SEAT Actor-only Team Try-out Experimentation of July 2008

- SEAT Report 2008-
(Final version: 19/02/2009)

Rianne Gouman *

Marten Kampman **

Masja Kempen *

Roelof Kuijpers ***

Niek Wijngaards *

* D-CIS Lab / Thales Research & Technology Netherlands
P.O. Box 90, NL 2600 AB, Delft, The Netherlands
E-mail: {rianne.gouman, Masja.kempen, niek.wijngaards}@icis.decis.nl
Web: www.decis.nl

** Internship at D-CIS Lab / Thales Research & Technology Netherlands
Institution: Universiteit Utrecht, Department of Information and Computing Science
E-mail: marten.kampman@cs.uu.nl
Web: www.cs.uu.nl

*** Internship at D-CIS Lab / Thales Research & Technology Netherlands
Institution: Universiteit Utrecht, Department of Philosophy
E-mail: roelof.kuijpers@phil.uu.nl
Web: www.phil.uu.nl

Abstract

This white paper contains the report of a first try-out experiment regarding assessing the effectiveness of actor-agent teams within the SEAT project in the CDM cluster in the ICIS research program. This document describes the background, experimental design and measurements and analysis of a pilot experiment conducted in July 2008 with an actor-only team.

To summarize our major findings briefly:

- RISK is promising as a tool for AAT experimentation purposes, but still requires major improvements (e.g. in its stability) to be suitable as reliable tool for experimentation.
- The experimental setup was fine, although the tutorial could have been more inter-active.
- The participants were positive about the experiment and were willingly to seriously invest their time in the scenarios and questionnaires.
- The three scenarios were practical and usable for this experiment. Through the experiment we gained more insight in the requirements for scenarios.
- The results of measurements are to be further explored together with the results of actor-agent and agent-only experiments in order to determine comparative performance measures.
- A questionnaire was used to assess the participants' opinions on RISK and the experiment. Although the questionnaire can still be refined it already gives an idea of the subjective experiences of the participants.

Contents:

1 Introduction

Since software agents became common in artificial intelligence, they have been the focus of much multi-disciplinary research. One of those research domains is the relation between agents and humans. Some examples of research foci within this research domain are: human agent cooperation, 'social agents', and personal assistant. All those topics have their own specific research challenges: human-agent cooperation, for example, looks into: what form could such a cooperation best take within a specific context: should agents be directed by a human leader, can humans accept orders from agents, do agent-human cooperations work best as mixed initiatives? In a sense, these and other research approaches all explore the boundaries of human-agent interactions. Such a human-agent interaction research domain is of great importance if humans are to live and communicate with agents in the near future as can be expected given the outlook on innovations that we now have.

In Delft, The Netherlands, the Delft Cooperation on Intelligent Systems (D-CIS Lab) multi-disciplinary research consortium combines Human Sciences and Information & Communication Technologies to explore human-agent synergies. D-CIS Lab is an open research partnership of Thales Nederland, the Delft University of Technology, the University of Amsterdam and the Netherlands Organization for Applied Scientific Research TNO. D-CIS Lab research focuses on systems-of-systems, consisting of humans and artificial systems involved in collaborative decision making under chaotic circumstances. One of the major application domains for this research is crisis management, in which decision making takes place under time pressure, is based on uncertain, incomplete and conflicting information, consequences of decisions cannot be foreseen, and the environment changes in unpredictable ways.

One of D-CIS Lab's main projects is the Interactive Collaborative Information Systems research project (ICIS). ICIS is to develop better techniques for making complex information systems more intelligent and supportive in decision making situations. Research within the ICIS project is versatile: its sub-projects focus on a wide range of themes ranging from architectures, MAS coordination, situational awareness, decision making, and human-agent teaming, to mention just a few. Within the SEAT sub-project Thales Research & Technology Netherlands is looking into Sustained Effectiveness of Actor-Agent Teams; a research theme that is based on the assumption that human actors and agents can productively and closely cooperate in teams (or: small groups), and that they will do so in the (nearby) future. Part of SEAT's current research is Actor-Agent Teaming (AAT) experimentation.

In this context a first major AAT try-out experiment was conducted on July 22, 2008. This document describes SEAT's aims with Actor-Agent-Team experimentation, the use of the RISK simulator, the setup of the try-out experiment, the first results, and the challenges that lie still ahead. Because of the small scale of the 'experiment', the experiment's explorative character and instabilities in the RISK simulator SEAT's initial experiment can best be regarded as a 'try-out experiment'. It should be understood as an exploration of tools and techniques to experiment with actor-agent teaming. As such this document contains no firm conclusions about actor-agent teams. The results we envisioned and obtained with this try-out experiment in July 2008 are insights in AAT experiment tooling, experimental design, measurement and analysis.

2 SEAT and Actor-Agent Team experimentation

The SEAT project focuses on Sustained Actor-Agent Team Effectiveness. SEAT's research is driven by the expectation that actors and agents can form a team that works towards a shared goal. This expectation is based on the following 3 assumptions:

- **Software agents can be designed in such a way that they can become part of a (human) team**, in other words, can be integrated into an actor-agent team. According to the definition of an actor-agent team in (Kempen, *et al.* (2007)) agents in an AAT are aware of the team's shared goal, have an understanding of sub-goals and tasks, have an awareness about their team members, can execute their own tasks in such a way that they contribute to the team's overall effectiveness and performance, and can interact in such away that they support the team's performance.

- **Such agents can meaningfully contribute to the effectiveness of a human team**, if agents are to be integrated in such a way in a human team that an AAT will be realised.

- As the effectiveness and performance of human teams is measurable, **the effectiveness & performance of an AAT is also measurable**.

It is the effectiveness of AATs that is of interest for SEAT's research, as becomes clear from its main research question:

How to sustain effectiveness in actor-agent teams in dynamic environments?

Thus, two main SEAT research objectives are:

- the understanding and measurement of team effectiveness and performance
  - e.g.: how can team effectiveness be measured in an experimental setting? How can actor-agent teaming be measured? What research methods and techniques can be used to experiment with actor-agent teaming?
- team behavior for human actors only, for agents and for actor-agent teams
  - e.g.: how do humans behave as team players who have to perform in a crisis management environment? can agents be integrated into a human team as reliable team players? Which social factors contribute to optimal team performance? How can agents be optimized to make them team good team players?

It is not evident how the effectiveness of an individual or team is to be measured, let alone the effectiveness of agents and AATs. As becomes clear AAT research looking into effectiveness and performance is challenging. It requires:

- an AAT (whether simulated or real-world)
- an environment in which experiments with the AAT can be conducted
- a methodology for AAT experimentation that contains
  - relevant scenario's for the experiments
  - measurement tools
  - hypothesis
  - operationalization of the research questions
- data analysis

These research objectives require a setting to experiment with actor-agent teaming and performance measurement. For research on effectiveness of actor-agent teams SEAT makes use of the RISK simulator, which will be further explained in the next chapter.

3 RISK

3.1 RISK Characteristics

- **3D-visualisation**
  The RISK Simulator (release 3.1) (http://forge.decis.nl/projects/risk) is a discrete event simulator in which participants / players can control a character (e.g., a human shaped 'avatar') in a 3D-world. The 3D-visualisation is at this point rather abstract. The graphics within the user interface can not be compared to those within current pc games, but are in our opinion sufficient to take up the role of a digital alter ego in RISK, and perform basic tasks within crisis management scenarios.

- **Time**
  Time in RISK progresses in discrete time-ticks. Each time-tick is about 5 seconds of real-world time. Although this might give human users the feeling of a small 'delay' in their actions, it gives them time to use the communication mechanisms in RISK (sending text messages).

- **Roles and actions**
  The prototype-world 'Demoville' can be used and modified, but new worlds can also be created. A world in RISK typically contains streets, houses, other buildings, trees, grass, cars, fences, fires, and other characters (such as victims). Humans can take up the role of paramedic of firefighter by connecting a graphical user interface (GUI) to a character. Paramedics can perform 'First aid' actions and (when victims have been first aided) they can 'Stabilize' victims. Firefighters can also perform First aid on victims but they can not heal victims further (i.e., not use stabilize). Firefighters have the ability to extinguish fires, which is impossible for paramedics to do (see for an overview table 1 hereafter). In order perform such actions the objects (fires or victims) have to be selected ('clicked on'). By selecting a fire or victim the status of this object (very ill victims, small fire, etc.) immediately becomes clear.

Overview of possible actions per role in RISK:	
Paramedic	**Firefighter**
Observe	Observe
Observe far away	Observe far away
Examine a specific entity	Examine a specific entity
Locomote (walk)	Locomote (walk)
Chat	Chat
First Aid	First Aid
Stabilize	-
-	Extinguish fire

Table 1: Possible actions per role in RISK

- **Characters**
  Through a human brain connection in the graphical user interface (see Appendix A for an explanation of the Human Brain Connector) players can let their character walk, observe and respond to fires and victims. When multiple players are logged on to human brain connectors each player controls its own character and is able to encounter the other characters in the RISK world. During the development of RISK the characters have been given names: 3 firefighters-characters are: Roelof, Rob & Arne. 3 paramedic-characters are Thomas, Riccardo, and Marten. The

characters are recognizable by their names depicted above their heads. In the RISK World Editor new characters can be created, but for this experiment the 6 present characters were used.

- **Communication**
  Within RISK all players can communicate with each other by means of text messages that can be written into a text /typing box in their GUI, and then can be send to one, several or all other players. All messages that are sent will be depicted in the communication box above the typing box. As such, RISK provides the SEAT researchers with a world in which can be experimented with players who have to perform crisis management related tasks in teams. Artificial agents can be incorporated into the scenario in RISK to investigate the workings of an actor-agent team: team formation and behavior, team performance and effectiveness.

- **User interfaces**
  RISK has several main user interfaces that are depicted here below in figure 1:

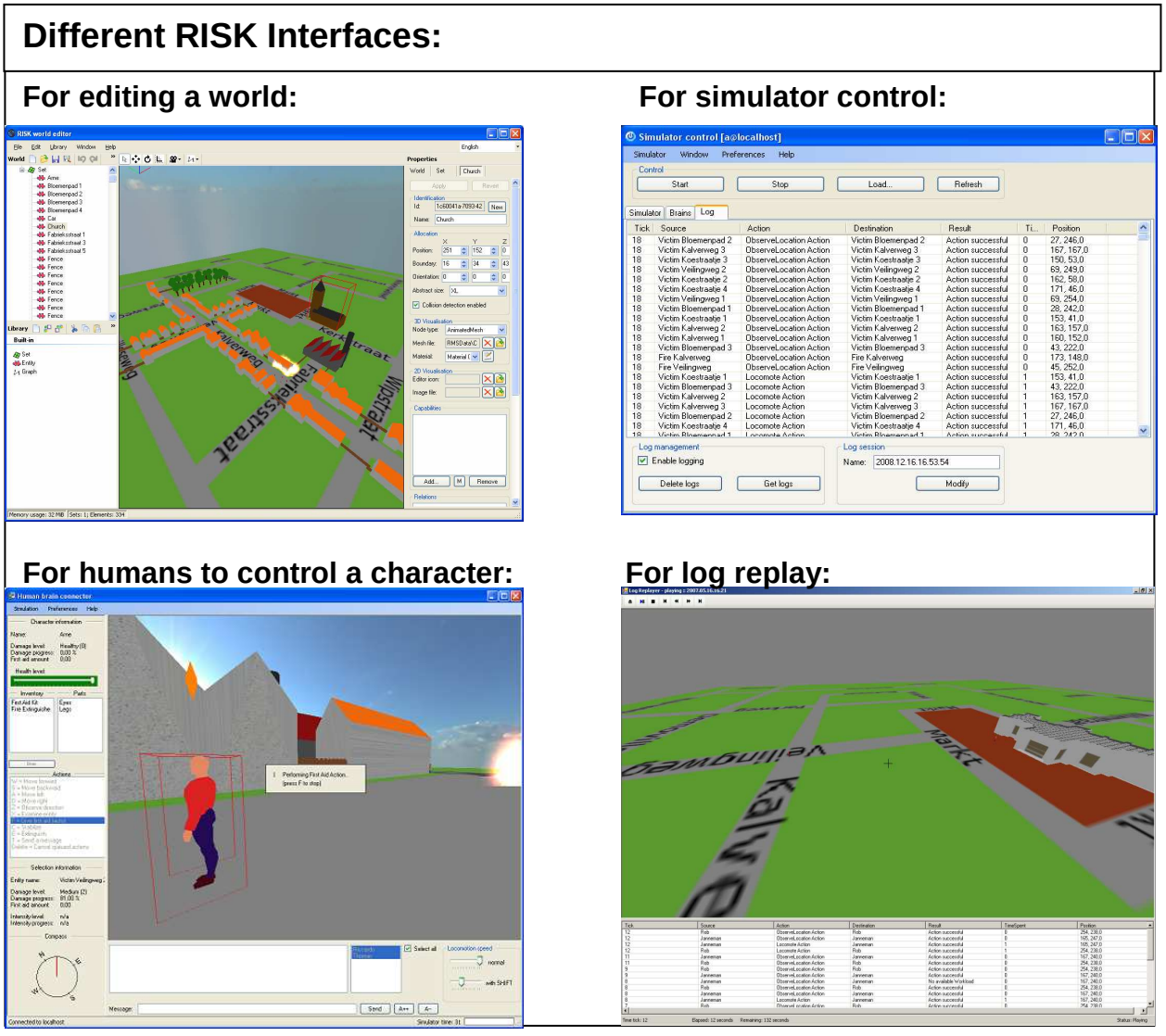


Figure 1: Different Interfaces of RISK

3.2 AAT developments in relation to RISK

To be able to use RISK for Actor-agent teaming experimentation required effort on actor-agent teaming, especially on coordination and communication. For this purpose Peter Oost (student at the Group for Human Media Interaction, Dep. Of Electrical Engineering, Mathematics and Computer Science, University of Twente) combined Machinetta (a framework for team-coordination) with agents in RISK in 2007 (Oost, 2007).

In January 2008 graduate student Roelof Kuijpers (student at the Department of Philosophy Msc. in Cognitive Artificial Intelligence at Universities Utrecht) began modeling agents for a first actor-agent team. He integrated 2APL agents with Machinetta, worked on their interface to the RISK environment, and then realized the text-based actor-agent communication which is essential for actor-agent teamwork (Kuijpers. (2009) Msc. thesis expected).

In May 2008, the graduate student Marten Kampman (student at the Department of Information and Computing Sciences, Msc. Agent Technology at Universiteit Utrecht) joined the SEAT team with a research focus on comparative experimentation with actor-agent teams and actor-only teams, using scenarios in the RISK environment. Kuijpers worked in close cooperation with Kampman, as Kampman addresses the challenge of extending the AAT from Kuijpers to function in specific scenarios for experimentation purposes (Kampman. (2009) Msc. thesis expected). A step towards that direction was the try-out experiment of July 2008, of which the experimental design will be discussed hereafter.

4 Experimental design

To work towards an integration of actors and agents in a team SEAT first aimed to experiment with *actor-only* teams in RISK. The first actor-only team experiment took place on place on July 22 2008. It was realized thanks to the work of students Kuijpers and Kampman as is mentioned in section 3.2. With the first actor-only experiment SEAT aimed to obtain a basic understanding of experimentation with RISK and its related challenges: the participating subjects, scenarios, experimental setup, and measuring the results.

4.1 Scenario's

In this experiment three scenarios are used to study the behavior of a team. With three scenarios it would be possible to experiment with different levels of difficulty (making it more or less difficult to find and respond to victims and fires. Because of time constraints no more than 3 scenarios could be presented to the participants. However, a number of 3 scenarios already could provide 3 different data-sets.

For each of these scenario's Kampman adjusted Demoville (e.g. blocking certain streets with fences, building a recognizable market square, tower, etc.), and adjusted the amount and locations of victims and fires, and their status. It was estimated that the participants in the experiment could extinguish the fires and rescue the victims within approximately 15 minutes in each scenario. During the experiment this appeared to be correct except for the second scenario in which it was too difficult for the participants to find their way. At the end of their 15 minutes time they were acknowledged 5 minutes extra time, but this could not prevent the fact that most victims already died when some participants reached them. The 3 scenarios that were created by Kampman are fully depicted in Appendix B. In short they concern:

- **1. Fire in DemoVille**
  - smoke fumes
  - number of fires and their exact locations unknown
  - make sure all fires are extinguished and all victims are helped

- **2. Toxic Fumes**
  - fire at a factory
  - toxic cloud floated towards town center
  - many people at the Markt are in serious danger
  - extinguish fire at the factory
  - help victims
  - handle other dangerous situations discovered accordingly

- **3. Tornado**
  - tornado has left a path of destruction
  - multiple fires and multiple victims
  - locations are unknown.
  - make sure all fires are extinguished and all victims are helped

4.2 Participants

While preparing the experiment it had to be determined how many people would be invited to participate in the experiment. On the one hand, more participants could yield more data on how people behave in RISK and perform in the scenarios that were created, on the other hand too much data would only complicate our first data-analysis and more participants would complicate the communication too much, it was expected. We therefore decided to invite 6 people to join the experiment within the week previous to the run.

The 6 people that joined the experiment were all more or less familiar to us and each other because they work in the same building: 2 are colleagues, 2 are long-time visiting PhD students (from European countries) and 2 were graduation students; 4 of them are male, 2 female. Although we did not inform about it specifically their ages lie between 20 and 35 years. We did not inquire about computer gaming experiences in this try-out of the experiment.

As a token of gratitude the participants received a box of chocolates the day after the experiment.

4.3 Tutorial

The Tutorial that was used to make the participants familiar with RISK is shown in Appendix C. It consisted of a handed-out paper that contained a screenshot of RISK; explanations of the controls; the interface and the general situation in RISK: controlling characters, extinguishing fires and rescuing victims. Kampman explained the Tutorial to the participants by talking them through the explanations in about 10 min.

4.4 Experiment

The experiment took place in the afternoon of July 22. When the participants were all present Kampman invited them to the introduction to explain the purpose and the schedule of the experiment. The tutorial was discussed with the participants, leaving them time for questions.

Then the participants were divided in 2 groups: 3 participants joined student Kuijpers and one researcher to take place behind three computers on the 1st floor, the 3 other participants joined Kampman and another researcher at three computers downstairs. It was impossible for participants to look on the computers of the others, and they were requested to communicate solemnly through the chat function in the simulator, a request to which they responded well.

When all participants were logged in they were requested to use RISK for 15 min. just to get a feeling of the simulator and its controls. During this training session the students and researchers answered questions from participants. When all participants seemed to have a basic understanding of RISK the first 15 minute scenario was initiated, followed by a 15 minute questionnaire, and a 5-10 minute break. These were followed by a second scenario, questionnaire and break; and then a third scenario, questionnaire, and a small additional questionnaire with some general questions (as is shown in Table 2 below).

Duration	Tentative time schedule Actor-only experiment
10 min.	Introduction
10 min.	Explanation Tutorial
10 min.	Start-up (take place, log on, etc.)
15 min.	15 minute free try-out
15 min.	Session 1 (using scenario 1)
15 min.	Questionnaire 1
05 min.	Small break
20 min.	Session 2 (using scenario 2)
15 min.	Questionnaire 2
10 min.	Small break
15 min.	Session 3 (using scenario 3)
15 min.	Questionnaire 3
10 min.	Additional questionnaire – general questions
15 min.	Short debriefing

Table 2. Tentative time schedule Actor-only experiment

The experiment ended with a short briefing for all who were involved. The entire experiment took about 3 hours. During the experiment the participants were not pressed to hurry, they were given enough time to finish a questionnaire or relax for a moment. Only the sessions were restricted to 15 minutes and stopped by the Simulator Control (Kampman).

5 Measurements and analysis

Most experiments are being conducted to perform measurements of some kind, gain data, analyze those data and then learn something out of this process. Often specific goals / research questions / hypotheses guide the measurements. In our case the approach to our experiment was so explorative that hypotheses were yet impossible to construct beforehand. Thus, as was explained in the introduction the experiment was a try-out, a test to use the RISK tooling and to understand what kind of data can sensibly be obtained from AAT sessions in RISK. To attain this objective two kinds of measurements were conducted:

- automatically all actions of the players were logged on every time-step and saved in RISK log files
- questionnaires captured the participants' opinions on relevant topics (such as: workload, use of RISK, etc.)

These measurements and the related results are described in the following section.

5.1 Data from RISK log files

RISK log files (figure 2 below) contain information on all actions performed by all characters at any time-tick. It also shows where an avatar was when he did something, and whether his action was successful. From the files can also be distracted how long characters performed certain actions and on which characters or objects they performed their actions (e.g. "Thomas performing 'extinguish' on a fire at Kalverweg 23 for 12 time-ticks"). All log files related to this experiment can be obtained from the first author, with an estimated 20 pages in length the files from the 3 scenario's are simply too large to be completely added to this document).

Tick	Source	Action	Destination	Result	TimeSpent	Position
12	Rob	ObserveLocation Action	Rob	Action successful	0	254, 238,0
12	Janneman	ObserveLocation Action	Janneman	Action successful	0	165, 247,0
12	Janneman	Locomote Action	Janneman	Action successful	1	165, 247,0
12	Rob	Locomote Action	Rob	Action successful	1	254, 238,0
11	Janneman	ObserveLocation Action	Janneman	Action successful	0	167, 240,0
11	Rob	ObserveLocation Action	Rob	Action successful	0	254, 238,0
9	Rob	ObserveLocation Action	Rob	Action successful	0	254, 238,0
9	Janneman	ObserveLocation Action	Janneman	Action successful	0	167, 240,0
8	Janneman	ObserveLocation Action	Janneman	No available Workload	0	167, 240,0
8	Rob	ObserveLocation Action	Rob	Action successful	0	254, 238,0
8	Janneman	ObserveLocation Action	Janneman	Action successful	0	167, 240,0
8	Janneman	Locomote Action	Janneman	Action successful	1	167, 240,0
7	Rob	ObserveLocation Action	Rob	Action successful	0	254, 238,0

Time tick: 12 Elapsed: 12 seconds Remaining: 132 seconds Status: Playing

Figure 2: RISK log file

The first experiment using the RISK environment and six human team players has delivered some extensive logs from which the following information has been retrieved using Perl scripts:

1. the paths taken by each individual entity in the environment (visualized by a 'path tracer')
2. the activities/actions performed by each entity
3. the status of each fire and victim over time.

Each of those findings is to be discussed below:

5.1.1 The paths taken by each individual entity in the environment (visualized by a 'path tracer')

Figure 3: Screenshot of RISK Path Tracer

A review option called 'tracer' visualizes the paths taken by each individual entity in the environment. The tracer can show participants' paths at a particular time step in the scenario, or it can 'replay' the entire scenario, showing participants' paths over time. This helps to gain insight in the exploration strategies and areas visited by different participants. Obviously this visualization tool is only informational and doesn't provide any statistical value.

Figure 3 depicts a screenshot of the tracer of the situation at the end of the first scenario (time step 196). It shows that almost the whole area (between the dark-blue roadblocks) is explored and that all victims and fires are found by the team. The names of the characters and their specific colors are shown in the upper left corner.

5.1.2 The activities/actions performed by each entity

The number of performed actions indicates how active each team or individual team member has been in a scenario. It also provides an insight in increase or decrease of activity in comparison with other scenarios. In this case chart 1 shows specific type of actions (e.g. chat, extinguish) that all team members conducted per scenario.

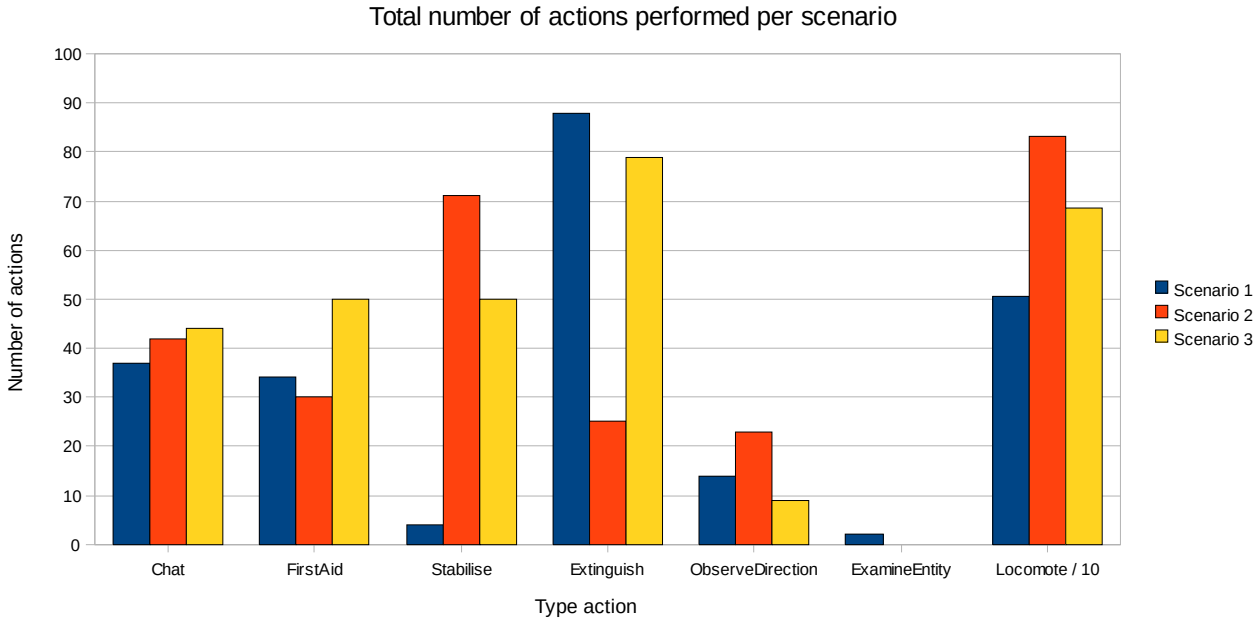


Chart 1: Total number of actions per scenario

The same diagram can also be arranged for individual team members or sub-teams. Only in the 3rd scenario the participants invested time in team formation, an attempt that was frustrated by simulator problems: characters that got stuck in 'each other'. The analysis of sub teams was thus impossible in this experiment.

Chart 1 depicts that chat actions increased with each scenario which indicates that the team communicated more and more. Another observation is that in the first scenario almost no victims are saved as can be seen by the low number of stabilize actions. This is in line with our impressions (also based on remarks of some participants) that the participants initially did not understand the difference between First aid and Stabilize, and concentrated on First aiding victims.

It is obvious that the number of actions performed have no value on themselves. They have to be understood within the context of each particular scenario: extinguish-actions are inextricably bound up with the number and intensity of fires, just as First Aid and Stabilize are bound up with the number and condition of victims.

5.1.3 The status of each fire and victim over time.

In RISK the status of each victim or each fire within a specific scenario can be analyzed. We expect that such an analysis is useful to understand actor-only team or actor-agent team performance. To know the status of a fire or victim by itself is rather irrelevant. The information becomes more important if the event is related to the humans (and future agents) that are to cope with the incident: for example if RISK log files can reveal how one sub-team managed to reach an incident faster than the other sub-team, or extinguish a fire quicker. The RISK path tracer and communication logs can provide additional and useful information to understand team performance and efficiency within the simulated environment.

An analysis of the actions of individuals and sub-teams, the status of fires and victims, and other relevant log data is to be done combination with data from a follow-up actor-agent experiment, and possibly an agent-only experiment. Results from these experiments and analysis are to appear in Kampman's Msc.

thesis on Actor-Agent teamwork and experimentation, (expected Q2 2009), and in a final report on SEAT's AAT experimentation.

5.2 Communication actions

The RISK Log files recorded all communications that were exchanged by the participants (see Appendix D). An understanding of these communications can help to improve the communication between humans and agents that we foresee in AATs. Based only on the unprocessed results the following can be seen:

- In scenario 1 a total of 37 messages were sent, in scenario 2 a total of 42 and in scenario 3 a total of 44;
- The duration of each scenario was about 200 time steps (of 5 seconds);
- Often messages consist of two or more (sub-)sentences;
- 18% of all messages contain one or more spelling or grammar error(s).

Speech acts

An understanding of all communication required a thorough analysis of all individual communicative actions ('speech acts').To make such an analysis possible Kuijpers tagged all communicative actions according to the speech act tags defined by Traum (1994) and Traum *et al.* (2003). This resulted in the following classification (as depicted in table 3):

Speech act labels for RISK communication analysis	
Speech act tag	***Meaning***
Inform	Sender presents receiver with new information in an attempt to add a new mutual belief
Info-req	Sender asks receiver to provide information that sender is missing but suspects that receiver may know; imposes an obligation on receiver to respond
Suggest	Sender proposes a new item as part of a plan
Request	Like a suggest, but also imposes an obligation to respond
Order	Sender, who is superior to receiver in the social structure, orders receiver to do something; imposes an obligation to perform the action that is its content
Accept	Sender agrees to a proposal by receiver
Reject	Sender rejects a proposal by receiver

Table 3. Speech act labels for RISK communication analysis

Depending on the number of sub-sentences messages received, one, two or more tags. For example, the message "I have 2 dead (!) victims, and one 96% victim that I am first aiding now. Can a paramedic heal this guy? (Veilingweg)" received the tags "inform" and "request".

In chart 2 below the percentage of performed speech acts indicates for each speech act how relatively often it has been performed in a scenario, so the total of percentages per scenario is 100%. In this case chart 2 shows specific types of core speech acts (e.g., inform, suggest) that all team members performed per scenario. The same diagram could also be arranged for individual team members or sub-teams.

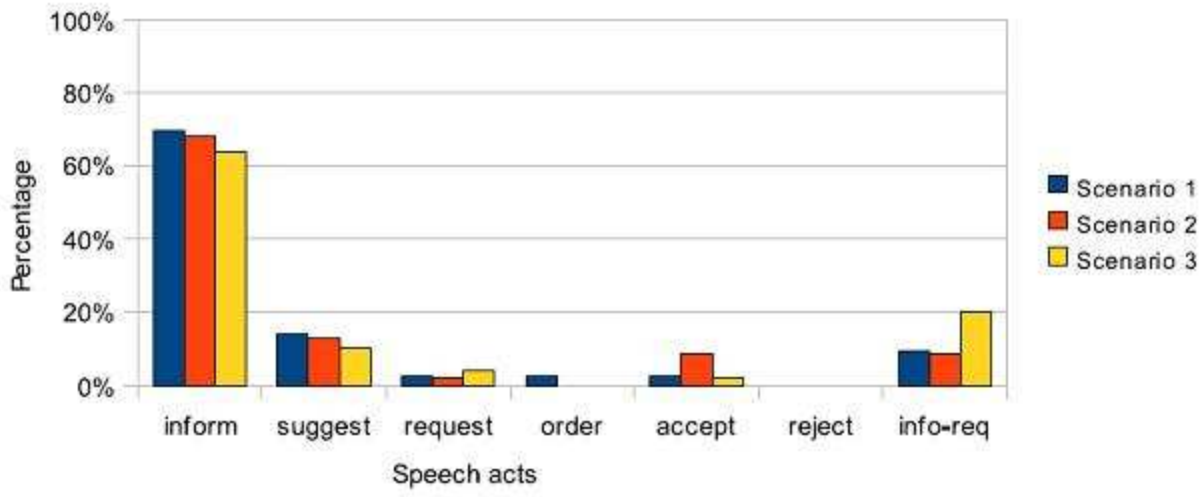


Chart 2: Percentage of speech acts performed

Observations regarding chart 2 are the following:

- Rejection of bids (e.g., task offers) is absent which suggests extreme cooperativeness of the participants in task negotiation.
- A great part (about 2/3) of the messages was sent to exchange information. Information exchanges contribute to the situational awareness, so it is likely that this is somehow reflected in the results. However, regarding their own situational awareness the participants in general indicated in the questionnaire that their situational awareness was neither high nor low (see section 5.3.2).
  Three possible reasons for a moderate situational awareness might be:
  - the information exchanges contained not enough information for a good situational awareness
  - the information that circulated was unclear
  - there was a case of information overload among some or all participants

  Participants' remarks in the questionnaire support the 1st explanation. They mention: 'having not enough communication between the team members' and 'having too little structured communication'.
- Another observation is that suggestions, requests and orders do not always have an explicit accept or reject message as result. It is obvious that this does not necessarily mean that they were neglected. One could accept implicitly by visibly doing the thing being asked and/or informing about doing it. For example, Roelof did ask: "Fire and 2 victims (yellow) on Hoefijzer. Can 1 paramedic join me?" This request was being answered by Thomas (who is a paramedic) with: "I am coming to hoefijzer". The answer, which is an information exchange, contains implicitly an accept message to the request.
- The 3rd scenario contained twice as much information requests compared to the other scenarios. This might be contributed to the fact that characters Thomas and Roelof were stuck for a while, which resulted in information requests like "can you move?" and "I am still stuck. How is everybody else doing?"

Grounding speech acts

Complementary to the analysis of speech acts alone, is the analysis of the acknowledgement of speech acts. This is based on the assumption that speech acts need to be grounded (e.g., acknowledged) first before having their full effect. Grounding acts will often be parts of messages that include speech acts, for example an answer or acceptance will ground the info-request or request that it relates to. Again following Traum's (1994) classification Kuijpers defined the grounding acts as follows (in table 4):

'Grounding act' labels for RISK communication analysis	
Grounding act tag	***Meaning***
Init	An initial message component of a discourse unit
Cont	A continuation of a previous act performed by the same communicator. It is expressed in a separate sentence, but is syntactically and conceptually part of the same act
Ack	An acknowledgement claiming or demonstrating understanding of a previous utterance
Req-ack	Attempt to get the receiver to acknowledge the previous utterance. This invokes an obligation on receiver to respond
Repair	Changes the content of the current discourse unit. This may be either a correction of previously communicated material, or the addition of omitted material which will change the interpretation of the communicator's intention
Cancel	Closes off the current discourse unit as ungrounded

Table 4: Grounding speech acts for RISK communication analysis

In chart 3 (below) the percentage of performed grounding acts indicates for each grounding act how relatively often it has been performed in a scenario (the sum of percentages per scenario is 100%). In this case chart 3 shows specific types of grounding acts (e.g., init, ack) that all team members performed per scenario.

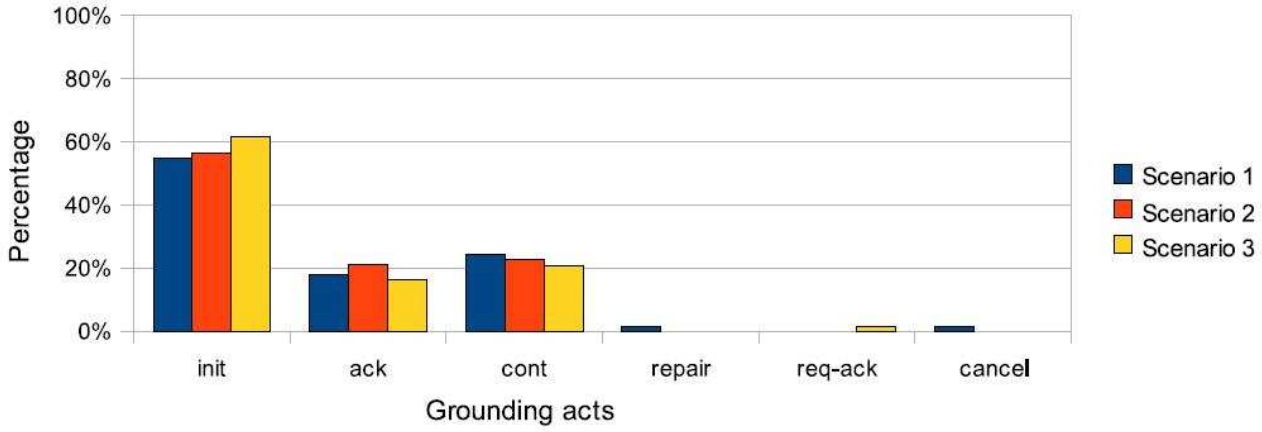


Chart 3: Percentage of grounding acts performed

Chart 3 depicts that:

- less than half of the initiated dialogues (init) have been grounded by an acknowledgement (ack)
- there were hardly any attempts to get the addressee to acknowledge the previous message (req-ack). So, it seems that an initiator often assumes that something is understood, without requiring an acknowledgement. Reasons for this could be that:
  - the visible behaviors of addressees are interpretable as acknowledgements
  - one acknowledgement acknowledges two or more initial grounding acts. For example, in the 3rd session Roelof and Thomas communicated at time-tick 171:

    171 Roelof (All): I am not stuck anymore! Thomas: how ya doin?

    174 Thomas (All): up and running again

Roelof's sentence contains two initiations (one for speech act "inform" and one for speech act "info-req") that need to be grounded. Thomas answered with "up and running again", which is an acknowledgement on the latter part of Roelof's message, but can also implicitly acknowledge the former part.

Understanding grounding/acknowledgement of different types of messages

It is obvious that the percentages of the grounding acts performed have no value on themselves. It would be more interesting to know *what kinds of messages are* acknowledged and what kind of messages are not or less often acknowledged. A classification of all messages based on their content resulted in 6 categories. Chart 4 shows the 6 categories of messages and for each scenario the percentage of messages in a category that has been acknowledged. The most striking issues in chart 4 are:

- Looking in more detail scenario 1 has no action requests (it had one, but the request was cancelled later), so neither request acceptances, and second, in scenario 3 none of the action requests were acknowledged and therefore there are no request acceptances to be acknowledged.
- Requests and messages concerning plan construction have been acknowledged more often than acceptances and information exchanges. To explore this further:
    - The former two happen in dialogues in which the receiver participates actively, the latter in dialogues in which the receiver participates passively. The consequence for agent in the team is that agents do not always know if messages are understood correctly when they are interacting with humans, especially in dialogues in which the human participates passively. Thus, agents would be required to assume that a message is understood correctly, or explicitly ask for an acknowledgement.
    - Probably the acknowledgement of a message also depends on whether dialogue participants observe each other. We have investigated this by defining the 'problem areas' that contain fires and/or victims. When two or more rescuers are in the same problem area at the same time then they see each other. We have analyzed the communication that happened between such 'sub-teams'. Hardly any of the messages in the (few) sub-team communications were acknowledged. This confirms our expectation that when team-members observe each other, little communication is necessary between them.

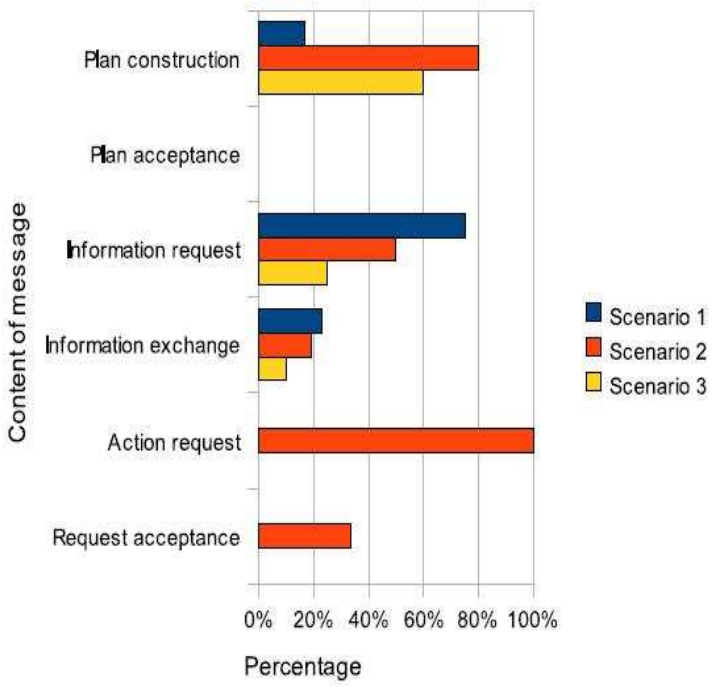


Chart 4: Percentage of messages acknowledged

Towards a Human-Agent Interface
The analysis of human communication in RISK reveals important issues that are to be kept in mind for optimalisation of human-agent communication:

- Since spelling and grammar errors occur often, a human-agent interface must be able to handle these. In the simplest case this can be done by sending a repair request when the human-agent interface encounters a failure during the interpretation.
- The data give examples of sentences that agents should be able to deal with. These can be used as test samples for testing of the human-agent interface.
- To keep the interpretation and translation process by the human-agent interface relatively simple, it is necessary to restrict the messaging to one sentence per message. In the analyzed data this is not the case, which made the (manual) tagging a difficult procedure.
- The data indicate that humans often extract information from visible behavior. Ideally, an agent does same, but it is more pragmatic to avoid implicit answering as much as possible. Solutions to handle this problem could be
    - presenting two answer possibilities ('accept' and 'reject') in the case of a task offer
    - force a participant by the interface to accept or reject an offer first before he/she can continue with other actions

Further work on the human-agent communication is described in Kuijpers' master-thesis (Kuijpers. (2009) Msc. thesis on Actor-Agent team coordination and communication. Utrecht University (expected March 2009)).

5.3 Questionnaire

5.3.1 Focus of the questionnaire

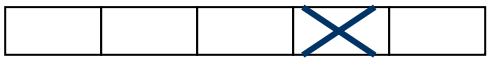
The questionnaire measured 6 themes that are deemed relevant for measurement of the human experiences within experimentation with AAT performance:

- Scenario
    - Was the scenario realistic, difficult, understandable and challenging?
- Situational awareness
    - Does the participant think he/she and the team had a good situational awareness?
- Organizational awareness
    - Does the participant have a good organizational awareness?
- Monitoring
    - Did the participants check on each other's feelings and performance?
- Performance
    - How was performance perceived?
- Workload
    - How was the workload perceived?

After each scenario all participants filled in a questionnaire (thus each participant filled in 3 questionnaires). An additional questionnaire measured participants' opinions on communication and RISK. The complete questionnaires can be found in the appendices E and F. Both questionnaires contained structured and open questions and all results are discussed in the following section. There were 35 questions in sum. The open questions could be answered in a text box, the structured questions could be answered by indicating a choice on a 5-point scale:

1 = not at all good / very little / etc.
2 = a little ….
3 = neutral ….
4 = rather ….
5 = very ….

For example:



(no, not at all) *1* *2* *3* *4* *5* *(yes, very much)*

5.3.2 Results regarding teamwork and performance

The results of this questionnaire are an insight in the subjective experiences of the participants regarding the 6 themes of the questionnaire. Because of the relatively small scale of the experiment the results can only be considered as general tendencies within this small group of participants. The subjective experiences were as follows:

Scenario

The scenarios were not considered as very unrealistic nor very realistic, or difficult. The (2nd 'Toxic Fumes' scenario) scenario was thought to be the most realistic. All three scenarios were perceived as quite understandable and challenging.

Situational awareness

Regarding their own situational awareness the participants in general indicated that their situational awareness was neither high nor low. Although some were not unsatisfied with their situational awareness, several reasons for a poor situational awareness were mentioned: being not able to obtain real-time information, being not used to the user interface, being not yet a good team, having not enough communication between the team members, having too little structured communication, and finding it difficult to locate fires. After playing the second scenario several participants indicated that communication was better, and one player described: "situational awareness grows, but could be better".

Regarding *the team's* situational awareness the participants felt that they were able to share information, and that their communication improved in the second and third scenario, but more information still was needed, for example on road-blocks, and progress of others.

Organizational awareness

Organizational awareness was measured with only two questions: 'whether the roles of other team members were clear', and 'whether people knew whom to ask for information or something to be done'. Both issues were quite positively valued. One participant mentioned that the social structure of the firefighters was not important to him as a paramedic. People appreciated that others told them who they were (firefighter or paramedic), where they were and what they were doing/going to do. The fact that, at the time that the experiment took place in RISK, there was no visual distinction in the user interface between firefighters and paramedics (all were wearing a blue pants and a red shirt), was bothersome for the participants understanding of each other's roles.

In the open question to describe the team's organization participants commented in more detail on their organization. One participant presumably described his view on an ideal organization of a team: "leadership > a team > sub-teams > roles". The other participants indicated for the first scenario that there was an "ad hoc organization" in which they initially would communicate to all in general, and in which everyone used a spread & search tactic for themselves. Also, that there were no real teams created because walking & talking within the simulator required much attention. If there was any organization at all, it was based on proximity, and the natural distinction between firefighters and paramedics, another participant commented.

It was very interesting to see however, that organization did emerge in the second scenario: "right from the start we created sub teams: paramedics went to the market, firemen also split up". This emerging organization was perceived as better than the lack of organization in the first scenario. One participant mentioned that the avatar named Roelof "became (a little bit) team leader".

After the 3rd scenario the participant that controlled the character named Thomas wrote: "I worked out a plan to go in teams of two (a paramedic & firefighter) to three different locations. This plan was followed pretty good, I think". Some of the others indeed mentioned that that there were indeed teams of 1 paramedic & 1 firefighter created, but did not mention explicitly the leadership of Thomas.

At their general remarks two participants made striking remarks on teaming: one found it difficult to decide whether or not to put effort in organizing the team when so little is known about size/amount of

victims and fires. Over-organizing could be a pitfall in his opinion. The other stated his opinion that teams should make action plans before getting involved in critical situations; that they should look for the fire as a team, each in different directions; and that they should agree on clear commands (e.g. about locations: S= south, E = east, etc.)

Monitoring

Regarding the monitoring of other participants' progress: the participants were asked about monitoring feelings, and monitoring performance, both to see whether they checked on the others, and whether they felt that others checked on them. Monitoring of each other's feelings (informing whether others were bored, okay, frustrated, stressed, etc.) was extremely low in this experiment. Monitoring of performance (how people are doing on tasks) was more frequently done, but still not a standard behavior, according to the participants' own perception.

Performance

The perception of performance (accomplishing tasks) showed a positive trend: for both their own performance and the team's performance participants felt that the performance increased with each scenario. At the same time satisfaction with one's own and with team performance increased. Still people found it rather difficult to reach the team's goals, as also became clear from participants' remarks.

The participants were asked to comment on the most difficult and easy tasks for both themselves and the entire team. There were no differences in participants' answers for themselves or for the team. The most difficult tasks were locating the fires and victims. Naturally the participants who had been a firefighter searched for fires and the paramedics searched for victims, although some participants searched for both fires and victims. To know where those victims or fires were, and secondly to reach them was often difficult. The most important reason being a lack of proper information: "no idea where they could be", "no exact location known", "hard to get real-time information", "very little information available". The road-blocks that were part of the second scenario also troubled people. Some participants mentioned their desire to 'perform on time' and to reach an 'optimal performance'. One found "searching and staying together as a team" the hardest mission. The participant with the avatar Thomas that took on some team leadership (see 'organizational awareness') described from then on 'planning and teaming everybody' as his most difficult task because the others had the tendency to go straight to fires and victims.

To extinguish the fires was perceived as very easy, it was just 'pressing a button', the participants met few challenges in this respect (e.g. lack of water/resources). When they were not obstructed by roadblocks, or when they received a clue about where victims/fires could be found, they perceived their tasks as much easier.

Workload

The entire workload of participants was measured by inquiring about feelings of:

- frustration
- time pressure
- 'working hard'
- mental activity

This measure was based on NASA's Task Load Index (Hart & Staveland 1988; NASA Ames Research Center).

As is shown by chart 5 the participants felt that all 4 characteristics of workload decreased almost with every scenario, whereas their satisfaction and the feeling that they accomplished their own goals increased. In addition participants had the feeling that the other participants experienced the same work-load.

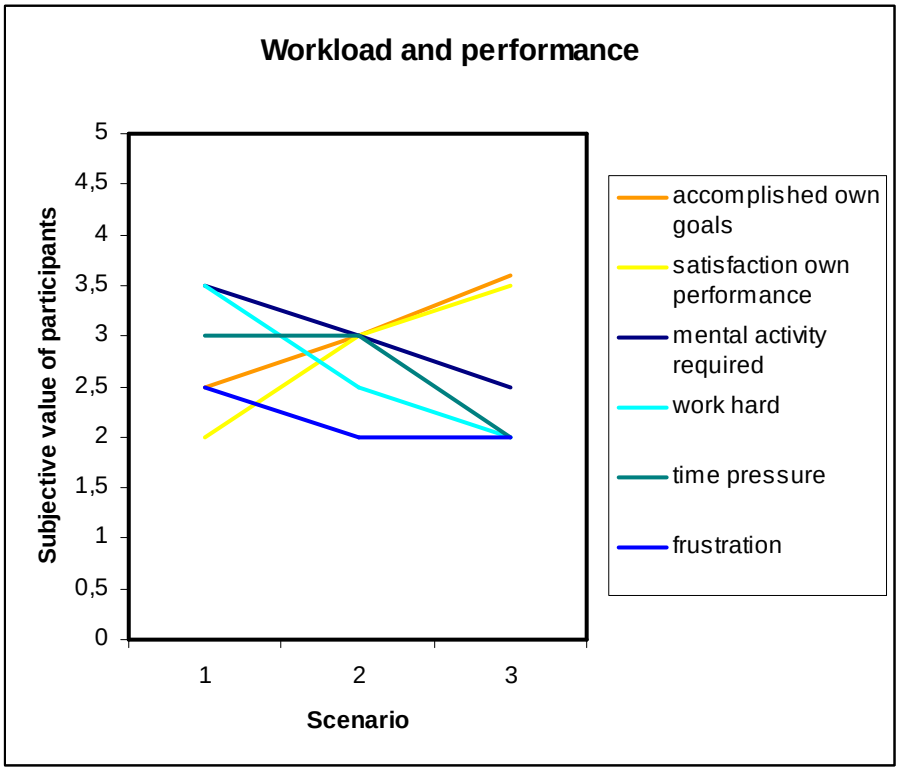


Chart 5: Opinion of participants regarding workload and performance per scenario

5.3.3 Results regarding communication and RISK

Communication

The participants were positive about their communication within the experiment. The feeling that they communicated sufficiently about events, plans and decisions was neutral, but the participants were convinced that their communication improved with each scenario, and even more during the entire experiment. This is line with their feeling that all their skills improved with each scenario.

Participants had few ideas about how to improve communication. They explained that different types of messages were send to one or several team members:

- wondering how one of the others was doing (Marten to Thomas: “will you need any help?”)
- ask someone for help
- ask someone about the location of a fire or victim
- give relevant information to nearby team members (Roelof telling Arne: “Arne, you are now going in the wrong direction, follow Rob!”).

The possibility to send a message to all other participants at the same time was used to inform others about problems with the simulator (“I am still stuck”; “up and running again”) and spreading a general update on the status and location of fires and victims (“I see a fire at the Molenweg 11 !!”; “Paramedic going to Veilingweg”).

RISK

With the use of a large RISK screenshot (see appendix E) that was depicted in their questionnaire the participants were requested to indicate any comments they could think of. This resulted in the following suggestions:

- To permanently show the location and performance of other team members on the RISK user interface
- To have a joystick to control the characters
- To hear a 'beep sound' when a message comes in
- To have the option to select the 'other team' as a destination of communication messages, instead of only one specific character or all at the same time
- To have the cursor immediately in the box for message typing when pressing 'T' in the user interface for sending a message

The most important remarks were:

- different roles of characters (civilian, firefighter, and paramedic) should be made visible by different clothing instead of the red sweaters and blue pants that the avatars now have.
- Secondly to get rid of the time ticks that are used in the simulator (4 of 6 participants remarked this). Participants were not happy with the simulator not running smoothly, and with queued actions, it was found to be not practical and a little 'weird'. One of the participants "got a very detached feeling towards the simulator".

General feedback on the experimentation

Feedback on the experimentation was positive. The RISK tutorial was found not so understandable, but in general it was thought to be sufficient to become acquainted with RISK for the first time and then perform the experiment. Participation in the experiment was considered to be quite nice and interesting, and not so difficult. One participant remarked: "Impressive work is done on the simulator (I know it is difficult)". Another understandingly wrote down "Because it is a large and difficult software package, there are still several bugs, but this will probably be fixed later on I guess…"

6 Discussion

Based on the 'try-out' experiment as it was described in this document, we briefly discuss the main findings that are deemed relevant for AAT experimentation with the RISK simulation. These findings concern the RISK simulator itself, and all major aspects of the experiment and its results. We consider these findings as lessons for following AAT experiments.

RISK

As experimentation tool RISK was unfortunately too unstable to ensure a smooth Actor-only try-out experiment (based on the use of version 3.1). During the training-session and the 3 sessions some participants' characters simple got 'stuck' (on the grass or into another character) as a result of which they could not do anything anymore. For character Roelof, for example, the most difficult task in the 3rd scenario was "getting out of Thomas". Besides characters that got stuck somewhere, participants also suffered 'general RISK jams' and log-on problems (making all others wait or log on again). With only six participants these defects seriously hindered the try-out, to the disappointment researchers, students and participants.

Besides these problems RISK functioned well as a tool and received positive comments. The user interface is relatively easy to use and become acquainted with. The remarks about RISK (see section 5.3.3: Results: RISK) provides some valuable comments about the use of RISK.

Try-out experiment & setup

The try-out experiment clearly was not a real experiment in the pure sense of the word. Instead of having hypotheses, well defined variables and other characteristics of experiments, it was more explorative, testing to see how things worked. Besides our own explorative approach, the RISK simulator was stuck so often during the experiment that a 'real' experiment would have been frustrated too badly to take any results into account.

For this try-out experiment the experimental setup was good. The entire try-out took almost three hours (which is quite long) but the participants did not seem to have been bothered by this time schedule. The three scenarios permitted a comparison between the different stories with specific numbers of fires and victims.

Participants

The participants liked the try-out experiment, up to really enjoying their participation. They cooperated very well with the experiment while playing their roles and answering the questionnaires. None of them complained about the frequent problems that occurred with the RISK simulator, although the experiment and their performance were sometimes hampered by those disturbances.

Because they were all familiar to us (but not with RISK or AAT research) we could easily invite them for participation, and set a date and time. The participation of 'new' participants might be more difficult if our AAT experiments are to run more frequently in the near future. Students from the Technical University of Delft might be future participants: this university is very close-by and Thales NL has many connections with TU Delft researchers. We have to find out in the future whether this would be an option to involve participants.

Tutorial

Regarding the tutorial and explanation of RISK to the participants it seems that using RISK is relatively easy to learn (as we once experienced ourselves). However: the tutorial was first discussed in general and then the participants had the time to practice in RISK for 15 min. It might have been more beneficial for the participants' understanding of the RISK user-interface and controls if they were guided through a tutorial behind their computers. We might change this in future experiments to see if such a hands-on tutorial provides participants with a better explanation of RISK. Such a hands-on tutorial should be supported by

close assistance of students and researchers who are to make sure that every participant fully understands the RISK user-interface.

Scenario

The scenarios as they were used for this try-out experiment were able to keep the players busy for the 15 minutes duration of the sessions. In one session the participants received 5 minutes longer to fulfill their tasks, in another scenario most victims at the market had passed away too early/were out of health, much to the disappointment of the paramedics who could do nothing more to save them.

Obviously there is a precarious balance between a scenarios' design and the possibilities for players to reach their goals. At this moment the following factors play a role within the AAT scenarios:

- The duration of a scenario
- Information that is given to participants beforehand
- The number of victims; their 'health status' (slightly wounded, badly hurt, etc.); and degradation rate
- The number of fires; their 'status' (smoldering fire, large fire, etc.); and degradation rate
- The number of players
- The players' roles (firefighter, paramedic)

The resources that players have their disposal (water for firefighters and plasters, bandages, medicines for paramedics) were not yet taken into account for experimentation, but can be a valuable addition to enhance difficulty for the participants, maybe induce some feelings of 'stress', and the issues of team member cooperation, coordination or even competition.

Measurement

The measurement of 'performance' and 'effectiveness' of an AAT is a challenging research quest. We aim to learn –through explorative experiments- how research questions, hypothesis and measurements can be used to gain further AAT understanding.

This experiment has yielded many data which can be used in different ways. At this moment it is possible to determine the location and activity of every participant on every time-step, we can look at the status of fires and victims and the communication between participants. How all these data can be combined to derive conclusions about performance is still difficult and can be further explored in the follow-up experiments with actors and agents in RISK

Communication

The analysis of the actor-only communication during this experiment reveals important issues that are to be kept in mind for optimalisation of human-agent communication: handling spelling and grammar errors; the usability of real communication data for testing in a later phase; the desired simplicity of the interpretation and translation process by the human-agent interface relatively simple; and the difference between actors and agent in handling information visually. From the analysis of the communication logs implications were drawn that are to improve the human-agent interface, and communication.

Questionnaire

The questionnaire was a success. The participants patiently took the time to fill in questions after each session although the 2nd and 3rd questionnaire were very for a large part similar to the first one. Their response gave an insight in the subjective opinions about the experiment, and rendered feedback on RISK's user-interface, RISK communication options and the try-out experiment in general.

The questionnaire is an easy tool to gather qualitative, subjective data and can easily be used and adapted for re-use. If more work is done on how to measure teamwork dimensions such as organizational awareness, monitoring, workload, and on performance and situational awareness the questionnaire could be a more complete measuring tool for these concepts. This does require more extensive research into measures regarding those concepts. Secondly the analysis of the response could be less time-consuming if the questionnaire would be administered by computer instead of on paper.

Teaming

Due to SEAT's research focus teaming is an extremely important research theme. In this try-out experiment teaming played a relatively small role, because of the explorative character of the try-out. The participants were not natural teams 'by themselves', nor were there artificially teams 'created' by the researchers.
Too many other factors in this try-out experiment (such as use of RISK, scenario's, communication etc.) were unclear to invest a great effort in teaming participants.

However, because teaming of participants is an important issue reflection on teaming is essential.
A team can be considered as a "distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal/objective/mission, who have each been assigned specific roles or functions to perform, and who have a limited life-span membership" (Salas, 1992, p. 4) (see for a discussion also Gouman *et al*. 2007). Relevant questions are: how to create a 'team awareness' among the participants? In other words: how to give them the feeling that they belong to the same team that shares the same goal? How to make them committed to that shared goal? How to measure team awareness within our simulated AAT? How to relate agents to human team awareness? These are challenges for future experiments.

7 Outlook on further experimentation

The results from SEAT's first actor-only pilot experiment as described in this document bring the authors a step closer to full-grown AAT experiments. Further AAT experimentation is a specific SEAT research objective. To reach this objective more development is required on software and experimental methodology.

Regarding software we envision improvements on:

- RISK as a simulation tool (e.g. RISK's stability, debugging, capacity to handle more participants and agents)
- Improving the actor-agent interface and actor-agent communication
- Integration of agents into an AAT
- Improving and stabilizing the connection between Machinetta, RISK and actor-agent interfaces

Regarding experimentation we aim to:

- Conduct more explorative actor-only, actor-agent and agent-only experiments and learn 'from the data'
- Reports of actor-agent-team and actor-only-team try-out experiments are expected in Q1 of 2009
- Improve AAT effectiveness and performance measurement-techniques
- Improve data-analysis
- Compare the results of all series of try-out experiment to adjust experimental setup and measurement techniques for next series of experiments in Q2 of 2009.

Alea iacta est!

8 Acknowledgements

The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project at the D-CIS lab, supported by the Dutch Ministry of Economic Affairs, grant no: BSIK03024.

For his contribution to the Machinetta agent teaming framework the authors thank Peter Oost (University of Twente). For his work on agent modelling, and the integration of Machinetta and 2APL with RISK the authors thank Roelof Kuijpers. For his contributions to the integration of Machinetta and 2APL with RISK, the design of scenarios for initial Actor-Agent-Team experimentation; the setup of SEAT's first Actor-only experiment; and the analysis of its results the authors thank Marten Kampman. They also express their gratitude to Frank Dignum (Utrecht University) for his advice and guidance to Roelof Kuijpers and Marten Kampman.

The authors' gratitude also goes out to Thomas Hood (TRT-NL), Hugo Burm (Y'All) and Elmer van Chastelet for their work on RISK.

Finally we are indebted to the 6 participants who invested their time willingly into our experiment, thereby helping us in the process to understand more about actor-agent-teaming.

9 Bibliography

Gouman, R., Wijngaards, N., Kempen, Masja. (2007) Basic Human Group Characteristics: exploring implications for actor-agent teams. Internal document within the ICIS Project.

Hart, S. G., & Staveland., L. E. (1988) Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload.* Amsterdam. The Netherlands: Elsevier.

Kampman. (2009) Msc, thesis on Actor-Agent teamwork and experimentation. Utrecht University (expected Q2 2009)

Kempen, M., Wijngaards, N., Gouman, R., (2007) Sustained Effectiveness of Actor-Agent Teams: Explorations and Demarcations. Internal document within the ICIS Project.

Kuijpers. (2009) Msc, thesis on Actor-Agent team coordination and communication. Utrecht University (expected March 2009)

NASA Ames Research Center: Human Performance Research Group. NASA TASK LOAD INDEX (TLX). v. 1.0. Computerized Version. Instruction manual. Moffett Field, California

Oost, H.B. (2007) Agent team coordination with Machinetta. (Report on an internship) University of Twente.

Salas, E., T.L. Dickinson, S. Converse & S.I. Tannenbaum. (1992) Toward an understanding of team performance and training. In R.W. Swezey & E. Salas [Eds.] Teams: Their Training and Performance. Norwood (NJ): Ablex, p. 3–29.

Traum, D. R. (1994) A computational theory of grounding in natural language conversation. Technical report. Rochester, NY, USA.

Traum, D. R., Rickel, J., Gratch, J., and Marsella, S. (2003) Negotiation over tasks in hybrid human-agent teams for simulation-based training. In: Proceedings of International Conference on Autonomous Agents and Multiagent Systems (AAMAS), pp. 441- 448. ACM Press

Appendices

Appendix A: Explanation of RISK Human Brain Connector

1. Here you can see the 3D view of the simulation from the perspective of the controlled brain. You can activate the view by clicking on it and leave it by pressing the Escape key. You can look around by moving the mouse.
2. This is the character information panel. Character properties of the controlled brain, such as damage level, are displayed here.
3. The inventory panel shows the tools that the controlled character carries.
4. Shows a list of actions that can be performed by the character. Actions are performed by pressing keys on the keyboard. (No need to press Enter afterwards!) The configured key for an action is shown in this list in front of the action and can be changed from the menu: Options|Key configuration. If a key for an action is pressed the action will be submitted to the simulator and highlighted in the list. When the simulator has actually performed the action, the menu item will be de-highlighted.
5. The selection information panel displays properties of the selected entity such as damage level for human characters and intensity for fires.
6. The compass shows the direction the character is facing.
7. Command line to chat with others users in the simulation.
8. Lets you adjust the velocity and shift of the character being controlled.

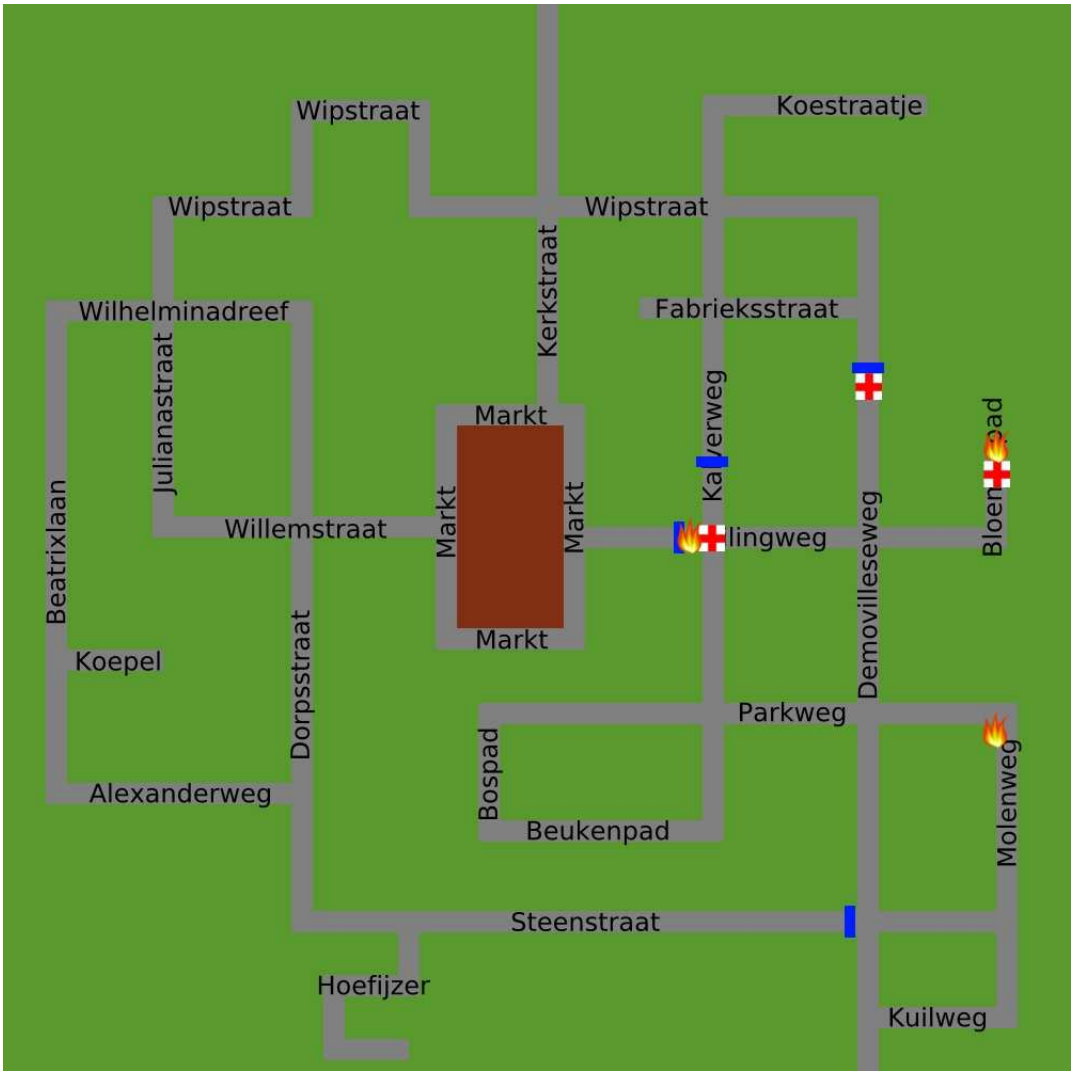
Appendix B: Actor-only experiment scenario's

1. Fire in DemoVille

"Already from a distance of 5 km are smoke fumes visible emerging from the small town of DemoVille. The emergency room has been alerted and has sent 3 firemen and 3 paramedics to investigate the situation. How many fires and their exact locations isn't known. However the location of the fires and victims is situated in the southeast of DemoVille.

Starting position: intersection of the Kuilweg and Demovilleseweg."

N.B.: This scenario contained 3 fires and 10 victims



Map of Demoville scenario 1

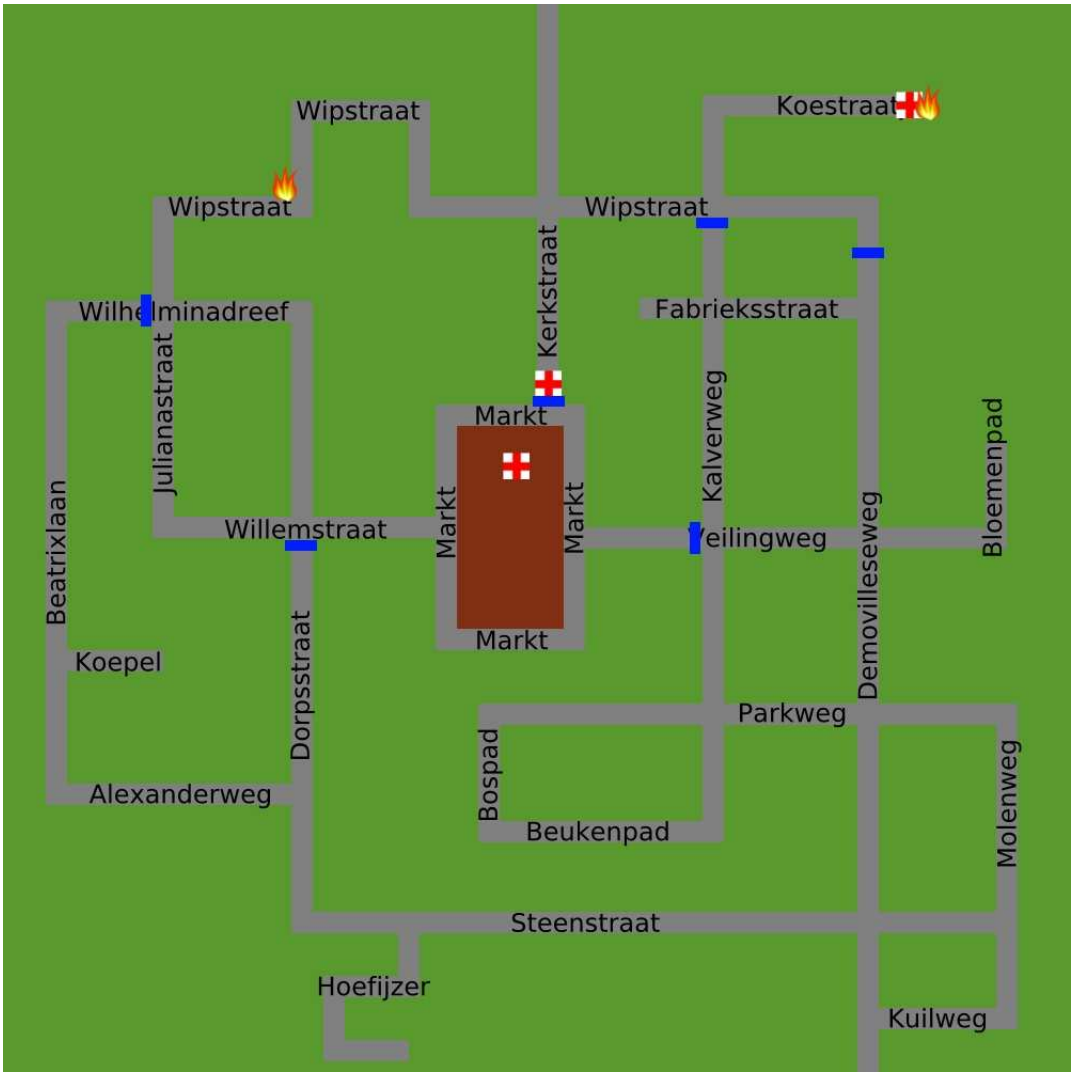
2. Toxic Fumes

Due to a serious fire at a factory in the North of DemoVille, a toxic cloud has been formed and has floated towards the town center. It is known that there are many people located at the Markt who are in serious danger. Make sure that the fire at the factory is extinguished and the victims are helped.

May there be any other dangerous situations discovered, handle them accordingly.

Starting position: end of the Kerkstraat.

N.B.: This scenario contained 2 fires and 13 victims



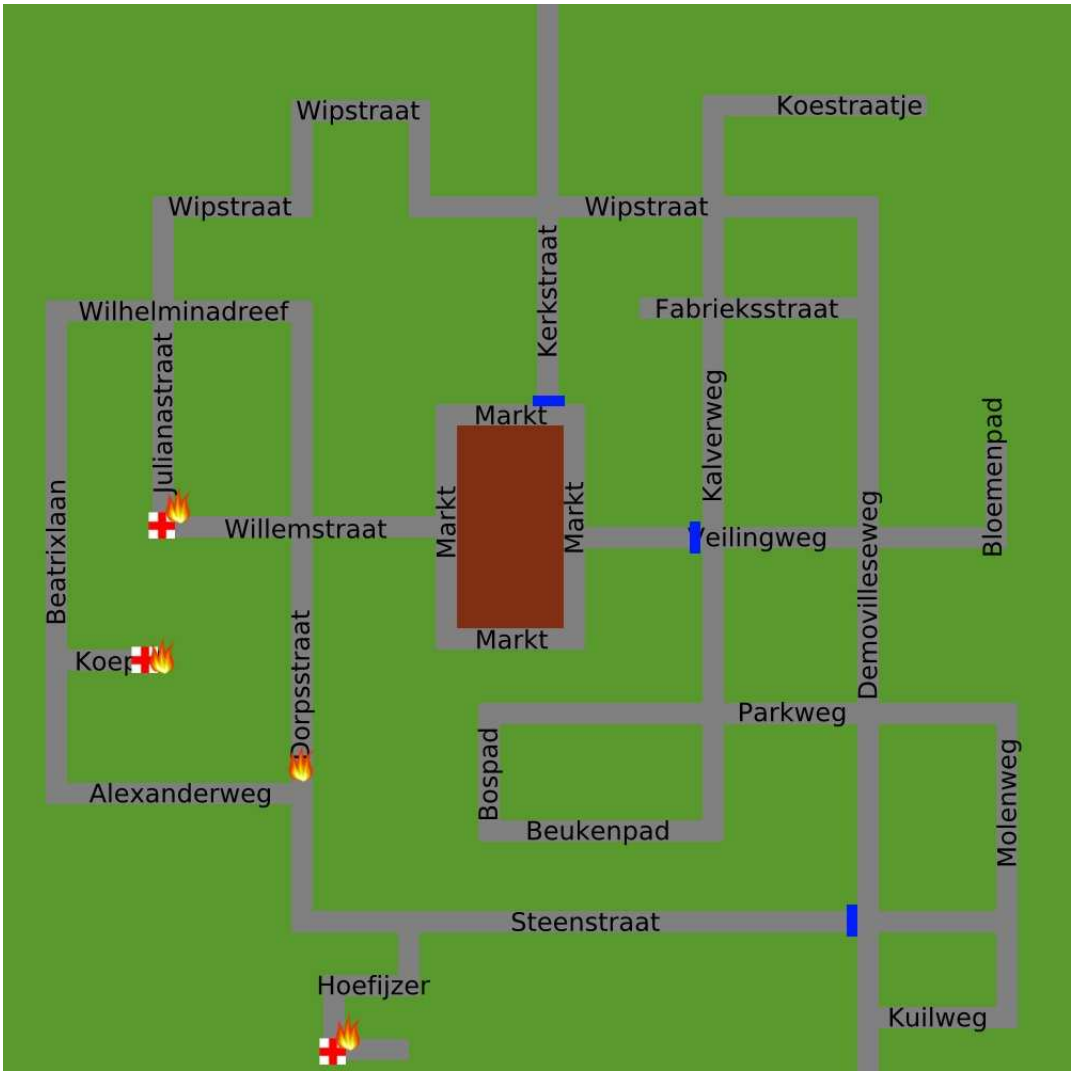
Map of Demoville scenario 2

3. Tornado

Because of the effects of global warming a tornado of the 5th category has struck DemoVille. This tornado has left a path of destruction at the west part of town. There are multiple fires and multiple victims, the total amount and their locations are unknown. Make sure all fires are extinguished and all victims are helped.

Starting position: intersection of the Steenstraat and Demovilleseweg.

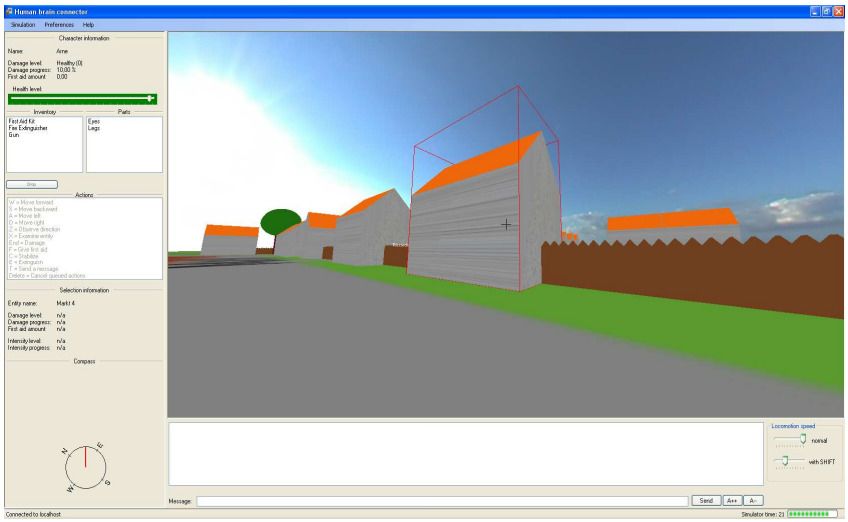
N.B.: This scenario contained 4 fires and 10 victims



Map of Demoville scenario 3

Appendix C: RISK Tutorial

Tutorial



Controls:
W – Move forward
A – Move backward
S – Move left
D – Move right
Z – Observe direction
F – Give first aid
C- Stabilize
E – Extinguish

Multiple actions are possible at the same time step for example:

Walk & Observe, Extinguish & Chat, etc.

Interface:
Character information – Displays name and health status
Inventory – Objects in possession
Actions – Available commands, to chat press esc and type in the message bar.
Selection information – Displays name and damage status the selection
Compass – Gives the current orientation
Messages – Displays all messages sent and allows sending messages to certain receivers
Simulator time – Displays the time ticks

Situation:

3 firemen:	Thomas, Riccardo and Roelof
3 paramedics:	Arne, Rob and Marten
1 fire	located at Koestraatje 18
4 victims	located somewhere at the Koestraatje

The health of the victims involved decreases and the intensity of the fire will increase over time.

First aid: applying first aid to a victim will decrease or stop the amount of damage taken. Note: this will not increase the health. This can be done by both a fireman and paramedic and should be applied in close range of the victim.

Stabilize: the stabilizing action will increase the health of the victim to a higher level. This action can only be performed by a paramedic. When a victim is 'healthy' stabilizing isn't necessary anymore.

Extinguish: extinguishing a fire can be done only by a fireman in close range of the perceived fire.

Objectives:
1. Walk around and use the observe direction action.
2. Communicate with teammates, directed (to a subgroup) and to all
3. Apply first aid to a victim
4. Fireman: move towards the fire and start extinguishing
5. Paramedic: stabilize the four victims involved

Appendix D: Logs of communications

This appendix shows the log files of all communications between the six participants during the three sessions. The left column ("Time") shows at what timetick the chat action was send, the second column ("Source") shows which players send the chat action, and the third column ("message") shows between brackets the destination for the chat, plus the actual message.

While communicating participants could send a message to one specific participant or to all other participants, by selecting with their mouse one name from the list, or clicking on the "all" button.
The log files show that most participants used to communicate to all other players at once, instead of selecting players for specific communication. We presume, as some participants also remarked that this feature of the chat function was not obvious to them when they participated in this try-out experiment.

1st session: communication about fires in Demoville

Time	Source	Message
14	Roelof	(All): Hi All, as it is in the southeast (we are now southwest) we should run down Demovilleseweg?
18	Riccardo	(All): We should go to the west, that's where the tornado was
19	Rob	(All): I think so
23	Arne	(All): I think we should split in teams !
24	Riccardo	(All): Oh wait, I read the wrong scenario...
32	Riccardo	(All): Southeast it is, sorry
34	Thomas	(All): I also read the wrong scenario
38	Riccardo	(All): But I have the correct one now...
47	Marten	(All): I am near by the Riccardo, but have not see vicitim
48	Riccardo	(All): Anyone who sees fires or victims, please report to all
50	Rob	(All): Does anybody see the fire?
59	Thomas	(All): I don't see any fire im at the crossing of demovilleseweg and kuilweg
59	Roelof	(Arne): Arne, you are now going in the wrong direction, follow Rob!
66	Arne	(All): I dont see anything eighter
70	Roelof	(All): Thomas: it should be on the southEAST, we started on the southWEST - so we have to go through the demovilleseweg first!
70	Riccardo	(All): No fires at the crossing Kuilweg Molenweg
77	Rob	(Roelof): roelof, did you see any fire there?
80	Roelof	(All): Victims! End of the Demosevilleseweg!!!
82	Roelof	(All): (they are standing next to cars...)
91	Roelof	(All): but the roads end with the cars, there is a fence. I will first go look for the fire.
92	Arne	(All): I see a fire at Molenweg 11 !!

96	Marten	(All): I also see the fire
104	Thomas	(All): I am on my way Roelof
118	Roelof	(All): The Demovilleseweg is closed between junctions Veilingweg and Fabrieksstraat
120	Roelof	(All): There is a fire and victims on the Veilingweg!
131	Riccardo	(All): Paramedic going to velingweg
134	Arne	(All): ok roeloff, i'm on my way. I extinguished the fire at Molenweg
137	Roelof	(All): I have 2 dead (!) victims, and one 96% victim that I am firstaiding now. Can a paramedic heal this guy? (Veilingweg)
139	Marten	(All): I see!
146	Roelof	(All): No paramedics needed anymore at the veilingweg, they're all dead...
162	Roelof	(All): Are the fires at the molenweg extuinguished?
166	Arne	(All): yes molenweg is extinguished
167	Marten	(All): Riccardo, I am flowing you. Have you saw any viciticm?
172	Arne	(All): Fire and Victems at Bloempad 3 - at the Factory
174	Rob	(Roelof, Thomas): Thomas, your patient is dead.
178	Riccardo	(All): Paramedic going to Bloempad
181	Arne	(All): I am extinguishing the fire right now

2nd session: communication about fire at a factory and victims on market square

Time	Source	Message
10	Riccardo	(All): Okee, alle paramedics als subteam naar de markt denk ik?
15	Roelof	(All): (english!) yes, all paramedics to the markt, all firemen to the fire (where is it?)
24	Thomas	(All): Keep in mind that a paramedic could be needed by the fire
31	Marten	(All): I am following you!
32	Riccardo	(All): The firemen can report if any victims are found there
44	Thomas	(All): That is right
46	Rob	(All): i am a fireman and I go on Wipstraat West, please some other fireman go look for the fire in Wipsdtraat east
50	Arne	(Rob): I'm will do that
58	Roelof	(All): Rob, I am already on Wipstraat west, could you do the other? (I am already past the first corner)
60	Thomas	(Riccardo, Arne): i found 1 victim straight ahead of the start
65	Riccardo	(Thomas, Marten): performing first aid with thomas

75	Thomas	(Riccardo, Arne): 3 other victims are a bit further but you will have to go around some blocking cars
82	Marten	(All): (Thomas) Where are you,Thomas?
82	Rob	(All): Ok I will go in fabriekstraat
90	Riccardo	(Thomas, Marten): Kerkstraat blocked, going round
93	Roelof	(All): Found fire!!! Along Wipstraat-West.
98	Thomas	(All): I am on markt street, straight ahead of the starting point
108	Thomas	(All): Im at Markt Kerk
109	Roelof	(All): I am extuinguishing a big fire on the Wipstraat, it is the second corner, near the 't' in the first Wipstraat word.'
120	Riccardo	(Thomas, Marten): I can't find my way around, streets are blocked with fences (klaverweg, demovilleseweg
120	Arne	(Roelof): I'm on my way, almost there
121	Marten	(All): I think we are blocked again, Riccardo
129	Riccardo	(Marten): Lets try to find another way,
135	Marten	(Riccardo): OK
144	Roelof	(All): The fire on the Wipstraat that I had found is now extinguished. I am now continuing along the wipstraat.
156	Thomas	(Riccardo, Marten): the 5 victims who are on the market that i see are all dead
157	Roelof	(Arne): Arne, you doing the Wilhelminadreef?
161	Arne	(Roelof): Yes, i will go down this street
164	Riccardo	(Thomas, Marten): Ok, several roads were blocked, so we couldn't find our way to the other side of the market in time
167	Roelof	(Arne): OK, I'll do the Julianastraat.
176	Rob	(All): How is the sirtuationon the Markt?\
181	Roelof	(Arne): Hi Arne, nothing found in my corner.
181	Riccardo	(All): A few victims saved, but others were unreachable in time
183	Arne	(Roelof): Nothing here also
199	Riccardo	(Thomas, Marten): Well I've found a way round, but too late...
206	Thomas	(Riccardo, Marten): Too bad
209	Marten	(Riccardo): I still have not reach the Markt
212	Roelof	(All): Firemen, I will check the Veilingweg.
213	Riccardo	(Thomas, Marten): I see some firemen walking around there, so they have been able to get there
219	Thomas	(All): I can't perform first aid and stabilisation anymore
228	Roelof	(All): Veilingweg is obstructed halfway, can't go further. Going back to the markt.

232	Marten	(Thomas): I am neary by you, but can not see viciticms

3rd session: communication about fires and victims in western Demoville

Time	Source	Message
18	Riccardo	(All): How do we team up?
25	Rob	(All): I go west to dorpstraat
25	Arne	(All): Ok, so we should probably all go down on steenstraat, then split up in a south and north team ?
26	Roelof	(All): I think, walk along the Steenstraat, and then split from there? There is no specific fire/sick people place.
27	Marten	(All): I agee
37	Roelof	(All): I will check out the Hoefijzer on the left.
47	Riccardo	(All): Who will be the north team?
53	Roelof	(All): Fire and 2 victims (yellow) on Hoefijzer. Can 1 paramedic join me?
54	Thomas	(All): teams of 2 paramedic/fireman. team 1 goes to beatrixlaan, team 2 willemstraat and julianstraat, team 3 dorpstraat. that's what i think is best
57	Thomas	(All): i am coming to hoefijzer
59	Arne	(All): i will go to the Beatrixlaan
63	Riccardo	(All): I will follow Arno
64	Roelof	(All): 3 victims here, performing first aid on 1. Others are orange and yellow.
65	Marten	(All): I will in team 2
66	Arne	(All): I already see a fire on junction alexanderweg - dorpstraat
78	Riccardo	(Arne): I will go ahead to beatrixlaan while you put out the fire
84	Rob	(All): I anm estinguishing the fire in dorpstraat
96	Roelof	(Thomas): Thomas, I am standing ' in ' you, can you move?
105	Riccardo	(All): Victims and fire in Koepel, 4 victims
107	Thomas	(All): i can't move either
108	Marten	(Thomas): will you need any help?
110	Arne	(Riccardo): i'm on my way
125	Arne	(Riccardo): I'm extinguishing the fire
133	Marten	(None): I have give first aid to victim 2 and 3
138	Thomas	(All): Hoefijzer 3 is still on fire
140	Roelof	(Thomas): thomas, can you do the red patient? I am extinguishing the fire.
143	Roelof	(Thomas): F
144	Rob	(All): Thomas, are you asking for help?

145	Roelof	(All): Hoefijzer fire is out!
150	Arne	(All): Koepel fire is out !
151	Thomas	(Roelof): Victims are already dead
152	Roelof	(All): all the hoefijzer victims are either dead or already first-helped.
155	Riccardo	(All): 1 victim stabilized at koepel, 3 dead
156	Thomas	(All): I am still stuck. how is everybody else doing?
156	Roelof	(Marten): Marten, can you heal my patient who has already been first aided?
162	Marten	(All): I have give first aid to vicitim 2 and 3
164	Arne	(All): i extinguished a fire at koepel, i'm now going up on Beatrixlaan
164	Rob	(All): I go and investigate the east end of the town, is anybody there?
171	Roelof	(All): I am not stuck anymore! Thomas: how ya doin?
174	Thomas	(All): up and running again
179	Riccardo	(Arne): Still following you!
193	Riccardo	(None): left or right?
198	Riccardo	(Arne): victims on the right and fire
199	Marten	(All): I heal the vicitm 2 and 3, are they you have mentioned?

Appendix E: Questionnaire on performance

QUESTIONNAIRE AAT EXPERIMENT - Session 1

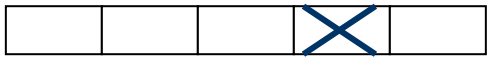
Introduction

- You have just completed a session in the AAT experiment. In addition to the data that we hope to obtain from RISK we are very interested in your opinion on some topics. This questionnaire therefore addresses several issues that are considered of importance to team performance.

- For reasons of comparison between scenario's, performance and teaming we will handout this questionnaire after each scenario has taken place (3 times in total).

- We expect that completing the questionnaire will take about 7 - 9 minutes.

- Most questions can quickly be answered by indicating your choice on a 5-point scale:
  1 = not at all good / very little / etc.
  2 = a little ….
  3 = neutral ….
  4 = rather ….
  5 = very ….

  For example: 

  (no, not at all) *1* *2* *3* *4* *5* *(yes, very much)*

- Your ratings and remarks will be dealt with anonymously, please feel free to add any kind of comments and help us improve our experiments.

Thank you in advance!

1. Scenario

To what degree do you find the scenario (not the simulator's user interface, but the story and events)…

1.1 … Realistic: ☐ ☐ ☐ ☐ ☐

(not at all) 1 2 3 4 5 (very much)

1.2 … Difficult: ☐ ☐ ☐ ☐ ☐

(not at all) 1 2 3 4 5 (very much)

1.3 … Understandable: ☐ ☐ ☐ ☐ ☐

(not at all) 1 2 3 4 5 (very much)

1.4 … Challenging: ☐ ☐ ☐ ☐ ☐

(not at all) 1 2 3 4 5 (very much)

2. Situational awareness

Situational awareness can be considered as an awareness of the current state of the environment & ongoing events, based on previous and real-time information.

2.1 Do you have the feeling that you continuously had a good situational awareness?

☐ ☐ ☐ ☐ ☐

(no, absolutely not) 1 2 3 4 5 (yes, all the time)

2.2 Why, or why not?

2.3 Do you think the other team members had a good situational awareness?

(no, absolutely not)	1	2	3	4	5	(yes, all the time)

2.4 Why, or why not?

3. Organizational awareness & leadership

3.1 According to your perception, how did the team organize itself (e.g. sub teams, roles, leadership, …) ? Please draw or explain this:

3.2 Were the roles (e.g. firefighter, team leader, paramedic, …) that *all* other team members performed during this mission *completely* clear to you?

(no, absolutely not)	1	2	3	4	5	(yes, completely)

3.3 Why, or why not?

3.4 When you needed information or something to be done, did you know whom to ask?

(no, not at all)	1	2	3	4	5	(yes, always)

4. Monitoring

4.1 Did other team members ever ask you how you felt (whether you were okay / bored / stressed / frustrated / …)?

(no, not at all)	1	2	3	4	5	(yes, very often)

4.2 Did you ask other team members about their feelings (whether they were okay / bored / stressed / frustrated / …)?

(no, not at all)	1	2	3	4	5	(yes, very often)

4.3 Did the other team members ask you how you were doing on your task(s)?

(no, not at all)	1	2	3	4	5	(yes, very often)

4.4 Did you check on the other team members' performance?

(no, not at all)	1	2	3	4	5	(yes, very often)

5. Performance

5.1 Do you think that the team reached its goals?

(no, none)	1	2	3	4	5	(yes, *all* goals)

5.2 Are you satisfied with the team's performance?

(no, not at all)	1	2	3	4	5	(yes, very much)

5.3 How difficult was it for the team to reach its goals?

(really easy)	1	2	3	4	5	(very difficult)

5.4 What was the most difficult task to accomplish?

……

5.5 Why?

……

5.6 And what was the easiest?

……

5.7 Why?

……..

5.8 Do you feel that you could accomplish your own goal(s)?

(no, not at all)	1	2	3	4	5	(yes, all goals)

5.9 How satisfied are you with your own performance?

(not satisfied at all)	1	2	3	4	5	(very satisfied)

5.10 What was your most difficult task to accomplish?

……

5.11 Why?

……

5.12 And what was the easiest?

……

5.13 Why?

……

6. Workload

6.1 How much mental activity was required (e.g. thinking, looking, deciding, searching, remembering, …)

(very little)	1	2	3	4	5	(very much)

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

1	2	3	4	5	

6.3 How much time pressure did you feel during your performance?

(very little)	1	2	3	4	5	(very much)

6.4 How much frustration did you experience? (e.g. irritation, stress, discouragement, etc.)

(very little)	1	2	3	4	5	(very much)

6.5 To what degree do you think all team members had the same workload?

(variable workloads)	1	2	3	4	5	(same workload)

7.1 Any additional remarks concerning this scenario and performance…:

Appendix F: Questionnaire on RISK & Communication

GENERAL QUESTIONS CONCERNING COMMUNICATION & RISK

Communication

C.1 In general, what was your motivation to send a communication to *one / several* team member(s) or to *all* team members?

***one / several* team member(s):**
***all* team members:**

C.2 Do you think that the team communicated sufficiently about all events that were going on during the scenarios?

(no, not at all)	1	2	3	4	5	(yes, very much)

C.3 Do you think that the team communicated sufficiently about plans and decisions?

(no, not at all)	1	2	3	4	5	(very much)

C.4 Do you think that your communication improved with every scenario?

(no, not at all) 1 2 3 4 5 (very much)

C.5 During the entire experiment, do you think that the team improved its communication (with every scenario)?

(no, not at all) 1 2 3 4 5 (very much)

C.6 According to your opinion, is there anything that could improve the communication between you and the other team members ?

RISK

R.1 Do you find the RISK tutorial understandable?

(Very easy to understand) 1 2 3 4 5 (difficult)

R.2 Was the RISK tutorial sufficient to become acquainted with RISK for the first time and then perform the experiment?

(very insufficient) 1 2 3 4 5 (perfect)

R.3 Did your skills within RISK (controlling the avatar, communicating, etc) improve with every scenario?

(no, not at all) 1 2 3 4 5 (very much)

To what degree did you find participation in the overall experiment …

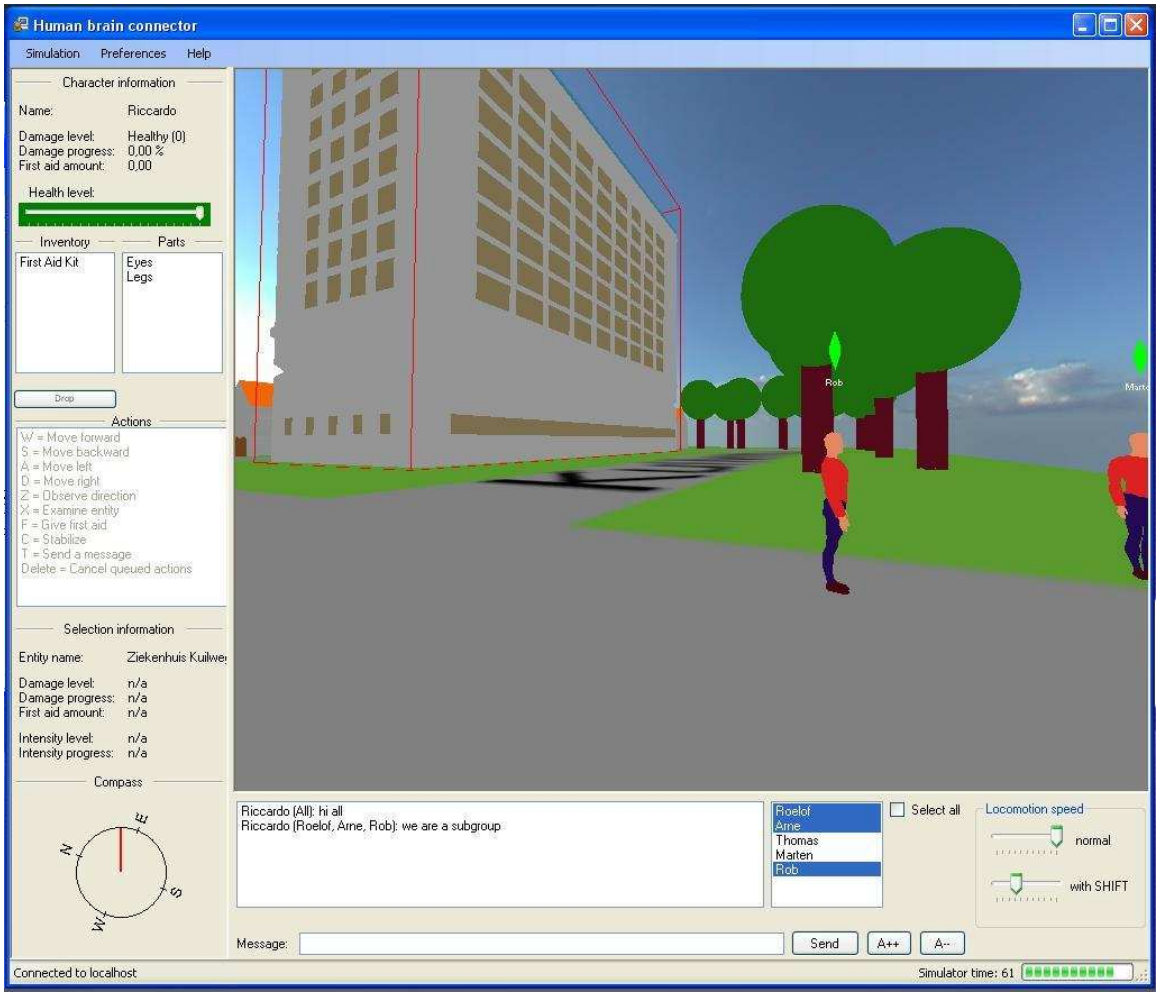
R.4.a …nice / fun

(no, not at all) 1 2 3 4 5 (very much)

R.4.b …. difficult

(no, not at all) 1 2 3 4 5 (very much)

This is a screenshot from RISK's user interface. Is there anything that can be improved? Please draw / indicate / explain your suggestions:



If you have any other positive or negative remarks, please share them with us…:

Remarks concerning the scenario's:
Remarks concerning the RISK Simulator:
Remarks concerning the experiment:

Any additional remarks…: