

# Actor-Agent Team Experimentation in the Context of Incident Management

**Rianne Gouman, Masja Kempen, Niek Wijngaards**

D-CIS Lab / Thales Research & Technology NL

{rienne.gouman, masja.kempen, niek.wijngaards}@d-cis.nl

## ABSTRACT

The collaboration between humans (actors) and artificial entities (agents) can be a potential performance boost. Agents, as complementary artificial intelligent entities, can alleviate actors from certain activities, while enlarging the collective effectiveness. This paper describes our approach for experimentation with actors, agents and their interaction. This approach is based on a principled combination of existing empirical research methods and is illustrated by a small experiment which assesses the performance of a specific actor-agent team in comparison with an actor-only team in an incident management context. The REsearch and Simulation toolKit (RESK) is instrumental for controlled and repeatable experimentation. The indicative findings show that the approach is viable and forms a basis for further data collection and comparative experiments. The approach supports applied actor-agent research to show its (dis)advantages as compared to actor-only solutions.

## Keywords

Experimentation, actor-agent teaming, simulation, performance indicators

## INTRODUCTION

Incident response and management can be characterized by situations that are out of the ordinary wherein humans can benefit from intelligent support. Advances in intelligent systems in incident response and management are reported in each of the past six ISCRAM conferences<sup>1</sup>. An important aspect of the research herein entails the *evaluation* of the collaboration between humans and artificial systems: how can we assess the added value of artificial (autonomous) systems? Our current contribution describes experimentation with humans and software agents applied to incident management from an actor-agent point of view.

The actor-agent point of view comprises the notion of Actor-Agent Communities (AAC). AACs are socio-technical information systems that deliver a solution for otherwise intractable information processing problems (Iacob, Nieuwenhuis, Wijngaards, Pavlin, and Van Veelen, 2009). AACs can be described as groups of humans and artificial autonomous systems that intimately work together to achieve a common goal (i.e. solve a problem). The concept of actor-agent teams (AATs) (Kempen, Wijngaards and Gouman, 2007) is related to actor-agent communities: both human and artificial entities collaborate, working towards a common goal. The usually smaller size of teams (as compared to communities) is of benefit to explore AAC concepts: it is much easier to construct an actor-agent team for experimentation.

Actor-agent teaming is rather novel as a focus in multi-agent-systems research. A welcome addition is a review of teamwork models by Sycara and Sukthankar (2006), wherein the challenges in realizing an actor-agent team are emphasized. Creating a shared understanding between actors and agents in an AAT is deemed the biggest overall challenge. Often, research with humans and agents (in teams) involves investigating the human in the role of supervisor, and the agents in the role of subordinates (with degrees of autonomy). An example is a human viewing a monitor with a geographic display, status information of agents controlling UAVs, and the agents' mission performance (Sycara and Lewis, 2004). In this setting the a priori hierarchical organization is a constraint on the

<b>Reviewing Statement:</b> This paper has been fully double blind peer reviewed
--

---

<sup>1</sup> See [www.iscram.org](http://www.iscram.org)

emergence of teaming aspects and the behaviors of the human and agents cannot be observed in the same manner. From our perspective, actor-agent teams can be more than actor-only teams with tooling support (such as a human supported by a spelling checker or a pda) and agents-only teams (with tooling). We presume that an AAT is more than the sum of its constituents and that agents can have a position in a team that is similar in (social) status to that of humans. From this perspective agents are not to replace humans in a team, rather they fulfill the role of (complementary) team-members by also contributing to the team's objectives.

Our challenge is to establish empirically whether AATs perform or behave differently (or better?) compared to teams consisting of humans only (actor-only teams: AOTs). In our research we focus on actor-agent teaming (currently without pre-defined organizational structures) to investigate effectiveness of actor-agent teams. Especially team performance over time, in changing circumstances such as those encountered in incident management, is of interest when comparing AOTs and AATs.

This paper describes our approach to investigate, evaluate and compare actor-agent teaming interactions with actor-only teams. This paper starts with an outline of our approach to actor-agent experimentation, followed by a brief description of the REsearch and Simulation toolKit (RESK). To illustrate our approach, a specific, small-scale actor-agent team experiment is described. This small experiment aims to compare performance between actor-only teams and actor-agent teams. Finally, our experiences with the approach and suggestions for improvements are discussed.

## ACTOR-AGENT EXPERIMENTATION

Repeatable experiments require a controlled environment so that a comparison between two situations (the controlled and a manipulated situation) becomes possible. The incident management domain, however, is inherently difficult for experimentation because of the dynamics of the incident management domain, the difficulty to conduct real-time observations (let alone to conduct experiments which 'cause' an incident), the unfeasibility of repeatable experimentation, and the constraints (in time and effort) to conduct research experiments during incident management exercises/training sessions. A further restriction is that controlled experiments are impossible: one cannot manipulate a variable (e.g., substitute human team-members by (autonomous) agent team-members) and "handle" the same crisis. Nevertheless, real-world, or 'natural', experimentation in incident management is a 'one-shot' means to collect useful data.

Besides constraints to experimentation in the incident management domain, the presence of agents as team-members in an AAT also poses challenges regarding the experimental design and analysis. Within the distributed AI and multi-agent communities, there is no established principled experimentation approach – as e.g. reported on studying experimentation reports in AAMAS-2005 (by Smit, Kempen, Meyer and Wijngaards, 2006). To explore actor-agent teaming we adopt the empirical approach to acquire data about the world, as commonly used in the natural sciences and social sciences.

Given the objectives of our actor-agent teaming research an experiment is designed in which both actors and agents can perform incident management tasks (in our case: first aid provisioning and fire extinguishing) in a simulated environment. The classical experimentation designs from the natural and social sciences are used for our comparative experimentation as follows (see also Table 1 later). The same human participants partake in two team types: in two scenarios as an AOT (not involving the agents at all) and in two scenarios as AATs (also involving the agent participants) while keeping a comparable team size of six. A minimum of two groups of participants allows for a balanced design, given four scenarios for teams to act in. Part of the challenge is to determine performance indicators and to collect sufficient data for current and future analysis. The experiment is not devised to test theories on actor-agent teaming. Its explorative character is in line with the vision of Axelrod (1997, p. 17) on simulation:

"To appreciate the value of simulation as a research methodology, one should think of it as a new way of conducting scientific research. Simulation can be compared with the two standard methods of induction and deduction. Induction is the discovery of patterns in empirical data. (...). Deduction, on the other hand, involves specifying a set of axioms and proving consequences that can be derived from those assumptions. (...). Simulation is a third research methodology. Like deduction, it starts with a set of explicit assumptions. But unlike deduction, it does not prove theorems. Instead, simulation generates data that can be analyzed inductively. Unlike typical induction, however, the simulated data come from a specified set of rules rather than direct measurement of the real world. While induction can be used to find patterns in data, and deduction can be used to find consequences of assumptions, simulation modeling can be used to aid intuition. Simulation is a way of performing thought experiments."

Tooling is needed to support the experimentation design while also enabling actual team-performance. Our choice is a virtual environment in which a team is to act: a controllable environment in which scenarios can be repeated for different AOTs and AATs. To focus the comparison on team characteristics, instead of on special capabilities of actors (e.g., creativity) or agents (e.g., fast numerical problem solving), a 'level' playing ground is needed, wherein both entities can participate on equal footing. Both actors and agents are not assumed to be part of the simulation environment: they need to control their virtual embodiments.

Multiple virtual environments are available in which human and artificial entities can control avatars (and other objects), e.g. consider the packages Second Life and Unreal Tournament<sup>2</sup>. In such packages scenarios (or 'games') can be defined for a group of players (human and/or artificial) with emphasis on immersive graphical details. The extent of logging, as well as repeatability of the experiment may leave to be desired, next to a potentially large effort spent on graphic detail. Such immersive environments can be used to assess, e.g., the effect of map orientation on people's navigation skills instead of having participants walk outdoors (Smets, te Brake, Lindenberg and Neerincx, 2007). Furthermore, actors and agents need to be embodied and 'restricted' as in games often non-player characters are heavily scripted with full access to the current game state (including player positions and actions). In our actor-agent experimentation the embodied agents need to rely on their (possibly defeasible) perceptions to acquire information on their environment and team-members.

The envisioned tooling for actor-agent experimentation differs from agent-based simulations where a multi-agent system (in a virtual environment) represents a (micro-level) model and the emergent results are often studied at the macro-level (e.g., Sierhuis, Clancey and Hoof, 2007; Klügl, 2008). In such simulations data is collected, interpreted and analyzed to assess certain criteria ('performance'). The simulation of an AAT by an agent-based simulation is not our objective. Using an agent-based simulation framework (which is capable of handling human interaction with its agents) to host our AAT presents a number of difficulties, including actor and agent embodiment, shared virtual environment, communication & coordination and may require our agents to be modeled and implemented in a specific manner. Even when humans can interact with agents in an agent-based simulation, the net result is not in general an actor-agent team.

## **SIMULATED ENVIRONMENT & EXPERIMENTATION TOOLING**

The simulated environment for AAT experimentation is supported by a package of real-time simulation software geared toward simulation of incident management in urban settings: RESK (REsearch and Simulation toolKit<sup>3</sup>). For this purpose, a simulator component has been developed which can simulate an incident management environment in which actors and agents can both be situated and act. The requirements on the simulator are:

1. The simulator has to simulate aspects of the incident management domain sufficiently realistic without overloading the actors and agents with too much detail and intricacies.
2. The simulator needs to be capable to handle the actions of actors, 'accommodate' agents and their actions, and provide an environment in which sufficiently realistic actor-agent interaction can exist.
3. The simulator must log all communications between the actor or agent team-members, environment dynamics, and time and effect of actions that are conducted by actors or agents, etc.
4. The simulator must allow repeatable and controlled experimentation.

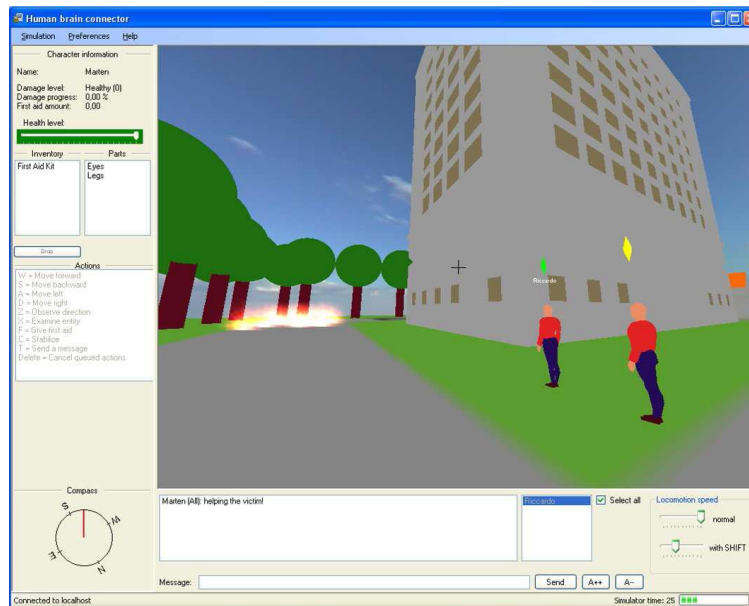
RESK is a discrete event simulator in which participants (actors and/or agents) control their embodiment (or 'avatar') in a 3D-world. Multiple visualizations are possible on the simulated environment, including a 3D first-person perspective and a 2D overview perspective. The graphical rendering is rather abstract and not to be compared with (commercial) games. Figure 1 shows the actor-interface to control an avatar (a 1<sup>st</sup> person perspective).

This simulated environment is sufficiently rich in terms of number of objects, uncertainty in observations and/or action success, but not in (graphical) detail to require problem solving capabilities of both actors and agents. In our natural world actors and (embodied) agents have rather different physical abilities; in the simulated environment these differences can be leveled: both actors and agents control similar 'avatars'. A specific operational setting is chosen that is corresponding to incident management: both actors and agents' avatars are capable of extinguishing

<sup>2</sup> See [www.secondlife.com](http://www.secondlife.com) and [openut.sourceforge.net](http://openut.sourceforge.net)

<sup>3</sup> RESK is open-source and developed within the ICIS project (see acknowledgements). RESK is available at <https://mantis.decis.nl> – registration required.

fires and/or providing aid to victims in a simulated village. The emphasis on operational activities, combined with the simplified objects in the simulated environment, simplifies the design of the agents as no general world knowledge is needed, nor are complex sensor and interpretation methods needed for agents to ‘understand’ their surroundings. Coordination within a team is (assumed to be) about planning, exploration and remediation. Communication is logged through the chat function – audio is currently not allowed. This design choice limits humans in their communication by taking away the ability to use speech. Understanding the consequences of different communication modalities is not part of our current research.



**Figure 1. RESK graphical interface for avatar control by an actor, including mobility and communication actions as well as perception actions (with rendering of results from avatar’s optical sensor).**

RESK is not intended as an immersive environment, neither in graphical quality nor in ‘responsiveness’: discrete time ticks are used. Each simulator time-tick corresponds to about 5 seconds of real-world time. This discrete time works well to support human participants to communicate (typing messages) without having to continuously control their avatar. Communication is by ‘chat’ messages, where the sender has the choice to send the message to all team-members or a specific subset. RESK also includes components for logging, scenario editing, adjustable ‘world-physics’ and adjustable effects of actions (e.g. actions might fail).

### SMALL ACTOR-AGENT EXPERIMENTATION EXAMPLE

Our approach to actor-agent team experimentation is currently illustrated by one small actor-agent team comparative experiment intended to investigate whether actor-agent teams *outperform* agent-only teams in an incident management context. The limited size of the experiment does limit the usefulness of its results as well as prevents a full validation of the approach and (RESK) tooling. The purpose of the small experiment is to *test* the approach, including the usability of the involved (RESK) tooling: allowing for later, larger-scale comparative experimentation.

First the hypotheses and their measurements are briefly described, followed by the specific experimental design, scenarios and experiment. A brief analysis and results conclude this section. A detailed description of the entire experiment is available at (Gouman, Kampman, Kempen, and Wijngaards, 2009).

### Hypotheses and analysis

The hypothesis is: *An Actor-Agent Team (AAT) has a better performance than an Actor-Only team (AOT)*. We postulated four indicators which are to be tested for their validity for comparing AOT and AAT ‘performance’. These indicators are:

1. Delay between reporting and extinguishing a fire or reporting and stabilizing a victim

2. Time of discovering and reporting fires and victims
3. Time of solving the incidents in a problem area
4. Amount of communications used to coordinate actions

These indicators result in the four following sub-hypotheses:

- A.1 An AAT has a *shorter* delay between reporting and extinguishing a fire or reporting and stabilizing a victim
- A.2 An AAT discovers and reports fires and victims *earlier* than an AOT
- A.3 An AAT solves the incidents in a problem area in a *shorter* time than an AOT
- A.4 An AAT uses *less* communication to coordinate their actions than an AOT

To test these sub-hypotheses all actions of the participants (actors and agents) at every time-step are logged in RESK log files, including chat communications, as well as logs of the simulated environment dynamics (including fire intensities and victims' health). These log files are available for analysis and can be 'played back' in a RESK log file viewer.

In addition four sub-hypotheses are defined regarding subjective experiences by the human participants:

- B.1 Team-members of an AAT experience *more* situational awareness than in an AOT
- B.2 Team-members of an AAT experience *more* organizational awareness than in an AOT
- B.3 Team-members of an AAT experience *more* monitoring on performance and task execution than in an AOT
- B.4 Team-members of an AAT experience a *better* overall performance than in an AOT

Questionnaires were used to assess the human participants' opinions on relevant topics (Forsyth, 2006; Levi, 2007; Salas, Dickinson, Converse, and Tannenbaum, 1992; West, 2004). These topics include scenarios, situational awareness, organizational awareness, monitoring of each other's feelings and performance, performance, workload, communication, and human-agent interaction. An additional questionnaire investigated human participants' opinions on communication and agent technology. Both questionnaires contained structured and open questions with a grand total of 35 questions. The structured questions could be answered by indicating a choice on a 5-point Likert- scale; the open questions could be answered in a text box.

## Experimental Design

The experiment was conducted in two sessions. The overall research objective of AAT comparative experimentation implies that a team is the largest unit of analysis and its individual team-members the smallest unit of analysis. In our experiment 12 actors and 12 agents in total participated in 2 groups (divided in 6 actors and 6 agents in each group). The 12 human participants consist of 6 direct colleagues, 2 long-time visiting PhD students (from European countries) and 4 graduation students; 8 of them are male, 4 female. Their ages lie between 18 and 50.. The 'homegrown' software agents are explicitly designed as team-members in a RESK-simulated urban setting (Kuijpers, 2009). The human team members communicate in English (also reflecting the non-Dutch origin of some participants); the agents communicate with the actors in (a limited vocabulary derived from) English. In this experiment, the agents are not the object of study: their effective teaming is studied here.

The six actors in group 1 participated in two scenarios as an AOT. Subsequently, this group was split forming two separate AATs consisting of three actors and three agents. It is important to note that within group 1 the *same* actors (participants a, b, c, d, e, f) participated in the AOT and in their two AATs. Similarly, within group 2 the *same* actors (participants g, h, i, j, k, l) participated in their AOT and in their two AATs. In actor-agent teams, the *same* agents participated (without learning capabilities). The design (i.e., first part of AOT then part of AATs) was repeated reversely for the six human participants in group 2, as Table 1 displays:

	Scenario	A	B	C	D
Group 1	Number of teams	1 team	1 team	2 teams	2 teams
	Type of team(s)	AOT	AOT	AAT	AAT

<b>Group 2</b>	<b>Humans participating</b>	6 human actors	6 human actors	3 human actors	3 human actors
	<b>Agents participating</b>	-	-	3 agents	3 agents
	<b>Number of teams</b>	2 teams	2 teams	1 team	1 team
	<b>Type of team(s)</b>	AAT	AAT	AOT	AOT
	<b>Humans participating</b>	3 human actors	3 human actors	6 human actors	6 human actors
	<b>Agents participating</b>	3 agents	3 agents	-	-

Table 1: Design of the actor-agent teaming experiment.

### Scenarios

For our experimentation the prototype-world “Demoville” is created, in which four scenarios are situated. Demoville contains streets, houses, other buildings, trees, surface (roads and grass), cars, fences, fires and avatars (including victims). Avatars are controllable RESK entities, under control by actors and agents outside of RESK during an experiment. Avatars of victims are controlled by scenario-specific ‘victim-scripts’ which regulate the decay of their health when not receiving medical treatment. Fires in the scenario are controlled by scenario-specific ‘fire-scripts’ which regulate the fire’s intensity and effect of extinguishing actions. All four scenarios distinguish two team-member avatar roles: paramedic and firefighter. Each scenario is played for 15 minutes, yielding four data sets per group for analysis. To keep the scenarios new for the participants each scenario takes place in a different section of Demoville with a different storyline to avoid learning effects. To keep the scenarios comparable to each other the number, location, and severity of fires and victims were kept similar. Actors and agents did not need to switch roles (paramedic vs. firefighter) between scenarios and teams. Scenarios A and C are designed to be relatively simple (e.g., fewer fires) than scenarios B and D. Table 2 presents an overview of the design guidelines of the scenarios:

Scenario	A	B	C	D
<b>Description</b>	Smoke fumes are emerging from the small town of DemoVille.	Due to a serious fire a toxic cloud has formed.	A tornado has left a path of destruction.	A public event has caused chaos.
<b># fires</b>	3	4	3	4
<b># victims</b>	8	8	8	8
<b>Distribution of # victims per fire</b>	4 / 2 / 2	2 / 2 / 2 / 2	4 / 2 / 2	2 / 2 / 2 / 2

Table 2. Overview of Scenarios for small Actor-Agent Team experiment

### Experiment

Before the experiment started the human participants were introduced to the purpose and the schedule of the experiment. Then a tutorial on the user interface of RESK was presented and demonstrated, including time for questions. The tutorial also explained the names and roles of the team-member avatars. A 20-minute training session allowed the human participants to become familiar with the RESK user interface. Subsequently they were seated in two rooms. The participants did not change desks or rooms during the experiment. The seating arrangement was such that the participants could not directly see each other. The actors were requested not to speak to each other and to communicate solely through the chat function in the simulator.

All scenario-based teaming sessions had a duration of 15 minutes and were accompanied by a brief fact sheet about the other team-members’ roles. Each scenario was followed by a questionnaire that also took about 15 minutes to complete. Between scenario A and B, and between scenario C and D there was a 5-minute break, between scenario

B and C all participants had a 15-minute break, after which they received a short explanation about splitting up into actor-agent teams or merging into an actor-only team. The experiment ended with a short debriefing for all who were involved. The entire experiment took about 3.5 hours to complete. 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 RESK scenario sessions were restricted to 15 minutes.

### Analysis and Results

Through analysis of the data the sub-hypotheses on performance comparison (A1, A2, A3 and A4) and subjective perception of actor-agent interaction (B1, B2, B3 and B4) could be assessed. As there were only two test groups with a limited number of human participants *no* elaborate statistical analyses could be performed: the analysis of the results should be considered indicative in the context of this small experiment. The RESK log files and the questionnaires rendered more data than can be discussed in this contribution. A complete discussion of the results can be found in (Gouman et al., 2009). The sub-hypotheses' assessment is summarized in Table 3 below.

Sub-hypotheses		Result
A.1	An AAT has a shorter delay between reporting and extinguishing or stabilizing a fire or victim respectively	Confirmed
A.2	An AAT discovers and reports fires and victims earlier than an AOT	Not confirmed
A.3	An AAT solves the incidents in a problem area in a shorter time than an AOT	Incomparable results <sup>4</sup>
A.4	An AAT uses less communication to coordinate their actions than an AOT	Confirmed
B.1	Team-members of an AAT experience more situational awareness than in an AOT	Not confirmed
B.2	Team-members of an AAT experience more organizational awareness than in an AOT	Not confirmed
B.3	Team-members of an AAT experience more monitoring on performance and task execution than in an AOT	Not confirmed
B.4	Team-members of an AAT experience a better overall performance than in an AOT	Not confirmed

**Table 3. Overview of sub-hypotheses assessment.**

An example of the analysis of actor-agent performance is the following:

#### *A.1 An AAT discovers and reports fires and victims earlier than an AOT*

The data confirm sub-hypothesis A.1. To be more precise: over all the scenarios the AOTs had a mean report time of 101 simulated time ticks versus 83 for the AATs. The reason for this is twofold. Agents have the capability of making an 'observe direction' action every five time ticks: when an agent perceives a fire or victim it will instantly broadcast this information, so the incidents are reported earlier by agents. Humans on the other hand, have the tendency to ignore this perceptive action and first walk towards a situation or already start extinguishing a fire or stabilizing a victim, after which they report this to their team-members.

An example of the analysis of the subjective perception of actor-agent interaction is the following:

---

<sup>4</sup> The comparison was rendered unfeasible, as the teams used a different order in which to solve the problem areas and didn't finish some problem areas while other teams did solve those problem areas.

### *B.3 Team-members of an AAT experience more monitoring on performance and task execution than in an AOT*

Sub-hypothesis B.3 is not confirmed. Pilot experiments already indicated that in the actor-agent team experiments social and task monitoring hardly were part of the teamwork (Gouman, Kampman, Kempen, Kuijpers, and Wijngaards, 2008). Sub-hypothesis B.3 describes the assumption that the presence of agents (who specifically communicate about their tasks and locations) might enhance the general feeling of task monitoring among the team-members. This, however, proved to be false: the participants had low scores on task monitoring.

Given these findings, the four postulated indicators seem valid for comparing AOT and AAT performance in the context of these scenarios and avatar roles. Additional data must be collected to further ground this validation. The main hypothesis of the experiment (*An Actor-Agent Team (AAT) has a better performance than an Actor-Only team (AOT)*) could not be convincingly verified. Comparison of the scenarios shows that the AATs do not really outperform the AOTs. It is important to note that they do *not perform worse* either. Further analysis shows that the current implementation of teaming and social capabilities can, and should be, improved.

### *Human participants' feedback*

In additional remarks the human participants commented on the experiment, on RESK, and on actor-agent teaming. Their remarks can be summarized as follows:

- The scenarios and tutorial were found to be understandable and some issues with the RESK actor-avatar-interface were mentioned. Participants needed time to get used to the simulator's discrete time-ticks, controlling of the avatar by the keyboard, navigation (through the streets of Demoville) and orientation. We assume this issue has not altered the results, as participants in the AOTs as well as the AATs had to cope with the same interface.
- Participants mentioned that they perceived little organization in both the AOTs and AATs. The reason given most often was that the participants were so busy with their tasks in the simulated scenario that the participants did not have enough time to look at the fact sheet and sometimes became confused about the other team-members' names and roles.
- The participants were (very) positive about the team's communication: they felt that the team communicated sufficiently about plans and decisions, even more so about the incidents in the scenarios. They also believed that their own communication to others and the entire team communication improved throughout the experiment.
- Regarding actor-agent interaction within the experiment most participants were neutral to questions about whether they explained information to agents, received feedback from agents, whether agents were being helpful, whether they understood the agents' intentions, etc. The participants claimed they followed the directions of humans more than those of agents.
- Regarding their personal beliefs about agent-technologies the participants were neutral on the positions "an agent can lead me" and "agents can work together in a team". The participants valued the idea that they "can treat agents like humans" positively, and supported the notion that an agent can help a human.

## **DISCUSSION**

Our approach for comparing teaming performance of AOTs and AATs has been illustrated with a specific, albeit small, experiment whose overall results are (unfortunately) inconclusive. The experimentation approach has currently proven useful: controlled and repeatable experimentation is now relatively easy to achieve for actor-only teams and actor-agent teams in incident management context. This is notably due to the functionalities offered by RESK and careful design of the incident management scenarios.

The combination of (objective) data collection through RESK logging and (subjective) questionnaires was sufficient for the current experimentation needs. The performance indicators as we determined in this small experiment seem to be useful for testing our current hypothesis. The questionnaires provided insight in the human participants'



situational and organizational awareness, social and task monitoring, the perception of their performance, and other relevant topics. Especially the open questions provided explanations to the participants' opinions. The degree to which teamwork can be achieved in the short time that humans perform within the scenarios while also becoming used to the simulator's interface is of some concern. With more data (acquired through more sessions with RESK) and improved agents, we should be able to unravel more potential measures and validate more subtle hypotheses.

RESK functioned well as an experimentation and simulation tool and received overall positive comments (Gouman et al., 2009). In addition, a number of improvements were elicited, most to do with details regarding the graphical interface and avatar-controls. For example the simulator's discrete time ticks were intended to allow actors the time to type chat messages without needing to continuously control their avatar. The participants needed to get used to 'non-smooth' avatar interaction and query of actions. These are features to be further tested.

The current set-up with RESK has delivered in our opinion a sound experimental toolkit for future comparative experimentation. Such further comparative experimentation can investigate the effects of:

- Different embodiments and 'physical' capabilities of agents (e.g., fire truck, UAV)
- Adjustable uncertainty (regarding effects of actions on perception, communication, object manipulation, locomotion, etc.).
- More complex scenarios with larger environments and multiple AATs.
- Resource usage by actors and agents for more complex team-performance metrics.

The current usefulness of our approach has a number of implications for actor-agent research. The added value of complex socio-technical systems, including actor-agent teams and communities, needs to be made explicit to both end-users (who can fulfill the role of actors) and businesses (who can develop the infrastructures and tooling to support actor-agent systems). The approach described here provides a specific format to highlight actor-agent interaction and the effectiveness thereof in comparison to actor-only teaming. The current (simplified) incident management context yields a number of simple performance indicators related to the scenarios and avatar roles' involved. Establishing more generic criteria with robust predictive 'powers' may lead to advances on endowing agents with the ability to monitor and predict team-performance. Furthermore, the relatively simple-to-reuse experimentation environment supports research and rapid prototyping of more advanced agents for actor-agent teams. Finally, the ability to re-analyse and reflect on data collected during past experimentations facilitates future experimentation and results analysis.

## CONCLUSION

This paper describes an approach for comparing actor-agent teaming and actor-agent teaming in an incident management context. Experimentation with actors and agents requires an environment in which factors that influence actor-agent interaction can be controlled and in which the experiments can be repeated. A simulated incident management environment is a reasonable alternative for such experimentation in comparison to (currently infeasible) repeatable and controllable actor-agent experimentation in the natural world. Our current, albeit limited, experiences with the experimentation approach prove to be useful for experimentation with actors and agents, data collection and comparison of results. The small experiment conducted to test-drive our approach has inconclusive experimentation results as we expected, given the limited scale of the experiment and the agents' relatively simple teaming capabilities. Further work on comparative AAT and AOT experimentation involves investigating an increment of the group size, improvement of the agents' team-oriented behavior, and the effects of scenario complexity and duration. The current experiences with the approach for actor-agent team experimentation enable further actor-agent research, development and comparative experimentation, together with improvements on the RESK tooling.

## ACKNOWLEDGMENTS

The research reported here has been conducted within 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 modeling, and the integration of Machinetta and 2APL with RESK the authors thank Roelof Kuijpers (Utrecht University). For his work on conducting actor-agent experiments the authors thank Marten Kampman (Utrecht University). The authors 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 Rob van der Most, Osman Sarar, Joris Sikkema (Hogeschool Rotterdam), Thomas Hood, Eddy van der Heijden (D-CIS Lab) and Elmer van Chastelet and Hugo Burm (Y'All) for their work on RESK. We are indebted to the participants who invested their time willingly into our experiment, thereby helping us in the process to understand more about AAT experimentation. Finally, the reviewers' comments led to a number of improvements in our text.

## REFERENCES

1. Axelrod, R. (1997). Advancing the Art of Simulation in the Social Sciences. *Complexity*, 3(2):16-22 Wiley Periodicals, Inc.
2. Forsyth, D.R. (2006) Group Dynamics, 4th Edition. Belmont, CA: Thomson Wadsworth.
3. Gouman, R., Kampman, M., Kempen, M., Kuijpers, R., Wijngaards, N. (2008) SEAT Actor-only Team Try-out Experimentation of July 2008. *D-CIS Research Report Series*, no. D-CIS RR/2008-001, available at <http://publications.decis.nl/>.
4. Gouman, R., Kampman, M., Kempen, M., Wijngaards, N. (2009) SEAT Actor-Agent Team Experimentation. *D-CIS Research Report Series*, no. D-CIS RR/2008-001, available at <http://publications.decis.nl/>.
5. Iacob, S.M., Nieuwenhuis, C.H.M., Wijngaards, N.J.E., Pavlin, G., Van Veelen, J.B. (2009) Actor-Agent Communities: Design Approaches. In Papadopoulos, G.A, Badica, C. (Eds.) *Intelligent Distributed Computing III*, SCI 237, pp. 237–242.
6. Kempen, M., Wijngaards, N., Gouman, R. (2007) Sustained Effectiveness of Actor-Agent Teams: Explorations and Demarcations. *D-CIS Research Report Series*, no. D-CIS RR/2007-002, available at <http://publications.decis.nl/>.
7. Klügl, F. (2008). A validation methodology for agent-based simulations. In: Wainwright R.L. and Haddad, H. (editors), *Proceedings of the 2008 ACM Symposium on Applied Computing (SAC)*, Fortaleza, Ceara, Brazil, pp. 39–43.
8. Kuijpers, R. (2009) Communication and Coordination in Actor/Agent Teams. *Msc. thesis*, Utrecht University, Dept. of Philosophy. Available at <http://publications.decis.nl/>.
9. Levi, D. (2007). Group Dynamics for Teams. Thousands Oaks, CA: Sage Publications.
10. Salas, E., Dickinson, T.L., Converse, S., Tannenbaum, S.I. (1992) Toward an understanding of team performance and training. In: Swezey, R.W., Salas, E. (eds.) *Teams: Their Training and Performance*. Norwood, (NJ): Ablex. pp. 3–29.
11. Sierhuis, M., Clancey, W.J., Hoof, R. v. (2003). Brahms: a multi-agent modeling environment for simulating social phenomena. In *Proceedings of First conference of the European Social Simulation Association (SIMSOC VI)*, Groningen, The Netherlands.
12. Smets, N.J.J.M., te Brake, G.M., Lindenberg, J., Neerincx, M.A. (2007). Influence of mobile map size and user capacities on situation awareness tested in a virtual environment. In: Van de Walle, B., Burghardt, P. Nieuwenhuis, C. (Eds), *Proceedings of the 4th International Conference on Information Systems for Crisis Response and Management ISCRAM 2007*, pp. 557–564.
13. Smit, A., Kempen, M., Meyer, A., Wijngaards, N. (2006), The Hidden Dangers of Experimenting in Distributed AI. In: Stone, P. and Weiss, G. (Eds), *the Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, Future University-Hakodate, Japan, pp. 1320–1322.
14. Sycara, K., Lewis, M. (2004) Integrating intelligent agents into human teams. In: Salas, E., Fiore, S. (eds.) *Team Cognition: Understanding the Factors that Drive Process and Performance*, Washington, DC: American Psychological Association.
15. Sycara, K., Sukthankar, G. (2006) Literature Review of Teamwork Models. *Technical report CMU-RI-TR-06-50*, Robotics Institute, Carnegie Mellon University.
16. West, M.A. (2004). Effective Teamwork: practical lessons from organizational research. Malden, MA: BPS Blackwell.