal underlying teamwork mechanisms. In this experiment three robot settings were tested ranging from the robot as a tool to the robot as a team player. The hypothesis was that team performance would increase when the robot behaves more like a team member. However, the results showed the opposite effect; the teams with the robot as a tool performed better. Possible explanations are an increase in mental workload for the humans, inter alia because interaction with the robot as a team player was not natural enough. Furthermore, it was hypothesized that SSA and team identification would be able to explain this increase in performance. However, SSA and team identification did not differ between the conditions. So the robot as a team player was not able to support better knowledge about the environment or induce higher team sense. Moreover, the decrease in performance could not be explained by SSA or team identification. Future research should, therefore, focus on other aspects of teamwork. The results of this experiment suggest that better vocal communication within a team could explain higher performance. In addition, cognitive workload should be included in future work. For example, workload should not increase with a better, smarter robot or results should be corrected for workload.

Appendices

In the appendices the Dutch versions will be given first followed by the English translations.

A Task description

De missie

Jullie zijn beide brandweermannen en zijn opgeroepen na een aardbeving. De aardbeving heeft een kantoorgebouw getroffen en in verband met naschokken mogen er nog geen mensen naar binnen. Gelukkig beschikt de brandweer over robots die wel al naar binnen kunnen.

Jullie gaan allebei een robot besturen om het getroffen kantoorgebouw te doorzoeken. De bedoeling is om zo snel mogelijk 6 slachtoffers te vinden. Daarbij is het ook van belang om een goede kaart te creëren met vrije routes en gevaarlijke objecten, zodat brandweermannen later de slachtoffers zo snel mogelijk kunnen redden. Daarnaast is het ook belangrijk om kort aan te geven hoe de slachtoffers er aan toe zijn.

De robots komen in een beschadigd kantoor terecht en kunnen vast komen te zitten. Pas dus op waar je je robot heen stuurt, want er kunnen geen mensen het gebouw in om de robot te bevrijden.

Jullie zijn samen een team. (conditie 3: Jullie zijn samen met de robots een team.) We hebben het net al even over samenwerking gehad, maar hier nog even kort wat handige puntjes:
- Houd elkaar op de hoogte
- Het is oké om elkaar opdrachtjes te geven
- Het is oké om om hulp te vragen, of hulp aan te bieden
- Bespreek even hoe je dingen aanpakt, wat jullie doel is (Dit kan tijdens de missie veranderen)
- Het gaat om jullie gezamenlijke prestatie, het maakt niet uit wie de slachtoffers vindt, het gaat erom dat jullie het samen zo snel mogelijk doen

In het kort: gezamenlijk zo snel mogelijk 6 slachtoffers vinden, denk als een brandweerman!

Omdat we het experiment wel door willen laten gaan, ook al kom je vast te zitten, bestaat er de mogelijkheid om opnieuw op te starten, wat normaal natuurlijk niet kan. Dit betekent wel dat je weer vooraan begint en ondertussen loopt de tijd gewoon door. Mocht je vast komen te zitten, roep dan de experimentleider.

English text

The mission

You are both firefighters and are called after an earthquake. The earthquake hit an office building and because of aftershocks no people are allowed to go inside. Fortunately, the fire department has robots that are allowed to go inside the building.

You are both going to control a robot control to search the destroyed office. The goal is to find 6 victims as quickly as possible. It is also important to create a good map with routes free of dangerous objects, so that firefighters can reach and save the victims as quickly as possible later on. Furthermore, it is important to briefly indicate the status of the victims.

The robots will enter a damaged office and can get stuck. So be careful where you steer your robot, because people cannot go into the building to free the robot.

You are (together with the two robots (This is added in condition 3))) a team. We already talked a bit about collaboration, but below some guidelines are listed:

- Keep each other informed
- It is okay to give each other assignments
- It is okay to ask for help or to offer help
- Discuss how you get things done, what is your goal (This may change during the mission)
- It is your collective achievement, it does not matter who finds the victims, the point is that you do it as fast as possible together

In short: find 6 victims together as quickly as possible, think like a fireman!

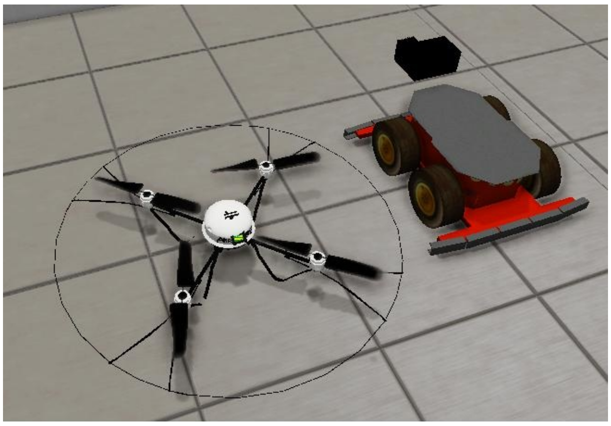
Because we want to experiment to go on, even if you get stuck, there is the possibility to restart, which is in a real situation not possible of course not. This means that you start at the beginning while time is ticking. If you get stuck, call the experimenter.

B Robot description

B.1 The robots – Condition 1

De rijdende robot = UGV (unmanned ground vehicle)
De camera van de robot zit midden boven op de robot. De camera op zichzelf kan niet bewegen. De robot heeft vier wielen en heeft moeite om verhogingen op te rijden. Aan de voor- en achterkant heeft de robot een bumper, die aan de zijkant een klein beetje uitsteekt. Hierdoor kan de robot ergens achter blijven hangen.

De vliegende robot = UAV (unmanned aerial vehicle)
De vliegende robot is vrij breed. De camera is schuin naar onder gericht. Als je de robot even niet bestuurd, blijft hij zweven op de plek waar hij was. Ook kan de robot zichzelf stabiliseren. Als je schuin vliegt en je doet even niets, zal hij vanzelf weer horizontaal gaan zweven.



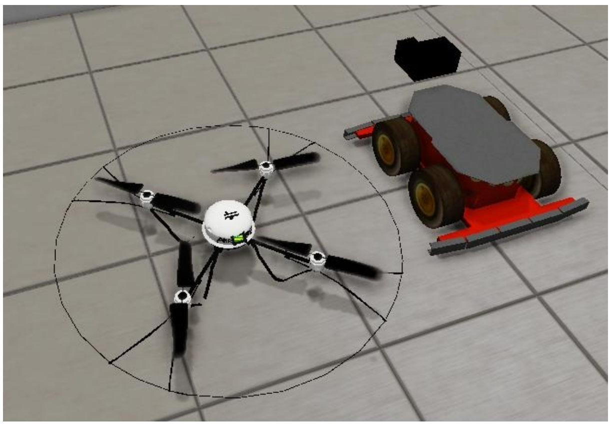
English text
The driving robot = UGV (unmanned ground vehicle)
The camera is positioned at the center on top of the robot. The camera itself can not move. The robot has four wheels and has difficulty to drive on slopes and platforms. On the front and back, the robot has a bumper on the sides that is sticking out a little bit. As a result, the robot can get stuck behind things.

The flying robot = UAV (unmanned aerial vehicle)
The flying robot is quite large. The camera is aimed/directed downward. If you do not operate/control the robot, it remains floating on the spot where it was. The robot can also stabilize itself. So, if you fly diagonally and you let go off the controller, it will automatically return to float horizontally.

B.2 The robots – Condition 2

De rijdende robot = UGV (unmanned ground vehicle)
De camera van de robot zit midden boven op de robot. De camera op zichzelf kan niet bewegen. De robot heeft vier wielen en heeft moeite om verhogingen op te rijden. Aan de voor- en achterkant heeft de robot een bumper, die aan de zijkant een klein beetje uitsteekt. Hierdoor kan de robot ergens achter blijven hangen. De robot heeft een aantal sensors om niet zichtbare voor mensen giftige gassen te meten, zoals koolstofmonoxide. De robot kan dit soort gassen, zijn status en taak veranderen op de overzichtskaart.

De vliegende robot = UAV (unmanned aerial vehicle)
De vliegende robot is vrij breed. De camera is schuin naar onder gericht. Als je de robot even niet bestuurd, blijft hij zweven op de plek waar hij was. Ook kan de robot zichzelf stabiliseren. Als je schuin vliegt en je doet even niets, zal hij vanzelf weer horizontaal gaan zweven. De robot heeft een aantal sensors om niet zichtbare voor mensen giftige gassen te meten, zoals koolstofmonoxide. De robot kan dit soort gassen, zijn status en taak veranderen op de overzichtskaart.



English text
The driving robot = UGV (unmanned ground vehicle)
The camera is positioned at the center on top of the robot. The camera itself can not move. The robot has four wheels and has difficulty to drive on slopes and platforms. On the front and back, the robot has a bumper on the sides that is sticking out a little bit. As a result, the robot can get stuck behind things. The robot has a number of sensors in order to measure toxic gases which are not visible for humans, such as carbon monoxide. The robot can add these gases, its status and role change in the overview map.

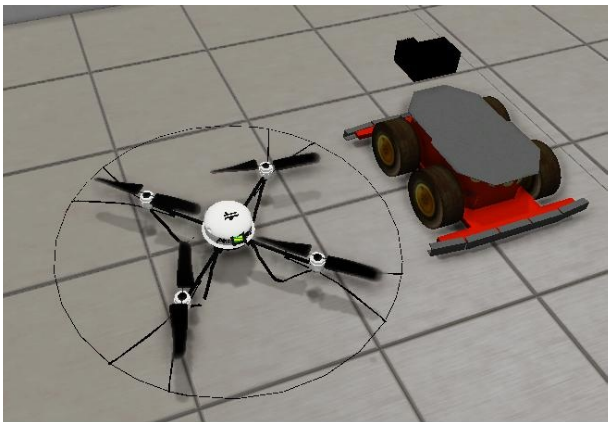
The flying robot = UAV (unmanned aerial vehicle)
The flying robot is quite large. The camera is aimed/directed downward. If

you do not operate/control the robot, it remains floating on the spot where it was. The robot can also stabilize itself. So, if you fly diagonally and you let go off the controller, it will automatically return to float horizontally. The robot has a number of sensors in order to measure toxic gases which are not visible for humans, such as carbon monoxide. The robot can add these gases, its status and role change in the overview map.

B.3 The robots – Condition 3

De rijdende robot = UGV (unmanned ground vehicle)
De camera van de robot zit midden boven op de robot. De camera op zichzelf kan niet bewegen. De robot heeft vier wielen en heeft moeite om verhogingen op te rijden. Aan de voor- en achterkant heeft de robot een bumper, die aan de zijkant een klein beetje uitsteekt. Hierdoor kan de robot ergens achter blijven hangen. De robot heeft een aantal sensors om niet zichtbare voor mensen giftige gassen te meten, zoals koolstofmonoxide. De robot kan dit soort gassen, zijn status en taak veranderen op de overzichtskaart. De robot is ook in staat om slachtoffers te herkennen en kan bepalen of hij door een opening past of niet, soms zal de robot daar iets over zeggen.

De vliegende robot = UAV (unmanned aerial vehicle) De vliegende robot is vrij breed. De camera is schuin naar onder gericht. Als je de robot even niet bestuurd, blijft hij zweven op de plek waar hij was. Ook kan de robot zichzelf stabiliseren. Als je schuin vliegt en je doet even niets, zal hij vanzelf weer horizontaal gaan zweven. De robot heeft een aantal sensors om niet zichtbare voor mensen giftige gassen te meten, zoals koolstofmonoxide. De robot kan dit soort gassen, zijn status en taak veranderen op de overzichtskaart. De robot is ook in staat om slachtoffers te herkennen en kan bepalen of hij door een opening past of niet, soms zal de robot daar iets over zeggen.



English text

The driving robot = UGV (unmanned ground vehicle)

The camera is positioned at the center on top of the robot. The camera itself can not move. The robot has four wheels and has difficulty to drive on slopes and platforms. On the front and back, the robot has a bumper on the sides that is sticking out a little bit. As a result, the robot can get stuck behind things. The robot has a number of sensors in order to measure toxic gases which are not visible for humans, such as carbon monoxide. The robot can add these gases, its status and role change in the overview map. The robot is also able to recognize victims, it can determine whether or not it fits through a passage, and sometimes the robot will say something about this.

The flying robot = UAV (unmanned aerial vehicle)

The flying robot is quite large. The camera is aimed/directed downward. If you do not operate/control the robot, it remains floating on the spot where it was. The robot can also stabilize itself. So, if you fly diagonally and you let go off the controller, it will automatically return to float horizontally. The robot has a number of sensors in order to measure toxic gases which are not visible for humans, such as carbon monoxide. The robot can add these gases, its status and role change in the overview map. The robot is also able to recognize victims, it can determine whether or not it fits through a passage, and sometimes the robot will say something about this.

C Walkie talkie usage

U kunt met de andere deelnemer communiceren en overleggen via de walkietalkie. Voordat u begint met praten drukt u de knop aan de zijkant in. Als u klaar bent met praten, laat u de knop los. Zolang u de knop heeft ingedrukt kan de ander niets tegen u zeggen.

Om efficiënt met elkaar te communiceren via de walkietalkies is het handig om de volgende termen te gebruiken.

Out	Ik ben klaar met praten en verwacht geen antwoord
Over	Ik ben klaar met praten en luister naar je antwoord
Roger	Ik heb begrepen wat je net zei
Say again	Herhaal alsjeblieft je laatste bericht
Affirm	Ja
Negative	Nee

English text
You can communicate with the other participants via the walkie-talkie. Before you start talking, press the button on the side. When finished talking, release the button. As long as the button is pressed, the other person cannot speak to you.

To effectively communicate with each other via walkie-talkies it is useful to use the following terms.

Out	I finished talking and expect no answer
Over	I finished talking and listen to your answer
Roger	I understand what you just said
Say again	Please repeat your last message
Affirm	Yes
Negative	No

D Text robot speech

Text robot speech in Dutch	**Translation**	**Played when (trigger)**	**Belonging to challenge**
De deuropening is erg smal, ik zou een andere weg zoeken	The doorway is very narrow, I would try to find another way	One of the first small doorways the UAV tries to enter	Goal negotiation
Als je op de pallets afrijdt kun je vast komen te zitten, ik zou een andere weg zoeken	If you are driving down to the pallets it is possible to get stuck, I would try to find another way	When UGV drives to the pallets	Goal negotiation
De ruimte naast het gat is erg smal, ik zou een andere weg zoeken	The space next to the hole is very narrow, I would try to find another way	When UGV drives to the hole	Goal negotiation
Doorwerken er wachten slachtoffers op ons	Hurry up, there are victims waiting for us	After 12 minutes to UGV and UAV	Directability
Rijd iets langzamer *(niet gebruikt)*	Drive a little slower *(not used)*	When UGV drives too fast *(not used)*	Directability
Vlieg iets langzamer *(niet gebruikt)*	Fly a little slower *(not used)*	When UAV flies too fast *(not used)*	Directability
Ik kan de medische status van het slachtoffer niet bepalen, kun jij me daarmee helpen?	I can not determine the medical status of the victim, can you help me with that?	When UGV finds it's first victim. When UAV finds it's first victim	Collaboration; ask for help

E Questionnaires

E.1 Demographics

Questions in Dutch	Translation
• Hoe oud bent u?	How old are you?
• Wat is uw geslacht?	What is your gender?
• Kent u de andere participant (deelnemer)?	Do you know the other participant?
• Hoe kent u de andere participant?	If yes, how do you know him/her?
• Heeft u een rijbewijs?	Do you have a driver's license?
• Hoeveel jaar heeft u uw rijbewijs al?	If yes, for how long have you had your driver's license?
• Hoeveel kilometerst rijdt u ongeveer gemiddeld per jaar?	If yes, how many kilometers do you drive per year (estimate)?
• Hoeveel ervaring heeft u met de onderstaande soort gamecontrollers? (er waren 3 plaatjes van verschillend gamecontrollers te zien)	How much experience do you have the following type of game controller? (3 pictures of several game controllers were shown)
• Heeft u thuis een spelcomputer?	Do you have gameconsole at home?
• Welke spelcomputer heeft u? (antwoordmogelijkheden: Xbox 360, Playstation 3, Wii, PC, other)	Which console do you have? (answer options: Xbox 360, Playstation 3, Wii, PC, other)
• Hoe vaak heeft u voor u mee deed aan dit experiment een robot bestuurd?	How often have you operated robots before?
• Hoe vaak speelt u first person shooter computer games?	How often do you play first person shooter computer games?

E.2 Situation awareness questions

SPASA questions

	Questions in Dutch	Translation
•	Het was gemakkelijk om te weten waar ik was en naar welke kant ik keek.	It was easy to know exactly where I was and towards where I was looking at.
•	Het was gemakkelijk om obstakels te identificeren en te ontwijken.	It was easy to identify and avoid obstacles.
•	Het was gemakkelijk om tijdsaspecten bij te houden.*	It was easy to keep track of time aspects.*
•	Het was gemakkelijk om bij te houden welk gedeelte al doorzocht was.	It was easy to keep track of the area covered.
•	Het was gemakkelijk om de slachtoffers die gevonden waren bij te houden.	It was easy to keep track of the victims that were located.
•	Het was gemakkelijk om te voorspellen wat er ging gebeuren.*	It was easy to predict what would happen next.*
•	Het was gemakkelijk om de missiedoelen te volgen.	It was easy to follow the goals of the tasks.
•	Het was gemakkelijk om mijn gang van zaken te veranderen, omdat ik me zeker voelde over de informatie die ik van de	It was easy to change my course of action because I felt confident about the information provided by the robot.
•	De informatie werd op een snelheid aangeboden die ik gemakkelijk kon waarnemen.	The information was provided at a rate I could easily perceive.
•	Ik had een goed begrip van de complete situatie.	I was able to have a good understanding of the holistic (global) situation

* This item was deleted to increase Cronbach's Alpha

E.3 Situations awareness questions

'TNO' questions

	Questions in Dutch	Translation
•	In het algemeen had ik een goed beeld van de omgeving	In general, I had a good idea of the environment.
•	In het begin was het moeilijk om een goed beeld van de omgeving te verkrijgen	In the beginning, it was difficult to build a picture of the environment.
•	Soms wist ik niet wat er gaande was in de omgeving	Sometimes, I lost track of what was going on in the environment.
•	Ik wist de hele tijd waar de robot was	I knew the whole time where the robot was.
•	Ik wist de hele tijd in welke richting de robot bewoog	I knew the whole time in which direction the robot moved.
•	Als ik een object zag dan wist ik waar het object zich ten opzichte van mij bevond	When I saw an object, I knew its position relative to me.
•	Als ik een object zag dan wist ik waar het object zich in de ruimte bevond (bijv. op de kaart)	When I saw an object, I knew its position in space (e.g. on the map).

E.4 Shared situation awareness questions

	Questions in Dutch	Translation
•	Welke kleur heeft de vloer van het gebouw?	What color is the floor of the building?
•	Wat bevond zich rechts van de UGV op het eind van de missie?	Which was located to the right of the UGV at the end of the mission?
•	Welke kleur heeft het plafond van het gebouw?	What color is the ceiling of the building?
•	Welke kleur hebben de meeste stoelen in het kantoor?	What color are the most seats in the office?
•	Welke kleur hebben de meeste muren in het kantoor?	What color are most walls in the office?
•	Hoeveel kamers hebben jullie samen doorzocht?	How many rooms have you searched together?
•	Hoeveel kamers heeft de rijdende robot (UGV) doorzocht?	How many rooms did the driving robot (UGV) search?
•	Hoeveel kamers heeft de vliegende robot (UAV) doorzocht?	How many rooms did the flying robot (UAV) search?
•	Welk percentage van het gebouw hebben jullie samen doorzocht?	What percentage of the building have you searched together?
•	Hoe breed is het kantoor? (meter)	How wide is the office?
•	Hoeveel tijd hebben jullie voor de missie nodig gehad? (minuten)	How much time do you have for the mission?
•	Hoeveel tijd zouden jullie nodig hebben voor het doorzoeken van het gehele gebouw?	How much time would you need to search the entire building?

E.5 Team identification questions

	Questions in Dutch	Translation
•	Ik voel me emotioneel verbonden met mijn team	I feel emotionally attached to my team
•	Ik voel me emotioneel verbonden met de andere persoon in mijn team	I feel emotionally attached to the other participant
•	Ik voel me emotioneel verbonden met de robots in mijn team	I feel emotionally attached to the robots
•	Ik voel in sterke mate dat ik tot het team behoor	I feel a strong sense of belonging to my team
•	Ik voel me in sterke mate betrokken bij de andere persoon in mijn team	I feel a strong sense of attachment to the other participant
•	Ik voel me in sterke mate betrokken bij de robots in mijn team	I feel a strong sense of attachment to the robot
•	Ik beschouw de problemen van het team als mijn eigen problemen	I feel as if the team's problems are my own problems
•	Ik beschouw de problemen van de andere persoon in mijn team als mijn eigen problemen	I feel as if the problems of the other participant are my own problems
•	Ik beschouw de problemen van de robots in mijn team als mijn eigen problemen	I feel as if the robot's problems are my own problems
•	Ik voel me onderdeel van een familie binnen het team	I feel like part of the family in my team

E.6 Manipulation check questions questions

	Questions in Dutch	Translation
•	Tijdens de missie had ik een goed overzicht van waar de andere deelnemer was	During the task I had a good overview of where my co-worker was
•	Tijdens de missie had ik een goed overzicht van waar de robot was	During the task I had a good overview of where the robots were
•	Tijdens de missie was ik in staat om te bepalen of de acties van de andere deelnemer invloed hadden op teamwork en de doelstellingen	During the task I was able to determine whether actions of the other participant affected teamwork and team goals
•	Tijdens de missie was ik in staat om te bepalen of de acties van de robot invloed hadden op teamwork en de doelstellingen	During the task I was able to determine whether actions of the robots affected teamwork and team goals
•	Ik was terughoudend met het zoeken van contact met de andere deelnemer, want ik was bang dat ik hem/haar zou onderbreken*	I was reluctant to seek contact with the other participant because I was worried that I would interrupt my co-worker*
•	Tijdens de missie had ik geen idee wat de andere deelnemer aan het doen was	During the experiment I had no idea what the other participant was doing
•	Tijdens de missie had ik geen idee wat de robot aan het doen was	During the experiment I had no idea what the robots were doing

* This item was deleted to increase Cronbach's Alpha

E.7 Coordination and communication questions

	Questions in Dutch	Translation
•	De werkzaamheden werden nauw gecoordineerd	The work done on tasks was closely harmonized
•	Ons team voorkwam dubbel werk	Our team avoided duplication of effort
•	Aaneengesloten taken werden goed gecoördineerd in ons team	Connected tasks were well coordinated in our team
•	Er was sprake van regelmatige communicatie binnen ons team	There was frequent communication within our team
•	Er was sprake van intensieve communicatie binnen ons team	There was intensive communication within our team

E.8 Trust questions

	Questions in Dutch	Translation
•	Ik kan rekenen op mijn team	I would be able to count on my team
•	Ik kan rekenen op de robots in mijn team	I would be able to count on the robots in my team
•	Ik kan rekenen op de andere persoon in mijn team	I would be able to count on the other person in my team
•	Als brandweerman/ brandweervrouw zou ik gebruik maken van de robots	As a firefighter, I would use the robots
•	Ik kan vertrouwen op mijn team	I can trust my team
•	Ik kan vertrouwen op de robots in mijn team	I can trust the robots in my team
•	Ik kan vertrouwen op de andere persoon in mijn team	I can trust the other person in my team
•	Ik ben voorzichtig met mijn team	I am wary of my team
•	Ik ben voorzichtig met de robots in mijn team	I am wary of the robots in my team
•	Ik ben voorzichtig met de andere persoon in mijn team	I am wary of the other peron in my team
•	Ik neem aan dat mijn team goed functioneert	I can assume my team will work properly
•	Ik neem aan dat de robots in mijn team goed functioneren	I can assume that the robots in my team will work properly
•	Ik neem aan dat de andere persoon in mijn team goed functioneert	I can assume the other person in my team will work properly
•	Ik zou het vinden van slachtoffers toevertrouwen aan mijn team	I would entrust my team to find victims
•	Ik zou het vinden van slachtoffers toevertrouwen aan de robots in mijn team	I would entrust the robots in my team to find victims
•	Ik zou het vinden van slachtoffers toevertrouwen aan de andere persoon in mijn team	I would entrust the other person to find victims

E.9 Satisfaction questions

	Questions in Dutch	Translation
•	Ik ben tevreden met het teamproces	I am satisfied with the team process
•	Ik ben tevreden met de resultaten van het team	I am satisfied with the results of the team
•	Ik ben tevreden met de andere leden van mijn team	I am satisfied with the other members of my team
•	Ik ben tevreden met de andere persoon in mijn team	I am satisfied with the other person in my team
•	Ik ben tevreden met de robots in mijn team	I am satisfied with the robots in my team
•	In het algemeen ben ik tevreden met de samenwerking binnen dit team	In general I am satisfied with the cooperation within the team

E.10 End questions

	Questions in Dutch	Translation
•	Wat is uw mening (positief of negatief) over de robot? Geef alstublieft ook aan waarom u dit vindt.	What is your opinion about the robot and why?
•	Geef 3 aspecten aan die verbeterd kunnen worden aan de robot.	Please give three aspects that could be improved (if possible)
•	Als u nog andere opmerkingen heeft, dan kunt u ze hieronder noteren.	If you have any other comments, please write them down below.

References

Ashforth, B., & Mael, F. (1989). Social identity theory and the organization. *Academy of management review*, *14*(1), 20–39.

Aubert, B., & Kelsey, B. (2003). Further understanding of trust and performance in virtual teams. *Small Group Research*, *34*(5), 575–618.

Baker, D., & Salas, E. (1992). Principles for measuring teamwork skills. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(4), 469–475.

Baron, R., & Kenny, D. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173-1182.

Barreto, M., & Ellemers, N. (2000). You cannot always do what you want: Social identity and self-presentational determinants of the choice to work for a low status group. *Personality and Social Psychology Bulletin*, *26*(8), 891–906.

Bolstad, C., Cuevas, H., Gonzalez, C., & Schneider, M. (2005). Modeling shared situation awareness. In *Paper presented at the 14th conference on behavior representation in modeling and simulation (BRIMS).* Los Angles, CA.

Bradshaw, J., Feltovich, P., Johnson, M., Breedy, M., Bunch, L., Eskridge, T., … van Diggelen, J. (2009). From tools to teammates: Joint activity in human-agent-robot teams. In M. Kurosu (Ed.), *Human centered design* (Vol. 5619, p. 935-944). Springer Berlin / Heidelberg.

Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Kidd, C., Lee, H., … Chilongo, D. (2004). Tutelage and collaboration for humanoid robots. *International Journal of Humanoid Robotics*, *1*(2), 315–348.

Bruemmer, D., Dudenhoeffer, D., & Marble, J. (2002). Dynamic autonomy for urban search and rescue. In *Proceedings of the AAAI mobile robot workshop.* Edmonton, Canada.

Casper, J., & Murphy, R. (2003). Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *33*(3), 367–385.

De Greef, D., T.E., Van der Kleij, R., Brons, L., Brinkman, W., & Neerincx, M. (2011). Observability displays in multi-teams. In *10th international conference on naturalistic decision making.* S. M. Fiore and M. Harper-Sciarini, Eds.

Drury, J., Scholtz, J., & Yanco, H. (2003). Awareness in human-robot interactions. In *IEEE conference on systems, man and cybernetics* (Vol. 1, pp. 912–918).

Endsley, M. (1988). Design and evaluation for situation awareness enhancement. In *Human factors and ergonomics society annual meeting* *proceedings* (Vol. 32, pp. 97–101).

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64.

Endsley, M., & Kiris, E. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(2), 381–394.

Fong, T., Thorpe, C., & Baur, C. (2003). Collaboration, dialogue, human-robot interaction. In *10th international symposium of robotics research* (pp. 255–266). Lorne, Victoria, Australia: Springer.

Gatsoulis, Y., Virk, G., & Dehghani-Sanij, A. (2010). On the measurement of situation awareness for effective human-robot interaction in teleoperated systems. *Journal of Cognitive Engineering and Decision Making*, *4*(1), 69.

Gawron, V. (2000). *Human performance measures handbook*. CRC Press.

Grootjen, M., Neerincx, M., & van Weert, J. (2006). Task-based interpretation of operator state information for adaptive support. In *Foundations of augmented cognition,* (pp. 236–242).

Guzzo, R., & Shea, G. (1992). Group performance and intergroup relations in organizations. In M. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology.*

Hart, S. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, pp. 904–908).

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., Onal, E., & Endsley, M. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing*, *10*(4), 409–430.

Kenrick, D., Neuberg, S., & Cialdini, R. (2010). *Social pscychology, goals in interaction.* Pearson.

Klein, G., Woods, D., Bradshaw, J., Hoffman, R., & Feltovich, P. (2004). Ten challenges for making automation a team player in joint human-agent activity. *Intelligent Systems, IEEE*, *19*(6), 91–95.

Mioch, T., Smets, N., & Neerincx, M. (2012). Predicting performance and situation awareness of robot operators in complex situations by unit task tests. In *ACHI 2012, the fifth international conference on advances in computer-human interactions* (pp. 241–246).

Murphy, R. (2004). Human-robot interaction in rescue robotics. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *34*(2), 138–153.

Neerincx, M. (2003). Cognitive task load design: model, methods and examples. *Handbook of Cognitive Task Design*, 283–305.

Nieva, V. (1985). *Team dimensions: Their identity, their measurement and their relationships* (Tech. Rep.). DTIC Document.

Riley, J., & Strater, L. (2006). Effects of robot control mode on situation awareness and performance in a navigation task. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, pp. 540–544).

Salas, E., Prince, C., Baker, D., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 123–136.

Scholtz, J., Antonishek, B., & Young, J. (2003). *Evaluation of operator interventions in autonomous off-road driving* (Tech. Rep.). DTIC Document.

Shapiro, D., Furst, S., Spreitzer, G., & Von Glinow, M. (2002). Transnational teams in the electronic age: are team identity and high performance at risk? *Journal of Organizational Behavior*, *23*(4), 455–467.

Smets, N., & Neerincx, M. (2011). *Domain analysis dr 5.1.1* (Tech. Rep.). TNO rapport.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction* (pp. 33–40).

Uhlarik, J. (2002). *A review of situation awareness literature relevant to pilot surveillance functions* (Tech. Rep.). DTIC Document.

Van Der Vegt, G., & Bunderson, J. (2005). Learning and performance in multidisciplinary teams: The importance of collective team identification. *The Academy of Management Journal*, *48*(3), 532–547.

Van Vugt, M., & Hart, C. (2004). Social identity as social glue: the origins of group loyalty. *Journal of Personality and Social Psychology*, *86*(4), 585-598.

Wickens, C. (2000). The trade-off of design for routine and unexpected performance: Implications of situation awareness. In M. Endsley & D. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 211–225). Mahwah, NJ: Lawrence Erlbaum Associates.

Wickens, C., Lee, J., Liu, Y., & Gordon Becker, S. (2004). *An introduction to human factors engineering* (2nd ed.). New Jersey: Pearson Education.

Yagoda, R. (2010). Development of the human robot interaction workload measurement tool (hri-wm). In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 54, pp. 304–308).

Yagoda, R., & Coovert, M. (2011). How to work and play with robots: An approach to modeling human–robot interaction. *Computers in Human Behavior*, *28*(1), 60–68.