

**INVESTIGATING INTERACTIONAL ISSUES OF AUTOMATED
PLANNING SUPPORT FOR DISASTER RESPONSE**

WENCHAO JIANG. BSC

Thesis submitted to the University of Nottingham for the degree of Doctor of
Philosophy

September 2015

ABSTRACT

This thesis contributes to the understanding of the potential socio-technical issues that can emerge from the interaction between responder teams and automated planning support, which in turn, leads to design implications for dealing with the emerged issues.

Recently, natural and man-made disasters in Haiti, Chile and Japan drew the attention of researchers of disaster management systems. A lot of efforts have been made to study the technologies that can assist human responders to improve their performance. In the Disaster Response domain, a Disaster Response team, which typically contains several incident commanders and field responders, are faced with the problem of carrying out geographically distributed tasks under spatial and temporal constraints in a quickly changing task environment. Recent advances in multi-agent technologies lead to the possibility of building planning support for team coordination by automating the task planning process. However, it is unknown how the planning support system can fit into the team organisation in a way that improves rather than hinders the team performance.

This PhD work adopts a serious mixed reality game approach as the vehicle to explore socio-technical issues in complex real world settings. We developed AtomicOrchid, an emergency response game to create a task setting which mirrors aspects of real world Disaster Response operation. In the game trials, participants are recruited to play as field responders and incident commanders to carry out rescue missions. Participants' experiences are observed and recorded as they coordinate with each other to achieve game objectives, with the support from an intelligent planner agent. The observations gathered in three AtomicOrchid field trials reveal how the organisational work is organised with and without automated planning support; how the

guidance from the planner supports/interrupts natural human workflow; how the human mediation (between planner and responders) played out to support the team work for better or worse.

In the first study, field responders and Headquarters (HQ) coordinate without support of the intelligent planner. The result showed the team planning is dominated by local coordination between field players with in a “situated” manner. The HQ is observed to successfully provide situational awareness to the field teams through remote messages. However, the HQ has little direct influence on the planning and actions of field teams. In the second study, an automated planner was introduced to guide field responders directly without involvement of incident commanders. At its best, the agent is observed to take over routine planning activities while the human focus on other issues such as finding teammates, targets and choosing the best routes. However, there also evidence showing the agent planning occasionally interrupts the workflow of the human team potentially because it fails to consider the social cost of task changes. In the third study, the system is modified to support incident commanders mediating task planning activities for field responders and the planning agent. We find that the human coordinator and automated planner agent can successfully work together in most cases, with human coordinators inspecting and ‘correcting’ the agent-proposed plans. However, occasional failures of planning are also observed due to: complacency; silent, missing or invisible information; and limited support for human planning.

The observations of the three studies are further analysed to generate design implications around the themes of balancing labour division, achieving common ground, facilitating accountability, and supporting situated actions.

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my supervisors Prof. Tom Rodden, Prof. Chris Greenhalgh and Dr. Joel Fischer for the continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis.

I would like to thank my family: my wife Jia Fu, and my parents for supporting me spiritually throughout writing this thesis and my life in general.

CONTENTS

| | | |
|----------|--|-----------|
| 1 | INTRODUCTION | 1 |
| 1.1 | Problem definition and objectives | 2 |
| 1.2 | Approach | 4 |
| 1.3 | Scoping | 5 |
| 1.4 | Research questions | 7 |
| 1.5 | Contributions | 8 |
| 1.6 | Publications of this thesis | 9 |
| 1.7 | Structure of the thesis | 10 |
| | | |
| i | BACKGROUND | 12 |
| | | |
| 2 | TASK PLANNING IN DISASTER RESPONSE | 13 |
| 2.1 | Task planning in Disaster Response | 13 |
| 2.1.1 | Defining Disaster Response | 13 |
| 2.1.2 | DR command structure | 15 |
| 2.1.3 | Task planning in large scale disasters | 16 |
| 2.2 | Command and control environment | 17 |
| 2.3 | Technological practices in Disaster Response | 19 |
| 2.3.1 | GIS-based planning support systems | 19 |
| 2.3.2 | ICT support for Disaster Response | 22 |
| 2.3.3 | Application of AI technologies | 25 |
| 2.4 | Summary | 28 |
| | | |
| 3 | INTERACTION BETWEEN HUMAN AND PLANNING SUPPORT SYSTEM | 29 |
| 3.1 | A socio-technical perspective on technology support | 30 |
| 3.1.1 | The socio-technical gap for planning support system | 31 |
| 3.1.2 | Summary | 33 |
| 3.2 | Other perspectives on human-system interaction | 34 |
| 3.2.1 | Automation and its impact on human performance | 34 |
| 3.2.2 | Human-Agent interaction | 37 |

| | | |
|-------|--|----|
| 3.2.3 | Summary | 41 |
| 3.3 | Games as an approach to study human-system interaction | 41 |
| 3.3.1 | Serious Games | 42 |
| 3.3.2 | Mixed Reality Games | 43 |
| 3.4 | Summary | 45 |
| ii | METHODOLOGY AND APPROACH | 47 |
| 4 | APPROACH | 48 |
| 4.1 | Adopting the socio-technical perspective | 49 |
| 4.2 | Serious Mixed Reality Games as a testbed | 51 |
| 4.3 | The AtomicOrchid Platform | 52 |
| 4.3.1 | Game mechanics | 53 |
| 4.3.2 | System architecture | 56 |
| 4.3.3 | The planning agent | 57 |
| 4.3.4 | Logging system | 58 |
| 4.3.5 | Iterative design and development | 59 |
| 4.4 | Exploring interaction designs with three AtomicOrchid studies | 60 |
| 4.4.1 | The on-the-loop interaction | 61 |
| 4.4.2 | The in-the-loop interaction | 62 |
| 4.4.3 | The three AO game studies | 63 |
| 4.5 | Collaboration with a professional Disaster Response organisation | 64 |
| 4.5.1 | Introduction of Rescue Global | 64 |
| 5 | METHODOLOGY TO INVESTIGATE HUMAN-AGENT INTERACTION | 66 |
| 5.1 | Ethnomethodological perspective | 66 |
| 5.2 | Interaction analysis | 67 |
| 5.3 | Data collection and handling | 69 |
| 5.3.1 | Video and audio recording | 70 |
| 5.3.2 | Log data handling | 71 |
| 5.3.3 | Message classification | 74 |
| 5.3.4 | Group interview | 75 |
| 5.4 | Analytic procedure | 76 |

| | |
|--|-----|
| iii STUDIES | 78 |
| 6 ATOMICORCHID STUDY 1: NON AGENT VERSION | 79 |
| 6.1 Introduction | 79 |
| 6.2 System description | 80 |
| 6.3 Study design | 82 |
| 6.4 Data analysis and results | 84 |
| 6.4.1 Results of message classification | 85 |
| 6.4.2 Responding to directives from HQ | 89 |
| 6.4.3 Local coordination without HQ | 93 |
| 6.4.4 Remote messages as a resource of situational awareness | 96 |
| 6.5 Discussion | 99 |
| 6.5.1 Division of labour | 99 |
| 6.5.2 Breakdown of remote coordination | 100 |
| 6.5.3 Implications on computational support | 100 |
| 6.6 Design requirements | 102 |
| 6.7 Summary | 105 |
| 7 ATOMICORCHID STUDY 2: THE HUMAN ON-THE-LOOP DESIGN | 107 |
| 7.1 Introduction | 107 |
| 7.2 System evolution | 109 |
| 7.2.1 The planning agent | 109 |
| 7.2.2 A feedback loop | 110 |
| 7.2.3 Interface improvements | 112 |
| 7.3 Study design | 112 |
| 7.4 Data analysis and results | 113 |
| 7.4.1 Assigning task assignments to existing teams | 114 |
| 7.4.2 Task assignments involving team reformation | 116 |
| 7.4.3 Task assignments involving task interruption | 119 |
| 7.4.4 Disagreement on task interruption | 121 |
| 7.4.5 The headquarters | 124 |
| 7.5 Discussion | 124 |
| 7.5.1 Division of labour between the agent and the human teams | 125 |

| | | |
|--------|---|-----|
| 7.5.2 | Hidden costs of team reformation and task interruption | 125 |
| 7.5.3 | Feedback to the agent | 127 |
| 7.6 | Design implications | 128 |
| 7.6.1 | Achieve common ground | 129 |
| 7.6.2 | Facilitate accountability | 129 |
| 7.6.3 | Balance responsibilities between humans and agent | 130 |
| 7.7 | Summary | 130 |
| 8 | ATOMICORCHID STUDY 3: AGENT-SUPPORTED IN-THE-LOOP DESIGN | 132 |
| 8.1 | Introduction | 132 |
| 8.2 | System evolution | 134 |
| 8.2.1 | Interfaces | 135 |
| 8.2.2 | The planning agent | 139 |
| 8.3 | Study design | 140 |
| 8.4 | Data analysis | 142 |
| 8.4.1 | Messaging system | 143 |
| 8.4.2 | Overview of task assignments | 144 |
| 8.4.3 | Confirming the plan | 145 |
| 8.4.4 | Just following the Plan | 147 |
| 8.4.5 | Correcting the plan | 149 |
| 8.4.6 | Changing the plan | 150 |
| 8.4.7 | Coping with the unexpected | 151 |
| 8.4.8 | When it all breaks down | 153 |
| 8.4.9 | Missing feedback from field | 156 |
| 8.4.10 | Division of labour in HQ | 157 |
| 8.5 | Discussion | 157 |
| 8.5.1 | How does the in-the-loop design play out? | 158 |
| 8.5.2 | Responsibility and complacency | 161 |
| 8.5.3 | Tacit confusions | 163 |
| 8.5.4 | Support human planning | 166 |
| 8.6 | Lessons from interaction design | 168 |
| 8.6.1 | Common ground | 169 |
| 8.6.2 | Supporting Mixed-Initiative planning | 170 |
| 8.6.3 | Interactional trouble | 171 |

| | |
|--|------------|
| 8.7 Professional feedback | 171 |
| 8.7.1 The Rescue Global feedback | 172 |
| 8.7.2 Conclusion and reflection | 176 |
| 8.8 Summary | 177 |
| iv CONCLUSION | 179 |
| 9 CONCLUSION | 180 |
| 9.1 Summary of contributions | 181 |
| 9.1.1 Interface evolution | 182 |
| 9.1.2 Key observations | 186 |
| 9.1.3 Interactional issues and lessons | 188 |
| 9.2 Limitations | 197 |
| 9.3 Future work | 199 |
| v APPENDIX | 202 |
| A APPENDIX TO ATOMICORCHID IMPLEMENTATION | 203 |
| A.1 System architecture | 203 |
| A.2 Data model | 206 |
| A.3 Agent-server protocol | 207 |
| B APPENDIX TO STUDY 1 | 208 |
| B.1 Game area and set-up | 208 |
| B.2 Participant demographic | 209 |
| B.3 Game event visualisation | 211 |
| C APPENDIX TO STUDY 2 | 213 |
| C.1 Game area and set-up | 213 |
| C.2 Participant demographic | 214 |
| C.3 Game event visualisation | 216 |
| D APPENDIX TO STUDY 3 | 218 |
| D.1 Game area and set-up | 218 |
| D.2 Participant demographic | 219 |
| D.3 Game event visualisation | 221 |
| BIBLIOGRAPHY | 223 |

LIST OF FIGURES

| | | |
|-----------|---|----|
| Figure 1 | Contributions | 8 |
| Figure 2 | Disaster Management Circle. Credit Wattegama [2012] | 14 |
| Figure 3 | Technological support for Disaster Response | 20 |
| Figure 4 | Tsunami response system, Credit Wattegama [2012] | 23 |
| Figure 5 | Actions and situated actions | 51 |
| Figure 6 | HQ and field players in AO | 53 |
| Figure 7 | The roles in AtomicOrchid (AO) | 54 |
| Figure 8 | The targets in AO | 54 |
| Figure 9 | Role-target mapping | 54 |
| Figure 10 | The radioactive cloud | 55 |
| Figure 11 | System architecture | 56 |
| Figure 12 | Input and output of the agents | 58 |
| Figure 13 | On-the-loop interaction design | 62 |
| Figure 14 | In-the-loop interaction design | 63 |
| Figure 15 | Log visualisation | 72 |
| Figure 16 | Replay system | 73 |
| Figure 17 | Analytic procedure | 77 |
| Figure 18 | The HQ interface | 81 |
| Figure 19 | The mobile responder app | 82 |
| Figure 20 | How responders addressed task allocation messages from HQ. | 90 |
| Figure 21 | Episode 1.1, JH (Behind Left), D2 (Middle Front), KY (Right behind) | 91 |
| Figure 22 | Episode 1.1, KY (Left) , MF (Right) holding mobile phones | 91 |
| Figure 23 | Episode 1.1, KYJH running into cloud | 92 |
| Figure 24 | Episode 1.1, KY, JH turning around to escape from the cloud | 92 |

| | | |
|-----------|---|-----|
| Figure 25 | Episode 1.2, MF (right), BR (middle) | 93 |
| Figure 26 | Episode 1.2, MF (right), BR (left) | 93 |
| Figure 27 | Episode 1.3, MF (right), BR (left) met with D2 (middle) | 95 |
| Figure 28 | Episode 1.4, MF pointing at a building | 97 |
| Figure 29 | Episode 1.5, JH (left) met team D2 (male, middle) and KY (female, right) | 98 |
| Figure 30 | On-the-loop interaction design | 108 |
| Figure 31 | Game engine handling plans from agent | 111 |
| Figure 32 | Mobile task interface in study 2 | 111 |
| Figure 33 | HQ and mobile interfaces in study 2 | 112 |
| Figure 34 | Mobile message interface in study 2 | 113 |
| Figure 35 | Overview of agent instructions | 114 |
| Figure 36 | Episode 2.1, CR (Left) and PC (Right) studying screen together after drop off. | 115 |
| Figure 37 | Episode 2.1, PC (Right) leading the way to new target. | 115 |
| Figure 38 | Episode 2.2, players from left to right: LT, SS, CR, PC. LT walking around the team, her body orientation suggesting attempts to leave the group. | 117 |
| Figure 39 | Episode 2.3, AW (right) leads the way, heading to target 44 as instructed. | 119 |
| Figure 40 | Episode 2.3, After the team received an instruction to disband, AW (right) and HB (left) simultaneously turn back and start walking back to the drop off zone, displaying bodily alignment. | 120 |
| Figure 41 | In-the-loop interaction design | 133 |
| Figure 42 | Messaging system design | 135 |
| Figure 43 | Mobile interface in study 2 | 136 |
| Figure 44 | Situational awareness interface operated by HQ2 | 136 |
| Figure 45 | Interfaces in study 3 | 137 |
| Figure 46 | Edit mode of task allocation interface | 137 |
| Figure 47 | Pop up messaging panel | 138 |
| Figure 48 | Task assignment in session 1 | 144 |

| | | |
|-----------|---|-----|
| Figure 49 | Task assignment in session 2 | 144 |
| Figure 50 | Episode 3.1, HQ1 (left) HQ2 (right) | 146 |
| Figure 51 | Episode 3.2, HQ1 (left) HQ2 (right) | 146 |
| Figure 52 | Episode 3.3, HQ identifies a task interruption, highlighted on right figure | 150 |
| Figure 53 | Episode 3.5, HQ2 points to a team | 152 |
| Figure 54 | Episode 3.5, HQ sends msg for task cancella- tion | 152 |
| Figure 55 | Episode 3.6, CE (left) grabs KM (right), start running into cloud. | 155 |
| Figure 56 | Episode 3.6, CE (left) grabs KM (right) walks around, looking into mobile screen, trying to locate target | 155 |
| Figure 57 | Overview of studies | 180 |
| Figure 58 | Interface of study 1 | 182 |
| Figure 59 | Interface of study 2 | 183 |
| Figure 60 | Task assignment interface study 3 | 184 |
| Figure 61 | Message panel in study 3 | 185 |
| Figure 62 | Mobile Interface of study 3 | 186 |
| Figure 63 | On-the-loop vs in-the-loop | 189 |
| Figure 64 | Technological illustration of AtomicOrchid | 204 |
| Figure 65 | Target distribution in study 1 | 208 |
| Figure 66 | Event visualisation of Study 1, Session A | 211 |
| Figure 67 | Event visualisation of Study 1, Session B | 212 |
| Figure 68 | Target distribution in study 2 | 213 |
| Figure 69 | Event visualisation of Study 2, Session A | 216 |
| Figure 70 | Event visualisation of Study 2, Session B | 217 |
| Figure 71 | Target distribution in study 3 | 218 |
| Figure 72 | Event visualisation of Study 3, Session A | 221 |
| Figure 73 | Event visualisation of Study 3, Session B | 222 |

LIST OF TABLES

| | | |
|----------|---|-----|
| Table 1 | Planning model and situated action | 34 |
| Table 2 | Result for MMDP, Myopic and Greedy algorithms, Credit Ramchurn et al. [2015b] | 57 |
| Table 3 | Overview of game results | 85 |
| Table 4 | Speech act classification | 86 |
| Table 5 | Themes of requests | 88 |
| Table 6 | Adjacency pairs | 88 |
| Table 7 | Overview of game results | 114 |
| Table 8 | Compliance with agent instructions by context | 126 |
| Table 9 | Result overview of study 3 | 142 |
| Table 10 | Task assignment overview | 143 |
| Table 11 | On-the-loop vs in-the-loop | 190 |
| Table 12 | Communications between components | 205 |
| Table 13 | AtomicOrchid DB model | 206 |
| Table 14 | Agent server protocol | 207 |
| Table 15 | Participants in session A, Study 1 | 209 |
| Table 16 | Participants in Session B, Study 1 | 210 |
| Table 17 | Participants in Session A, Study 2 | 214 |
| Table 18 | Participants in Session B, Study 2 | 215 |
| Table 19 | Participants in Session A, Study 3 | 219 |
| Table 20 | Participants in Session B, Study 3 | 220 |

ACRONYMS

- AI Artificial Intelligence
- AO AtomicOrchid
- BSG Command and control model of Bronze-Silver-Gold
- CSCW Computer Supported Collaborative Work
- C₂ Command and Control
- DR Disaster Response
- ERR Emergency Response Rooms
- EM Ethnomethodology
- GIS Geospatial Information System
- HACS Human Agent Collective Systems
- HAC-ER Human Agent Collectives for Emergency Response
- HAI Human Agent Interaction
- HCI Human Computer Interaction
- ICT Information and Communication Technologies
- LoA Level of Automation
- MAS Multi-Agent System
- MRG Mixed Reality Game
- PSS Planning Support System
- RG Rescue Global
- UAV Unmanned Aerial Vehicles
- USAR Urban Search And Rescue

INTRODUCTION

Coordination in Disaster Response ([DR](#)) operations such as Urban Search And Rescue ([USAR](#)) can be very challenging. In large-scale disasters, [DR](#) teams may have limited resource and personnel to deal with multiple incidents across a large impact area. Task planning and execution need to be carried out by geographically distributed [DR](#) teams in real-time against uncertainties in the environment. The challenges highlight the opportunity space of technology support for real-time task planning and execution.

Recently, various Information and Communication Technologies ([ICT](#)), ranging from communication infrastructures to social media platforms, have been playing increasingly significant roles in disaster management. Moreover, Multi-Agent System ([MAS](#)) researchers have devised various real-time task planning algorithms to automate planning in time critical task domains such as Disaster Response. The advances in both [ICT](#) and multi-agent optimisation algorithms lead to the opportunity of automated planning support in the [DR](#) domain. However, before we apply these algorithm to support [DR](#) operations, we need to understand how can we design the interaction between responder teams and the planning support in a way that improves, rather than hinders team performance. Empirical studies of Computer Supported Collaborative Work ([CSCW](#)) systems have shown that it is vital to study technology in use to understand potential tensions between social and technical aspects of a system. In particular, field studies of workflow support systems have revealed that technologies can disrupt rather than smooth workflow if they are not designed in a socially acceptable way [[Bowers et al., 1994](#)]. The aim of this work is to explore potential socio-technical issues surrounding a planning support system, which in turn, informs interaction design

of such systems.

This PhD work is sponsored by the ORCHID research project (EPSRC grant), which contributes to the understanding needed to build Human Agent Collective Systems ([HACS](#)) in the disaster domain. As computational systems becoming increasingly embedded into our life, the researchers from the ORCHID project envision a future in which people and computational agents operate at a global scale, forming human agent collectives. The ORCHID project aims to realise the vision of [HACS](#) by studying the science that is needed to understand, build and apply [HACS](#) that symbiotically interleave human and computer systems¹.

This chapter will give an overview of the PhD work, which covers research objectives, approach, research questions and contributions, followed by a list of publications related to this thesis and an overview of the thesis structure.

1.1 PROBLEM DEFINITION AND OBJECTIVES

In large scale disasters, a [DR](#) team may have limited resources and personnel to deal with large number of incidents across a large geographic area under time pressure. In this situation, task and team allocation become a grand challenge for the [DR](#) team. The responders and resources need to be assigned to teams and tasks in a way that minimises loss of life and costs (e.g., time or money). For instance responders with different capabilities (e.g., fire-fighting or life support) have to form teams in order to perform rescue tasks (e.g., extinguishing a fire or providing first aid). Thus, responders have to plan their paths to the tasks (as these may be distributed in space) and form specific teams to complete them. These teams, in turn, may need to disband and reform in different configurations to complete new tasks, taking into account the status of the current tasks (e.g.,

¹ www.orchid.ac.uk

health of victims or building fire) and the environment (e.g., if a fire or radioactive cloud is spreading). Furthermore, uncertainty in the environment (e.g., road connectivity, task status update) or in the responders abilities to complete tasks (e.g., some may be tired or get hurt) means that plans are likely to change continually to reflect the prevailing assessment of the situation.

Recent advances in multi-agent systems research led to a number of real-time simulation and optimization technologies, some of which have great potential to be adapted to support task planning for DR teams (Section 2.3.3). Although the opportunity space has been recognised, most multi-agent coordination algorithms have only been tested in computational simulations. None of them have been deployed to guide real humans in DR situations. However, extreme difficulties might be encountered when introducing new technology support for human teams. New technologies might not support, but might disrupt smooth workflow if they are designed in an organisationally unacceptable way. Much literature in CSCW has pointed out how ill-designed workflow management/automation system can lead to undesirable results, not only failing to improve work efficiency but also hindering human performance. Field studies of CSCW technologies have shown that it is vital to study technology in use to understand potential tensions raised for teamwork [Bowers et al., 1994].

This PhD work adopts a socio-technical view on the responder teams and their technological supporting systems. The term socio-technical system is used to describe a system that involves a complex interaction between humans, machines and the environmental aspects of the system (Section 3.1). Particularly, the interactional issues with a planning support system can emerge from the different ways in which human and system carry out planning. Typically, problem solving systems are implemented with a scientific planning model while human actions are more situated without plans as necessary prerequisites (Section 3.1). This difference can result in significant is-

sues of human-system interaction.

This thesis follows the long-standing tradition of empirical CSCW study, which investigates complex collaborative work settings and identifies implications for technology support. In order to build automated systems that support human task planning, sufficient empirical studies may be required to understand how can we bridge the gap between the technical support system and social aspects of the human team. The objective of this PhD work is to fill this gap by exploring and unpacking interactional issues surrounding interaction with the intelligent planning system.

1.2 APPROACH

To meet our research objective, we adopt a serious mixed reality games approach (Chapter 4) to create a game probe (i.e. AtomicOrchid) that enables studying team interaction with a planning support system in a disaster scenario whilst providing confidence in the efficacy of behavioural observations. Mixed-reality games are designed to bridge the physical and the digital world. Arguably, they serve as a vehicle to study distributed interactions across multiple devices and ubiquitous computing environments in the wild.

AtomicOrchid (Chapter 4) is a serious mixed-reality game designed to mirror aspects of real-world disaster operations. In this game, field responders use smartphones to coordinate, via text messaging, GPS, and maps, with headquarter players and each other. The players in the game face a distributed task planning problem with both temporal and spatial constraints. To achieve the game objectives, the players need to dynamically change their team configurations. The task planning process in the game is supported by a planning support agent. The planning support agent is based on a state-of-the-art coalition formation optimisation technology. Design and implementation of AtomicOrchid (AO) will be introduced in more details in chapter

4.

In order to explore the interactional issues in automated planning support systems, three studies were conducted with different research focuses. In the first study, field responders and headquarter coordinate without support of the automated planner. The study established a baseline performance of the game play and derived several requirements for planning support systems. In the second and third studies, an automated planner was introduced to support task planning with two different interaction designs. The second study adopted the human “on-the-loop” design pattern in which the planning agent automatically generates plans and instructs field players to execute plans. In the third study, we adopted the human “in-the-loop” design, in which every plan generated by the planning agent needs to be approved and edited before it is sent to field players for execution. More details of these two interaction patterns will be introduced in Chapter

4.

The work also adopted an ethnmethodologically-inspired approach for data analysis. We recorded both system logs and video of interaction in the field trials for analysis. To capture the distributed, concurrent nature of the interaction, four researchers with camcorders shadowed the field player teams. Qualitative interaction analysis was carried out on the collected data to provide thick descriptions of human-system interaction and unpack issues related to the interactions.

1.3 SCOPING

This thesis is relevant to several research areas.

- Human Computer Interaction ([HCI](#)). [HCI](#) is the overarching research area of the PhD work. This thesis follows a long-standing tradition of empirical [CSCW](#) study, which investigates complex

collaborative work settings and identifies implications for technology support. Two other sub domains of **HCI**, namely automation design and human agent interaction, provide framework models and terminologies to guide the interaction design in the field trials.

- Information and Communication Technologies in **DR**. With the vision of **HACS**, the current **ICT** for **DR** may eventually evolve into **HACS** in the future. The thesis aims to help realise the vision by providing design implications for **ICT** systems with automated planning support.
- Multi-Agent System (**MAS**). The multi-agent simulation technologies underpin the technical possibility of intelligent planning support, providing the opportunity space of human agent collective planning.

There are various **ICT** and **MAS** technologies designed to support disaster management activities in the different stages of the crisis circle including preparedness, response and recovery [Wattegama, 2012]. This thesis is going to limit the scope to operations of rescue and evacuation in the immediate aftermath of a disaster impact, which typically require high levels of team coordination and real-time task planning and execution.

The thesis also focuses on the issues related to a human team interacting with the automated planning support from a **HCI/CSCW** perspective. Although this work involves planning support agents based on multi-agent coordination algorithms, the effectiveness, performance and other technical issues of particular coordination algorithms are not the primary concern of this work.

As part of the ORCHID project, the AtomicOrchid (Chapter 4) serious game platform was developed as a research probe to trial human agent collective planning in the domain of Disaster Response. The AtomicOrchid platform consists of two major components (Chapter 4 for details): a game engine, and an embedded task planning agent. The

core game engine was developed, deployed and maintained by the author, whereas the task planning agent was developed by ORCHID research partners - Feng Wu and Sarvapali Ramchurn. Both Feng and Ramchun have expertise and research interest in the performance of task planning algorithm, while the author's research interest is the interaction between human and the intelligent task planning agent.

1.4 RESEARCH QUESTIONS

With the aim of understanding interactional issues surrounding automated planning support system, this work seeks to answer the following two research questions through field trial of AtomicOrchid:

- A What interactional issues will emerge if we try to automate the planning process in a Disaster Response team? In particular, the issues originated from the socio-technical gap often found in CSCW systems (Section 4.1), which can range from social, organisational to other interface design issues. This work aims at conducting an exploration of these issues in the interaction design space of automated planning support.
- B How can we design interaction to support human system collaboration in task planning? Following the first question, the emerging issues will need to be handled with appropriate interaction design. This work seeks to produce interaction design implications from field observation and interaction analysis, both of which are grounded in literature (Chapter 5).

Both questions will be answered with respect to the 2 different interaction patterns, namely human “on-the-loop” and human “in-the-loop”, which will be defined fully in section 4.4.

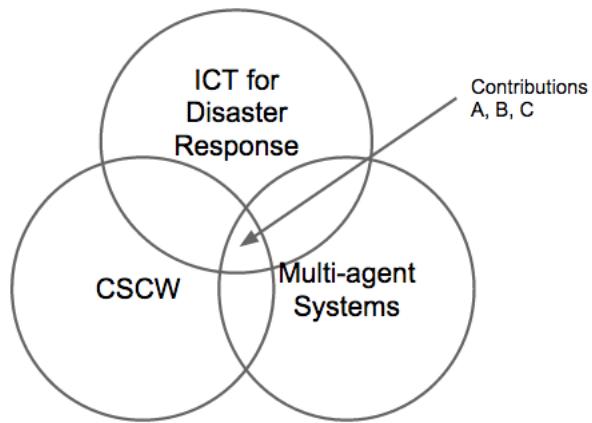


Figure 1: Contributions

1.5 CONTRIBUTIONS

This thesis contributes to the knowledge in the following areas (Figure 1):

- A The system prototypes.** Real-world interactive prototypes (i.e. AtomicOrchid) for investigating human-system interaction in a Disaster Response setting. An iterative prototyping process is conducted throughout the three AtomicOrchid studies, which results in interactive prototypes of planning support systems.
- B The interactional issues.** The field observation of serious game trials generate thick descriptions of human system interaction in the collaborative work settings of DR. The key observations reveal interactional issues surrounding automated planning support.
- C Design implications.** For each study, field observations are further analysed to generate design implications which contribute to future deployment of automated planning support system in the complex collaborative work setting of DR.

1.6 PUBLICATIONS OF THIS THESIS

Parts of the contents of this thesis have been accepted by peer-review for publication in journal and conference proceedings in the field of **HCI** and multi-agent system. The core contributions include:

1. The chapters [4](#), [5](#) present the approach and methodology employed to study interactional issues of planning support system. Some of the ideas in these chapters expands on the contents in:
Fischer, Joel E., Wenchao Jiang, and Stuart Moran. AtomicOrchid: a mixed reality game to investigate coordination in Disaster Response." In Entertainment Computing-ICEC 2012, pp. 572-577. Springer Berlin Heidelberg, 2012.
2. The exploration of requirements for building a coordination support system in chapter [6](#) has been published in:
Fischer, J.E., Jiang, W., Kerne, A., Greenhalgh, C., Ramchurn, S.D., Reece, S., Pantidi, N. and Rodden, T. (2014). Supporting Team Coordination on the Ground: Requirements from a Mixed Reality Game. To appear in: Proc. 11th Int. Conference on the Design of Cooperative Systems (COOP 14). Springer.
3. The exploration of interactional issues related to human on-the-loop pattern reported in chapter [7](#) has been published in:
Jiang, W., Fischer, J.E., Greenhalgh, C., Ramchurn, S.D., Wu, F., Jennings, N.R. and Rodden, T. (2014). Social Implications of Agent-based Planning Support for Human Teams. In: Proc. of the 2014 Int. Conference on Collaboration Technologies and Systems (CTS 14). IEEE.

Other contributions include:

1. Some results from studies in chapter [6](#) and [7](#) also appears in the journal article:
Ramchurn, S. D., Wu, F., Fischer, J. E., Reece, S., Jiang, W., and Roberts, S. J., et al. (2015). Human-agent collaboration for Disaster Response.

Journal of Autonomous Agents and Multi-Agent Systems.

2. The game probe ‘AtomicOrchid’ built in this PhD work is a central component of HAC-ER (Human Agent Collectives for Emergency Response) system developed as main demonstrator of ORCHID project (orchid.ac.uk). The demonstrator is presented in the paper:

Ramchurn, S. D., Simpson, E., Fischer, J. E., Huynh, D. T., Ikuno, Y., Reece, S., and **Jiang, W.** et al. (2015). HAC-ER: A Disaster Response system based on human-agent collectives. In AAMAS-15 : 14th Int. Conf. on Autonomous Agents and Multi-Agent Systems. [Best Paper Award].

3. The planner agent integrated in ‘AtomicOrchid’ is based on a novel multi-agent coordination algorithm. Evaluation of the algorithm is partially facilitated by the ‘AtomicOrchid’ trials:

Wu, F., Ramchurn, S. D., **Jiang, W.**, Fischer, J. E., Rodden, T., and Jennings, N. R. (2015). Agile Planning for Real-World Disaster Response. In International Joint Conference on Artificial Intelligence.

1.7 STRUCTURE OF THE THESIS

This thesis is structured as four parts. Part I surveys the relevant background literature. Chapter 2 will firstly give an overview of task planning activities together with command and control structures of DR teams, followed by a review of empirical studies related to command and control work settings. Chapter 2 then examines the state-of-the-art technology practices from 3 perspectives including planning support systems, ICT support, and application of Artificial Intelligence in DR. In Chapter 3, the relationship between technological support and human operators will also be examined by reviewing relevant literatures of CSCW systems, Automation design and Human agent interactions. The rest of the chapter give an overview of serious mixed reality games which underpins the research approach of this PhD work.

Part II develops the approach and methodology employed to study interactional issues with planning support systems in two chapters. Chapter 4 develops the framework of interaction patterns under which interactional issues can be explored. The rest of this chapter introduces serious mixed reality game as an approach to study socio-technical systems, followed by detailed description of the game used as testbed for this study, AtomicOrchid. Chapter 5 describes the methodology used to study the interactional issues. In particular, this chapter describes ethnographic observation and interaction analysis, which is supplemented by interviews.

Part III presents observational studies of this thesis. Chapter 6 reports the first observational study with AtomicOrchid. This version of AtomicOrchid does not have a planning support agent included. The study establishes baseline human performance of task planning and derives general requirements for communication support. Chapter 7 gives an account of the second observational study of AtomicOrchid. In this study, a planning agent was built into the game with the human on-the-loop interaction pattern. Chapter 8 reports the third field study of AtomicOrchid with the human in-the-loop pattern.

Part IV (chapter 9) concludes this thesis with a summary of discoveries, contributions, limitations and future work.

Part I

BACKGROUND

TASK PLANNING IN DISASTER RESPONSE

The overarching background of this thesis is disaster management, of which Disaster Response (**DR**) is one particular period immediately after the disaster impact. This chapter reviews the technological practices and relevant domain knowledge. First, section 2.1 reviews the relevant literature to develop an understanding of planning practices in **DR** operations. Section 2.2 reviews the relevant empirical studies to provide insight into characteristics of Command and Control (**C₂**) work settings. Section 2.3 gives an overview of technical practices in **DR** in the context of three detailed application areas, which includes Geospatial Information System (**GIS**) based planning support, Information and Communication Technologies (**ICT**) systems and Artificial Intelligence (**AI**)-based technologies.

2.1 TASK PLANNING IN DISASTER RESPONSE

To scope the task planning activities in **DR**, this section begins with a definition of Disaster Response, followed by a brief overview of command and control structures of **DR** organisations. The last section (2.1.3) examines the main characteristics of task planning in large scale Disaster Response.

2.1.1 *Defining Disaster Response*

Different countries and agencies may apply different rules and standards for defining phases of disaster management, but most of them agree that disaster management is carried out in a cycle. Figure 2 illustrates the a model of the disaster management cycle described by

[Wattegama, 2012]:



Figure 2: Disaster Management Circle. Credit Wattegama [2012]

1. Mitigation: any activity that reduces either the chance of a hazard taking place or a hazard turning into disaster.
2. Risk reduction: anticipatory measures and actions that seek to avoid future risks as a result of a disaster.
3. Prevention: avoiding a disaster even at the eleventh hour.
4. Preparedness: plans or preparations made to save lives or property, and help the response and rescue service operations. This phase covers implementation/operation, early warning systems and capacity building so the population will react appropriately when an early warning is issued.
5. Response: includes actions taken to save lives and prevent property damage, and to preserve the environment during emergencies or disasters. The response phase is the implementation of action plans.

6. Recovery: includes actions that assist a community to return to a sense of normalcy after a disaster.

DR operations refer to the actions taken during or in the immediate aftermath of the disaster. In this period, a significant number of individuals may be trapped and injured. A great amount of structural damage needs to be dealt with. Medicine, food and shelters are in high demand. This period calls for prompt action within an exceptionally short period of time [Wattegama, 2012]. Responder teams may find themselves with limited resources and need to make plans to utilise the resources in a timely and satisfactory manner [Chen et al., 2005, 2008].

2.1.2 DR command structure

Emergency response agencies typically employ a hierarchical command structure [Ramchurn et al., 2015a]. One widely used command and control structure is the Gold, Silver, Bronze model. In this model, decision making is divided into strategic, tactical, and operational levels. The teams responsible for each are referred to as “Gold, Silver, and Bronze” respectively. The decisions on main objectives of the response effort are made at the strategic (Gold) level. At the tactical level, the Silver command team decides on the allocation of resources and tasks to be carried out based on the specified objectives. At the operational level, Bronze First Responders (FR), on the ground, carry out those tasks. Information gathered on the ground is also passed back up from Bronze, through Silver, to Gold [Ramchurn et al., 2015a]. Some literatures [Chen et al., 2005, 2008] also generalizes the command and control structures as a generic two level model. The key characteristic of the two-level model is division between a remote coordination center and on site teams. On-site responders react to the immediate scene without a global picture, while the coordination center deals with strategic issues and works with a global picture,

leveraging external resources to help the on-site response.

2.1.3 *Task planning in large scale disasters*

One important characteristic of large-scale disasters is the presence of multiple spatially distributed incidents [Chen et al., 2005]. To gain insight into the problem of task and resource allocation in large scale disasters, we will firstly examine how a single incident is dealt with. The procedures for dealing with a single emergency incident have been documented by a number of field studies [Comfort, 2004; Dawes et al., 2004; Petrescu-prahova and Butts, 2005]. In Toups et al. [2011]’s study, fire emergency response to small-scale structural fires is depicted as follows:

Fire emergency response is undertaken by small teams distributed throughout the incident, coordinated by an incident commander (IC). Multiple response teams, or companies, are dispatched to any incident and cooperate around the fireground. A company officer leads each team, which consists of firefighters and/or engineers. Normally, each company is associated with a firefighting vehicle; an apparatus, such as an ambulance, engine, or ladder truck.

From the depiction of single incident emergency response, we can see that a combination of different resources (e.g. ambulance, fire engine, ladder truck) and skills (e.g. structural engineers, firefighters and medics) are deployed to the location of the incident. To deal with multiple incidents, the Disaster Response team has to coordinate spatially distributed resources and personnel to carry out operations (e.g. search, rescue and evacuation) [Chen et al., 2005]. That is, resources and responders need to be divided and combined into teams and deployed to handle distributed incidents. Depending on the number of incidents, response personnel may need to dispatch, deploy and redeploy limited resources. One major concern for task planning in

Disaster Response is how to efficiently allocate limited resources to multiple incidents with temporal and spatial constraints [Bradshaw et al., 2011].

Also, the task environment of DR is characterised by various uncertainties including, but not limited to hazard uncertainties, task-flow uncertainties, environmental and informational uncertainties [Chen et al., 2008]. Sudden and unexpected events may occur as the disaster situation unfolds. Therefore, fixed plans of actions for responders are unlikely to work. The uncertainties may need to be handled by improvisation, prioritisation, and dynamic sourcing of capabilities [Faraj and Xiao, 2006], which means dynamic changes of plans are necessary to deal with uncertainties in a dynamic task environment.

In summary, responders in DR need to carry out a set of interdependent activities under time pressure and spatial constraints. Both their resources (personnel and physical assets) and capacity for problem solving required for planning may be stretched in a large-scale, multi-incident disaster. To alleviate these problems, technological support in DR has long been studied by computer scientists.

2.2 COMMAND AND CONTROL ENVIRONMENT

Command and Control (C₂) environments have been a long-standing research topic in the field of Human Computer Interaction (HCI). Research on C₂ settings focuses on how operators manage and control safety-critical systems [Fischer et al., 2015]. Studies of C₂ work settings are particularly relevant to this thesis because most emergency response operations are carried out in a C₂ environment (section 2.1.2).

In the field of CSCW, ethnographic studies have been carried out to explicate work practices and organisational conduct in C₂ environments, with the aim of informing the design of technological support systems. These studies are conducted in a range of application areas

such as air traffic control [Richard et al., 1989], emergency response [Fischer et al., 2015], and military operations [Tolcher, 2005], with the aim of providing design implications for technological support. For example, Heath and Luff [1992] conducted a detailed examination of social organisation of collaborative work within a line control room of the London Underground. The analysis of Heath and Luff [1992] particularly focuses on interaction between different personnel as they coordinate a range of tasks and utilise various tools. The study revealed the methodical ways in which the participants surreptitiously monitor each other's conduct, and at the same time, render their actions visible and accountable to the other team members. Therefore, it is critical for computational systems to facilitate individuals are able to mutually monitor their co-participants [Heath and Luff, 1992].

Similar C₂ setting can also be found in the orchestration of Mixed Reality Game (MRG) experiences in which online players in the control room and field players 'on the ground' interact through pervasive technologies such as smart phones, GPS and other mobile sensors (Section 3.3). The ethnographic studies of MRGs of Benford et al. [2006]; Crabtree et al. [2004]; Koleva et al. [2001] focus on the distributed nature of C₂ environment. The observations in the studies reveal methodological ways in which players collectively battle the uncertainties and interruptions of technologies through distributed orchestration processes. The results of the studies lead to a set of design implications for handling uncertainties and interruptions in distributed C₂ environments.

Apart from ethnographic studies, the impact of technology support in the control room is also assessed by human factors researchers through quantitative evaluation of operators' performance [Grootjen et al., 2007], [Sharples et al., 2011]. For example, the study by Sharples et al. [2011] is conducted to assess the impact of recent automation in rail way signalling operations. The studies evaluated changes of operators' performance through statistical test and quantitative comparison of observational data. The results indicate that automation is

effective at reducing high level of interaction and workload. However, operators do seem to struggle to maintain awareness of system state and they are also unable to monitor the way in which automated support makes decisions in real-time.

While both qualitative and quantitative studies contribute to an understanding of human-system interactions in complex [C2](#) settings, this PhD work is primarily interested in understanding social organisations of work, and follow the tradition of empirical [CSCW](#) studies to unpack the social interactions in [C2](#) setting.

2.3 TECHNOLOGICAL PRACTICES IN DISASTER RESPONSE

To give an overview of current technological practices related to planning support, This section reviews three related research domains in the context of computational DR support, that is - [GIS](#)-based task planning systems for plan formulation and evaluation; Information and Communication Technologies ([ICT](#)) for information acquisition and management; and the Artificial Intelligence ([AI](#)) technologies for disaster simulation and plan optimisation.

This PhD work is primarily interested in real-time task planning systems that utilise [ICT](#) technologies as underlying infrastructures, and apply intelligent coordination algorithms for plan generation. This kind of system can be located in the overlapping area of the three research domains that are to be reviewed in this section (Figure [3](#)).

2.3.1 *GIS-based planning support systems*

A Planning Support System ([PSS](#)) can be defined as a suite of computational components that help planners explore and manage planning

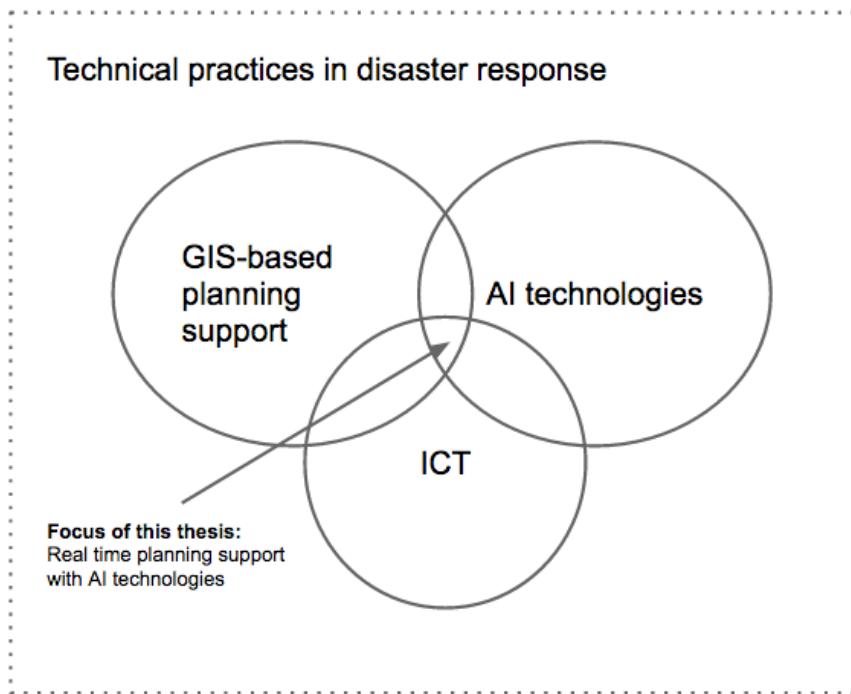


Figure 3: Technological support for Disaster Response

activities [Geertman and Stillwell, 2004]. Literature has documented a variety of planning support systems with a range of purposes such as land development [Pettit et al., 2003] and logistic scheduling [Miller, 1999]. In the context of Disaster Response, planning activities typically concern dispatch, routing and deployment of rescue resources (see section 2.1.3), which means GIS for spatial analysis may be a central component of a planning support suite. Therefore, this section will focus on reviewing GIS-based PSses.

While general-purpose GIS is only designed to handle geo-spatial data [Geertman and Stillwell, 2004], a PSS may have a wide variety of functionalities to support multiple aspects of the planning process, which may include, but is not limited to problem diagnosis, data collection, mining and extraction, data modelling, visualisation and display, scenario-building and projection, plan formulation and evaluation, and collaborative decision-making support [Geertman and Stillwell, 2004; Zerger and Smith, 2003]. A range of the GIS-enabled planning systems have been developed for various purposes such as vulnerability assessment, risk management and hazard mitigation [Cova, 1999; Schooley et al., 2010]. One example is the “Intergraph” devel-

oped by Victorian Emergency Services (Australia) [Intergraph Corporation, 2000] to support emergency dispatch services (e.g. ambulance, police and fire services). The system supports daily response activities by providing automated geo-referencing, routing, mapping, planning and analysis [Zerger and Smith, 2003]. Another example of GIS-based planning support is an evacuation planning tool for radiological disasters developed by [Eglese and Pidd, 1994]. The system combines simulation models and GIS software to model routes for evacuation in radiological incidents. It uses the spatial data structure and programmed simulation models to predict traffic flow through the road network under scenarios such as vehicle breakdowns and road closures. The system is designed for evaluating evacuation strategies prior to disaster. Real-time decision support is not possible because the simulation is computationally intensive.

'Real-time' task planning in DR is different from other long-term planning scenarios due to the time pressure and uncertainties involved. Responders in DR have to plan dynamically according to the disaster environment that is always quickly changing. The study of Zerger and Smith [2003] points to some technical and social impediments for applying GIS tools to support 'real-time' planning activities in DR. 1) Some spatial modelling and simulation technologies [Eglese and Pidd, 1994] can be very computational intensive which makes it unable to provide results in real-time. 2) Some of the planning systems adapted from GIS software, require a high level of training to operate. 3) There is a lack of social and organisational considerations in most of the existing GIS tools. For example, some PSS are based on a single computer terminal, which is unable to facilitate information sharing with multiple users in a control room and 'on the ground'.

The third point is aligned with the studies of C₂ task environments (in section 2.2). The C₂ environment is a collaborative setting which involves complicated social interaction and organisational conduct. Therefore, a PSS designed for a C₂ environment should consider supporting the social conduct of responder teams. Further, the compu-

tational support in a modern [C₂](#) environment (e.g. air traffic control [[Mercer and Homola, 2014](#)] and underground control [[Sharples et al., 2011](#)]) supports a range of activities such as information acquisition, analysis, decision selection and action implementation. Similar to other [C₂](#) environments, plan generation and action implementation in DR are typically fast paced and happening in parallel. Existing [PSS](#) such as those by [[Intergraph Corporation, 2000](#)], and [[Eglese and Pidd, 1994](#)] only provide isolated functionalities for supporting plan generation. To realize ‘real-time’ planning support, the author believes the [PSS](#) should be further extended or integrated with other system functionalities for action implementation, such as progress monitoring, conflict detection, and communication support.

2.3.2 *ICT support for Disaster Response*

Information and Communication Technologies ([ICT](#)) include communication infrastructures and software systems on top of these infrastructures. Communication infrastructures refer to communication channels such as radio, television, satellite, internet, text and voice communication over mobile phone. The basic communication infrastructures have long been utilised by responders to capture both soft data (generated by human) and hard data (from sensors) for their decision making [[Fischer et al., 2012](#)]. Apart from the infrastructures, [ICT](#) software systems are also playing increasingly important roles. Most modern DR practice may consist of both manual processes which directly relies on [ICT](#) infrastructures and (partially) automated process of data analysis and information management that are supported by [ICT](#) software. For example, the figure 4 illustrate a tsunami response system operated by the Asian Disaster Preparedness Center (ADPC) [[Wattegama, 2012](#)]. In this system, the technical components are comprised of a network of seismographic stations, sea-level gauges and deep-sea pressure sensors. A tsunami forecasting centre is equipped with seismic processing and modelling software. Tsunami warnings are disseminated through the communication links between national

centres and the people at risk, the links include, email, television, radio, cellphones and satellites. Another scenario for **ICT** support can be found in most modern Emergency Response Rooms (**ERR**). For instance, the **ERRs** in Amsterdam uses a **ICT** software called GMS (Integrated Emergency Response Room System) [Boersma et al., 2009]. GMS is used to connect the different information sources (e.g. a national digital radio network for public safety). However, the **ERR** in Amsterdam is not co-located, but divided by regions and functions (Fire, Police and Medical). As a result, the GMS used by **ERRs** are not connected and they are tailored differently to the various needs and requests of operators. As the Emergency Response Rooms (**ERR**s) systems are disconnected, information exchange between **ERR** is conducted manually through telephones [Boersma et al., 2009].

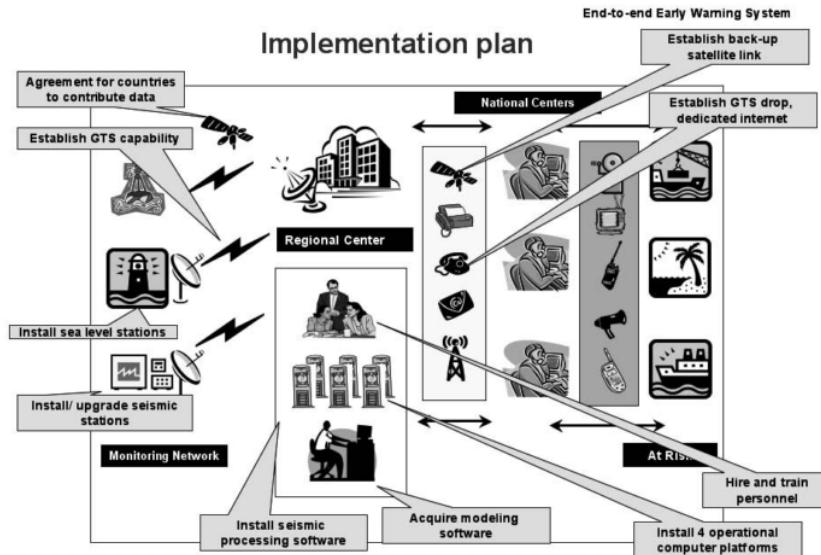


Figure 4: Tsunami response system, Credit Wattegama [2012]

With the proliferation of smart phones and ubiquitous computing technologies, the internet has become increasingly important in the Disaster Response domain. Web-based applications have been used in the Indian ocean tsunami for tracking missing people, coordinating donors, and recording locations of shelters [Wattegama, 2012]. The internet also expands the possibilities for public participation. For example, the use of microblogging tools such as 'twitter' in **DR** have become a popular research topic in the field of crisis informatics [Kogan

et al., 2015; Sarcevic et al., 2012; Starbird and Palen, 2010]. Microblogging tools enable the public to take not only a more active part in seeking information, but also in providing information to each other, as well as to formal response efforts. Crowdsourcing platforms such as Ushahidi [Morrow and Mock, 2011] are another example of ICT support for public participation in DR. The Ushahidi platform has been used in the Haiti earthquake as a volunteer effort to produce a crisis mashup. Information about the humanitarian crisis and the response that followed was aggregated in near real-time by volunteers from a variety of sources including: SMS, Web, Email, Radio, Phone, Twitter, Facebook, Television, List-servers, Live streams, Situation Reports [Morrow and Mock, 2011]. Similarly, crowdsourced map systems such as OpenStreetMap [Palen et al., 2015] have been applied in disasters (Haiti earthquake, Typhoon Yolanda) for volunteers to create updated crisis maps. Although public participation has played an important role in DR, its integration with formal response efforts is still a challenge [Palen and Liu, 2007], and a lot of research effort seeks to address the social, organisational and technical challenges of the integration [Dashti et al., 2014; Sutton et al., 2008].

In summary, ICT infrastructure and software are commonly used in Disaster Response. Two case studies (tsunami response system and emergency response room support) are used to illustrate the ICT supported DR workflows, which contains a mix of manual and automated processes. Public participation is another trend of ICT system support arising from increased global access to internet. The trend of public participation can be seen as an application of connected computer systems in DR. However, issues of supporting public participation are not the focus of this thesis, instead, this work is particularly interested in automated task planning support, which heavily relies on ICT infrastructure for data acquisition and plan implementation. Functionalities in ICT software also have many overlaps with that of task planning systems, in the sense that they all support responders to acquire, process and manage information for decision making in DR.

2.3.3 Application of AI technologies

In the field of [AI](#), machine learning, optimisation and agent-based simulation algorithms have increased in availability for Disaster Response. Disaster simulation technologies [[Okaya et al., 2014](#); [Scerri et al., 2005a](#)] have long been applied in various types of disasters such as fire [[Tang and Ren, 2012](#)], hurricane [[Vickery et al., 2009](#)] and earthquake [[Sobhaninejad et al., 2011](#)] to understand the process and impact of a disaster. State of the art simulation technologies are able to provide real-time simulations with high data resolution. For example, the simulation platform devised by [[Sobhaninejad et al., 2011](#)] is able to provide real-time simulation of ground motion and structural damage after an earthquake, in an urban area with 10^4 to 10^6 structures.

Evacuation simulation is also extensively studied in the literature. For example, [Pillac et al. \[2015\]](#) developed a simulation algorithm for large scale flood evacuation. The algorithm consider the dynamics of the flood and the state of the transportation network over time. The experiment showed this algorithm was able to provide plans to evacuate about 70,000 people in an impact region in real-time. Another example is simulation of indoor fire evacuation devised by [Tang and Ren \[2012\]](#). The algorithm combines modelling of building geometry, human behaviour, and fire field to provide intelligent decision support. While results of evacuation simulations are promising in lab experiments, some studies have pointed out that the simulation technologies over-simplified human behaviour [[Deconinck, 2013](#)]. For instance, in some countries, most people disregard evacuation orders. The complexity of human behaviour arises from social, cultural, political and psychological factors, which makes behaviour modelling a challenging topic for the [AI](#) community [[Provitolo and Dubos-paillard, 2011](#)].

Agent-based optimisation algorithms for multi-agent team coordination is also an active research area in the [AI](#) community. The

coordination challenges in time critical task environments inspired multi-agent researchers to investigate real-time algorithms that provide optimal or near optimal task plans that can be used to guide distributed human or robotic teams [Kitano, 2000]. Various algorithms have been designed to meet different requirements of team coordination. For example, Lagoudakis and Markakis [2005] provide a number of auction-based algorithms that allocate agents to tasks considering the best routes for each agents in the team. On the other hand, Scerri et al. [2005b] devised algorithms to allocate tasks based on token passing. Their algorithms focus on matchings of task and team capabilities, which ensures that the right agents are routed to the right tasks based on capability thresholds. Other algorithms [Ramchurn et al., 2010; Koes et al., 2005] also consider multiple criteria including task workload, deadline and team coalition effects. From a HCI perspective, the capabilities of these algorithms are also limited by a number of factors. First, the concerns about over-simplified human behaviour modelling can also be applied [Drury et al., 2009]. Second, because the algorithms are designed to deal with specified requirements (e.g. routing, task/skill matching) with simplified models of environments, they may be insufficient to deal with contingencies that arise from messy real disaster environments [Armenakis, 2012].

It is observed that most intelligent algorithms are only studied and evaluated in the research lab. Integration of the algorithms and actual DR practices is still a multi-disciplinary research challenge. With the increase in networked computers, sensors and the amount of data generated from different sources in real-time [Ramchurn et al., 2015a], there are increasing demands for intelligent computational support of data processing and task planning. As a case study, researchers from the ORCHID project have developed a prototype system, called HAC-ER (Human Agent Collectives for Emergence Response) [Jennings et al., 2014; Ramchurn et al., 2015a,b], which demonstrates how AI algorithms may transform the landscape of real-time task planning in DR.

The Human Agent Collectives for Emergency Response ([HAC-ER](#)) system consists of a set of connected components for real-time task planning support, each of which are powered by multiple machine learning and agent-based algorithms. First, a component called Crowdscanner is used to deal with vast quantities of unstructured data produced very rapidly on the internet as the disaster unfolds, such as text messages or photographs from web-based platforms such as Twitter and Ushahidi [[Morrow and Mock, 2011](#)]. The approach is to use a machine learning algorithm (for details, see IBCC [[Simpson et al., 2013](#)]) to fuse heterogeneous reports from both unreliable and trusted sources into a common picture of the disaster, or a heatmap of incidents. Second, multiple Unmanned Aerial Vehicles ([UAV](#)s) are deployed as mobile sensors to search or further inspect incidents reported by crowdscanner. The control of multiple [UAV](#)s is assisted by a multi-agent coordination algorithm (Max-sum [[Ramchurn et al., 2010](#)]). The algorithm is capable of quickly optimising the task allocation for [UAV](#)s to visit points of interest or conduct search in an area. Finally, responders and assets on the ground will need to be dispatched and deployed to deal with distributed incidents. So, another [HAC-ER](#) component¹ is designed to assist human operators in the control room to conduct real-time planning. The component is powered by a coordination algorithm based on MMDP modelling techniques [[Wu et al., 2015](#)], which takes into account the priorities of incidents, and locations of responders teams and incidents. The algorithm can produce computationally optimised task allocations for responder teams to attend as many incidents as possible with a time constraint. Further, algorithms and interfaces of all components in [HAC-ER](#) are designed in a way that accept input from human operators. The [HAC-ER](#) prototype depicts a possible future in which human operators and intelligent components collaboratively conduct task planning, and demonstrates the potential for applying [AI](#) technologies in [DR](#) task planning.

¹ based on the AtomicOrchid platform developed in this PhD work

2.4 SUMMARY

Relevant literature has been reviewed in this chapter. The studies of Disaster Response operations showed that the task planning activity is challenging due to time, spatial constraints and uncertainties in a dynamically changing disaster environment. Further, the DR planning process is typically embedded in a Command and Control (C₂) task environment. The literature has revealed C₂ settings are characterised by complex social interaction and organisational conduct. Therefore, technology support does not only need to provide solutions to complex coordination problems, but also needs to support social aspects of planning generation in complex C₂ settings.

There is various existing technology support designed for DR. First, GIS-based planning support combines GIS functionalities with environmental modelling techniques to provide decision support for various purposes such as floor and indoor fire evacuation. However impediments for applying GIS-based PSSEs are also identified in the literature. Second, ICT support systems historically played an important role in DR. ICT infrastructures and software are becoming increasingly connected at global scale, which changes the landscape of DR by expanding the possibilities for public participation in DR. Further, the AI community has also devised various machine learning and optimization technologies for DR operations. In particular, a number of multi-agent coordination algorithms have been devised to provide solutions for task allocation in DR. In the overlapping area of the three domains, there is an opportunity space for real-time task planning support in DR. A planning support system can utilise ICTs as its underlying infrastructures and apply multi-agent coordination algorithms for plan generation. However, the research for building such a system is still rare, leaving a gap in the literature that this thesis seeks to narrow.

INTERACTION BETWEEN HUMAN AND PLANNING SUPPORT SYSTEM

Applying task planning to support complex Disaster Response operations may not be a straightforward process. Some [HCI](#) literature [[Ackerman, 2000](#); [Bowers et al., 1994](#); [Niazhani et al., 2009](#)] has shown that introducing technological systems to support organisational work can be extremely difficult. Use of technologies may have unexpected negative impacts on human team performance. This thesis adopts a socio-technical view towards understanding the impact of planning support technologies on human team performance. In particular, in planning support system, the socio-technical perspective is concerned with issues that emerge from the gap between human and system problem solving models, namely the divide between the situated actions of human and the ‘scientific’ planning model held by the system [[Suchman, 1987](#)]. Meanwhile, two other [HCI](#) research areas are also concerned with issues of human system interaction from different perspectives: research on automation design concerns what and how to automate a working process by using computational systems; research on human agent interaction focuses on building software agents capable of teamworking with human operators. This section reviews these three perspectives to give a conceptual background to the design challenges that we may encounter when developing [DR](#) task planning support.

Further, studying human system interactions in a real world disaster could be very challenging because disaster conditions can not be reproduced easily. Therefore, researchers have long been using games as an approach for studying the impact of technology support. Section [3.3](#) reviews the strengths and weakness of using games as an

approach for studying the disaster work setting.

3.1 A SOCIO-TECHNICAL PERSPECTIVE ON TECHNOLOGY SUPPORT

A so-called socio-technical gap is often observed in research into Computer Supported Collaborative Work ([CSCW](#)) systems. The research field of [CSCW](#) addresses how collaborative activities can be supported by means of computer systems [[Carstensen and Schmidt, 1999](#)]. [CSCW](#) goes beyond building technology itself and investigates how people work within groups and organizations and the impacts of technology on those processes. The term socio-technical systems was originally coined by Emery and Trist [[Ropohl, 1999](#)] to describe systems that involve complex interactions between humans, machines and the environmental aspects of the system. The implication of this definition is that both social and technical factors e.g. people, machines and context need to be considered when developing such systems. Introducing a technology support system into an organization requires the technical and social aspects to be integrated, which can be a major challenge of building [CSCW](#) systems [[Ackerman, 2000](#)].

It is argued [[Ackerman, 2000](#)] that human activities are highly flexible, and contextualized and that computational processes and entities need to be similarly flexible. The social-technical gap is the divide between what we know we should support socially and what we can support technically [[Ackerman, 2000](#)], and the gap may result in serious negative impact on human activities [[Bowers et al., 1994; Abbott and Sarin, 1994](#)]. Therefore, it is vital to study technology in use to understand potential tensions between technical mechanisms and social life [[Bowers et al., 1994](#)]. In the context of Disaster Response, we believe the same is true for the application of automated planning support. Empirical works [[Kopena et al., 2008; Fischer et al., 2015; Zerger and Smith, 2003](#)] have revealed the complexity of social and technological processes in Disaster Response operations. Responders

with different roles have to engage in various interdependent activities that are distributed in both time and space (Section 2.1). In order to build automated systems that support human team coordination, it is vital to explore the social-technical gap and its impact.

3.1.1 The socio-technical gap for planning support system

In the context of planning support, the socio-technical gap can be specifically caused by the different ways in which human and systems plan and act. The view of purposeful action is adopted by many researchers and designers of intelligent systems, serving as the model of plan and action that is embedded in many problem solving systems [Allen, 1984]. The view of purposeful action is also embraced by behavioural science as the basis for the traditional view of rational human action. In this view, a plan is a prerequisite of and prescribes to action [Suchman, 1987]. However, the view of purposeful action is challenged by recent trends in sociology, arguing that the prescriptive significance of plans and intentions for actions is vague. Rather than determining the actions, the plans only serve as resources for human's actual deliberations on their actions [Suchman, 1987].

3.1.1.1 Planning model embedded in problem solving systems

The view of purposeful action aligns with the planning model from cognitive science [Suchman, 1987], which is embedded in many of problem solving systems. The model treats a plan as a sequence of actions to achieve an end goal. The purpose of problem solving is to find a way from initial state to an end goal state. In many problem solving systems, actions can be described by prerequisites, effects, and decomposition [Allen, 1984]. The prerequisite are the conditions in the initial state that triggers the action. The effects defines how the action can alter the initial state and decomposition of actions (into subactions) defines how the actions can be executed. Some problem solving systems are only designed for plan generation. When a

plan is constructed, the problem solvers finishes their work, assuming nothing will go wrong. However, contingent situations (or unanticipated conditions) are likely to emerge, requiring replanning [Suchman, 1987]. Therefore, modern autonomous systems are typically designed to conduct planning and monitor execution, which enables them to respond to contingent situations.

Recent development in AI (e.g. multi-agent systems) require the planning model to be extended from an individual agent to a multi-agent situation. To achieve concerted actions in a social situation with multiple agents, the planning model seeks to attach the knowledge of other agent's plans and actions as a type of environmental condition for planning. In this way, the basic view of a single, goal directed agent (e.g. BDI model of agents [Georgeff et al., 1998]) remains intact. To obtain the knowledge of other agent's plans, agents are required to conduct plan recognition in addition to planning and execution monitoring. The basic assumption of plan recognition is that agents as observers take actions of each other as evidence to form hypotheses of plans that explain the intentions and actions of each other.

The planning model is problematic when it is applied as a model of a human's "psychological process" of action. Studies have pointed out on many occasions that human actions do not appear to follow logical steps or rational plans [Suchman, 1987]. Researchers of cognitive science have argued that the issue is due to actions under vague intents and incomplete plans. As a result, the human planning activity is often considered an inferior version of scientific planning that need to be supported. However, many sociologists challenge the view by refuting the significance of plans in determining actions [Suchman, 1987]. The key argument points out that human action is essentially situated, not planned. This argument will be detailed in the next section.

3.1.1.2 *Situated actions of humans*

In contrast to problem solving systems, recent efforts in sociology argue that humans perform in a situated manner. The argument adopts the view of situated action as opposed to the view of purposeful action (at least in the past) held by most AI researchers [Suchman, 1987]. The starting point of the argument is that the plan does not prescribe the actions. The plans are part of a large context of circumstance of ongoing actions, serving as a resource for deliberation on actions [Suchman, 1987].

In the view of situated action, the plans are only representations of our actions in the form of imagined projection and recollected reconstructions, which do not have causal relation with the actions we take. Psychologists [Mead and Mind, 1934] suggested that, although a great deal of deliberation and discussion can be taken into the plan, we may quickly stop following our plans after a few actual actions. Mead and Mind [1934] also pointed out that our post hoc analysis of situated actions often make them appear to have followed rational plans. Situated actions are in fact highly dependent on specific occasions circumstances and social situations. In contrast to plan recognition, mutual intelligibility in a social situation is not achieved by using logical deduction and formal algorithms based on behavioural observation. Mutual intelligibility is maintained through every instance of interaction, rather than fixed social rules and constraints [Suchman, 1987].

3.1.2 *Summary*

The great divide between the planning model embedded in many problem solving systems, and situated actions performed by humans signifies the issues of human system interaction in complex socio-technical systems such as a Disaster Response management (Table 1). In contrast to fully autonomous systems, computational support sys-

tems have to interact with human to achieve mutual intelligibility and concerted actions. It is important to recognise that the common sense planning activities of humans are not an inadequate version of the scientific planning model that need to be improved or supported by systems. Instead, a human plans and acts in a way that is fundamentally different from machine problem solving. Therefore, a big challenge in designing technology support is how to bridge the socio-technical gap between situated action and the scientific planning model.

| | Plan | Interaction |
|--------------------|--|---|
| Purposeful actions | Prerequisite to and prescribes to actions. | Shared understanding is achieved through plan recognition. Plan recognition is achieved by logical formulae based on observations of other's behaviour |
| Situated actions | Do not prescribe actions. Resource of deliberation and representation of actions. | Mutual intelligibility is achieved through interactions embedded in the social situations. |

Table 1: Planning model and situated action

3.2 OTHER PERSPECTIVES ON HUMAN-SYSTEM INTERACTION

Two [HCI](#) research areas, namely automation design and human agent interaction, are very relevant to the development of autonomous systems that collaborate with humans. This section reviews literatures in these two areas, with the aim to further develop an understanding of issues and challenges of human-system interaction.

3.2.1 Automation and its impact on human performance

The aim of an automated support system is to replace the tasks originally performed by human with a machine [Bradshaw et al., 2011]. It

can be defined as the execution by machine, usually computer, of a function previously performed by a human [Parasuraman and Riley, 1997]. The tasks that can be automated used to be limited by technical capabilities, but this is no longer the case. With the quick growth of machine's speed and intelligence, the tasks that can be automated are rapidly increasing, including complex cognitive activities such as information analysis, planning and decision making [Parasuraman et al., 2000]. The boundary between human and machine capabilities has blurred. Automation designers have to make hard choices about what to automate and to what extent.

One traditional approach to automation design is to simply automate all system functions that can be automated easily in a cost-effective way, leaving all the remaining tasks to human operators. The main considerations in this approach are technical capability and cost. The assumption of this approach is that the automation of sub systems' functions can lead to optimisation of the whole system with no detrimental impact resulting from the automation. However this is not always the case, as a large body of empirical work in automation design [Manzey and Onnasch, 2012; Parasuraman et al., 2000] has shown. The benefits of automation may not always be realized but can be offset by some unwanted performance consequences resulting from an inappropriate use of such systems. These performance consequences include overreliance on automation, loss of situation awareness, and possible loss of skills needed to perform the automated functions manually in case of automation failure [Kaber and Endsley, 1995].

The recognition of negative automation impact leads to the challenge of interaction design for automated support. Some researchers have suggested a division of labour between human and automation according to their strength and weakness. As in Fitts [1951] list, a set of strengths and weaknesses of humans and machines is identified. However, some studies have pointed out the division of labour would not be as simple as a labour division according to strength and

weakness [Bradshaw et al., 2011]. By delegating task to a machine, the nature of the human tasks can be changed as well. Researchers point out that automation does not simply supplant human activity but rather changes it, often in a way that is unanticipated by the system designer [Bradshaw et al., 2011]. Alternative approaches such as the un-Fitts list [Hoffman et al., 2002] suggest that automation should be aimed to leverage and extend human capability, rather than replacing it. While both Fitts and un-fitts lists are useful as high-level guidelines for interaction design, the author believes it is important for system designers to study the current work settings and technologies in use, in order to gain insight into how the division of labour is naturally achieved [Crabtree et al., 2012], so that the detailed design implications can be drawn from insights into naturally occurring division of labour.

3.2.1.1 *Levels of automation*

Various framework models have been proposed to guide research on automation-induced performance consequences. The framework models allow for a standardized characterization of automated systems with regards to how functions are distributed between humans and machine. One commonly recognised model is known as the Level of Automation (LoA). Most tasks can be fully or partially automated, which implies that automation is not all or none, but can vary across a continuum of automation level [Wickens et al., 2010]. At the lowest level, all system functions are performed manually by human operators. At the highest level, the system is fully automated, taking over all system functions. In between this two extremes, there are 10 different levels of automation proposed by Wickens et al. [2010].

The LoA model can be combined with more detailed classification of automation types according to a cognitive information processing model [Parasuraman et al., 2000; Manzey and Onnasch, 2012]. In this classification, the human information processing work is divided into four stages, which can be supported by automation individually. The

four successive stages are referred to as information acquisition, information analysis, decision selection, and action execution. Each of these identified stages can be automated with a certain degree of automation. With respect to human performance consequences, it is generally assumed that higher level of automation will benefit the system by reducing the workload of human operators. In contrast, it is also assumed that medium LoA can keep the human in-the-loop, which in turn, prevents what has been referred to as out-of-the-loop unfamiliarity, that is, a loss of situation awareness and a loss of manual skills [Kaber and Endsley, 1995; Parasuraman and Manzey, 2010].

While the model has been commonly adopted to study one-to-one operator-system interaction (e.g. tele-operation [Schwarz et al., 2014]), the author believes that the model may be insufficient to capture the complexity of the interactions in the context of technology support for organisational work. For example, automation of a single function may have implications for the roles and responsibility of many participants in the system and fundamentally change social conduct within an organisation. However, despite its drawbacks, some concepts and vocabularies in the LoA model are adopted by this research to characterise the possible types and levels of the automated planning support (Section 4.4).

3.2.2 Human-Agent interaction

Human Agent Interaction ([HAI](#)) is a research area that has strong overlap with automation design, in that they are both concerned with the impact of automation (i.e. computational systems) on human performance. While automation design is trying to answer the question of what and to what extent the system should be automated, the study of human agent interaction concerns the issues of interaction design related to the development of human agent systems, with the aim to develop agents which are good at teamworking with human opera-

tors[Bradshaw et al., 2011; Sukthankar et al., 2011].

The term “software agent” is commonly defined as the software that can operate independently without constant human supervision [Vlassis, 2007]. Agent researchers investigate infrastructure, language and communication to realize coordinated agent software systems [Nwana, 1996]. More recently, the use of the word software agent has become much more diversified, as agent researchers have made attempts to investigate a broader range of classes and types of agents such as interface agents, collaborative agents and information seeking agents [Nwana, 1996].

Some HAI researchers have adopted a view that humans and agents are equally important team players in a human agent system [Sukthankar et al., 2011]. This highlights the importance of mutual interaction between human and agent that can enhance the competencies of both human and system. From this perspective, the aim of automation is no longer to "replace" humans but to achieve a human system symbiosis through effective human agent coordination. As researchers realize that effective human-machine symbiosis requires sophisticated interactions design between human and agent, the “social” issues of automation are thought to be as important as the technical issues [Bradshaw et al., 2011]. To build socially acceptable agent systems, research communities have formed around the topics of 1) interface agents and assistants, 2) adjustable autonomy, and 3) mixed-initiative interactions. The following sections review the literature of those three HAI research topics, to overview the state of the art techniques for building teamwork agent.

3.2.2.1 *Interface agents and assistants*

An interface agent is the software that actively assists a user in operating an interactive interface [Lieberman and Selker, 2003]. The metaphor of personal assistant is used to design the interaction between the human and the agent. The software behaves like a personal assistant who is collaborating with the user in the same work environ-

ment [Lieberman, 1997]. The assistant learns the user's goal, interests, habits and preferences (as well as those of his or her community) by observing the user's actions and behaviours, so that it can effectively provide suggestions or directly take interface actions to support the human performing tasks [Maes, 1994].

User modelling is a typical approach for implementing interface agents. The user model is the program's expectation of what a user will do, what they need to be told and how they will respond [Lieberman and Selker, 2003]. The most sophisticated way to acquire user models are adaptive, which means they can create dynamic behaviour without being pre-programmed. For example, the user interface could learn common misspellings and typos and remind a user of how to correct them [Lieberman and Selker, 2003].

The user modelling approach has been adopted to build various agent systems. For example, Maxims [Metral and Maes, 1998] is an agent that assists the user to organise their emails. Maxims learns to prioritize, delete, forward, sort and archive mail messages on behalf of the user. The agent continuously learns the user as the user deals with emails. The user modelling technique is also applied to implement a meeting scheduling agent [Kozierok and Maes, 1993]. The agent assists a user with the scheduling of meetings (accept/reject, schedule, reschedule, negotiate meeting times, etc.) based on continuous learning of the user's preference. Other examples of applications ranges from news filtering, music/book recommendation [Lieberman and Selker, 2003] to smart home control [Costanza et al., 2014].

3.2.2.2 Adjustable autonomy

'Adjustable autonomy' is a concept that emerged from research on human agent teamwork. While in traditional agent-based systems, the division of labour between human and agent is pre-programmed (i.e. a fixed level of autonomy), the system with adjustable autonomy can have reasonable dynamism with a sufficiently fine-grained range of automation levels. People want to maintain the boundary of automa-

tion at a point that minimizes their need to attend to interaction with the agent while providing them with a sufficiently comfortable level of assurance that nothing will go wrong [Bradshaw et al., 2003]. The concept of ‘adjustable autonomy’ has been widely adapted to guide a wide range of areas such as tele-operation [Schwarz et al., 2014; Goodrich et al., 2001], smart home control [Costanza et al., 2014] and the domain of space exploration [Dorais et al., 1999].

3.2.2.3 *Mixed Initiative Systems*

“Mixed initiative” is another approach adopted by researchers to design human agent interaction. Mixed-initiative refers to a flexible interaction strategy, where each human or agent can contribute to the task that it does best [Allen et al., 1999]. In the most general cases, the human’s and agents’ roles are not determined in advance, but opportunistically negotiated between them as the problem is being solved. At any one time, one might have the initiative controlling the interaction while the other works to assist it, contributing to the interaction as required [Horvitz, 1999].

Practitioners have devised algorithms, interfaces and applications that facilitate “Mixed Initiative” planning and control [Ferguson et al., 1996; Burstein et al., 2003; Hardin and Goodrich, 2009; Zimmerman et al., 2007]. One example is the interactive optimisation algorithm designed by [Yang and Lee, 2012] to support environmental planning. Problem-solving in environmental planning typically involves multiple criteria, which can be qualitative or quantitative. The interactive algorithm allows a human decision maker to actively participate in the optimization process, and thereby provides the means to include qualitative expert knowledge within its solution space searching process.

3.2.3 *Summary*

Through reviewing [HCI](#) research from two perspectives (automation studies and human agent interaction), we developed an understanding of the interaction challenges raised by integrating technology with humans' working processes. First, automation designers realised that technology can bring unexpected human performance consequence. Automation designers devised the [LoA](#) model as a way to view alternative ways that we can automate a process. Although there are limitations, the model still provides useful guidelines/terminologies as a starting point for designing human system interactions. Second, research on human agent interaction attempts to realise human system symbiosis by developing software agent with social capabilities of teamworking with humans. A number of interaction design guidelines and metaphors have been raised in the [HAI](#) literature, such as "agent assistant", "adjustable autonomy" and "mixed initiative interaction". The perspective of [HAI](#) also inspires the interaction design of the automated planning support component in the game probe studied in this PhD work.

3.3 GAMES AS AN APPROACH TO STUDY HUMAN-SYSTEM INTERACTION

This PhD work adopts a serious mixed reality game approach to investigate the socio-technical issues surrounding agent planning support for Disaster Response operations. This section will firstly review the literature of serious games to give an overview of the history and applications of serious games. In section [3.3](#), Mixed Reality Game ([MRG](#))s will be introduced, followed by a discussion of the potential for using [MRG](#)s to support ethnographic study of ubiquitous systems, which underpins the rationale of the serious [MRG](#) approach applied in this PhD study.

3.3.1 *Serious Games*

There are various definitions for the term ‘serious game’ in the literature. One issue that most agree on is that the term is concerned with the use of games and gaming technology for the purposes other than mere entertainment, such as education, training, healthcare, and advertisement [Susi et al., 2007]. Although serious games usually have the look and feel of digital games, they are often simulations of real-world events or processes. By engaging the participants with simulated environments and systems, the serious games allow learners to experience situations that are impossible in the real world for reasons of safety, cost, time, etc. [Squire and Jenkins, 2003; Meesters and Van De Walle, 2013]. Recently, the serious games have increasingly been applied in the area of Disaster Response. In particular, ‘serious games’ are developed for training and simulation of terrorist attacks, disease outbreaks, biohazards, traffic control and fire fighting etc [Susi et al., 2007; Squire and Jenkins, 2003].

One good example of a serious game for DR is the “Biohazard” developed by MIT Comparative Media Studies [Squire and Jenkins, 2003]. The video game is designed to train emergency responders to deal with toxic spills in public locations. Emergency responders work in teams to organize the response to a gas attack in a crowded suburban shopping mall. The game objective is to save as many civilians as possible under time pressure. In the game, players need to quickly assess the situation, divide into teams, and coordinate with each other to identify source of chemical spill. Players can thus practice recognizing the signs of different chemicals and viruses, examining victims’ symptoms, and observing, the pattern of chemicals spread in differing environmental conditions [Susi et al., 2007].

Another example of a DR serious game is the Team Coordination Game developed by Toups, Kerne and Hamilton [Toups et al., 2011], which teaches participants effective cooperation and communication, based on a zero-fidelity simulation of team coordination that focuses

on distributed cognition in lieu of concrete details, yet draws directly from fire emergency response work practice [Toups et al., 2011]. User study and evaluation of the coordination game have suggested that a serious game built with a low fidelity approach can still help responders to improve their coordination skills.

The advantages of serious games in the Disaster Response domain may be obvious. Creating high fidelity exercises for Disaster Response could be costly and dangerous, while serious games allow participants to repetitively practice without danger or high cost. For instance, the game Biohazard [Susi et al., 2007] enables players to experiment with a multitude of different task conditions and strategies with a simple change of some game variables.

However, there are also concerns that existing serious games often fail to capture the social aspects of DR operations. For example, it is argued that the game Biohazard failed to capture the complexity of victims behaviour in highly stressful, emergency situations. Therefore, the game may not be able to help responders to develop interpersonal skill (expressed through voice, and gestures) which are required to deal with panicked victims [Susi et al., 2007]. In fact, serious games are usually not designed to simulate all aspects of DR operations, but only to mirror some aspects of DR practices and processes (e.g. the zero fidelity coordination game). Serious games are thus not a replacement for field trials and other training methods, but a tool responders can use to explore ideas and talk about their practice.

3.3.2 *Mixed Reality Games*

A Mixed Reality Game (MRG) is one type of game which tries to bridge the physical and the digital [Benford et al., 2005] worlds. The term ‘mixed reality’ refers to virtual experiences being played out in real-world spaces. This kind of games typically use pervasive technologies, such as cellular phones, GPS, Bluetooth, wireless networks

and sensors, with the aim to blend virtual game events into people's life and real world environment. Some researchers have recognised the potential to adapt MRGs for training purposes in DR domains [Fischer et al., 2012], as a mixed reality game is thought to be a powerful way of exposing participants to learning experiences not otherwise possible.

Arguably, mixed reality games are also becoming an established vehicle to study distributed interactions across multiple devices and ubiquitous computing environments 'in the wild' [Crabtree et al., 2006; Benford et al., 2005; Fischer et al., 2012]. The emergence of ubiquitous computing features distributed interaction across a burgeoning array of small, mobile devices and online environments. Because a mixed reality game also creates such an ubiquitous computing environment, mixed reality games are thought to be an ideal platform for ethnographers and system designers to investigate issues of distributed interactions in future ubiquitous computing systems.

Further, Fischer et al. [2012] also argues that a DR support system potentially shares a set of characteristics with a MRG, which suggests that a MRG could be a platform for investigating issues of interactions in the DR setting:

1. Bridging the physical and the digital. Both DR as well as MRGs routinely bridge the physical and the digital as part of their actors' coordination [Benford et al., 2005]. DR for example makes use of the twitterverse to inform real world response e.g. [Sarcevic et al., 2012].
2. Orchestration. DR and MRGs are both highly orchestrated activities. Authoring and orchestration tools 'behind the scenes' of an MRG, as well as player interfaces, provide managers, players and spectators with different temporal and spatial views of the game world in order to support the experience [Crabtree et al., 2004]. These settings are surprisingly comparable to the 'control

room' of a Disaster Response operation, in their collections of sophisticated technological arrangements to communicate and coordinate real-time information streams, in order to create a holistic view amidst an immersive setting of interest.

3. On-the-ground and online. In both DR, as well as in MRGs, people on the ground often work with people online to solve a common problem. Sarcevic et al. [2012] shows how understanding online content can foster understanding of medical coordination challenges in DR on the ground. MRGs often leverage the fact that people on the ground and online have different views of the world, which are turned into different abilities within the game [Flintham et al., 2003].

These key characteristics illustrate the overlapping characteristics between time-critical, distributed coordination in mixed reality games and DR operation, which underpins the motivation for applying the mixed reality game approach to investigate socio-technical issues involved in developing future DR support systems [Fischer et al., 2012].

3.4 SUMMARY

There has been a long-standing tradition in the HCI community to study the relationship between humans and technology support. Studies in three HCI sub domains (CSCW, automation design, and human agent interaction) all recognised the socio-technical challenges raised by integrating technology with humans' working process. The studies in this PhD work aligns with the tradition of empirical CSCW investigation of complex collaborative work in organisational setting. On the other hand, research on automation design has developed the LoA model to frame alternative ways we can automate a process. Although there are limitations, the model still provides useful guidelines/terminologies as a starting point for designing human system

interactions. Further, the study of human agent interaction tries to tackle the interaction challenge by developing agent software with social capabilities of teamworking. Various approaches have been adopted by HAI researcher (e.g. interface agent, mixed initiative, adjustable autonomy), which can serve as guidelines for interaction design. These approaches have inspired the design of the planning support component examined in this thesis

Further, the study of Disaster Response operations is difficult due to its safety critical nature. The serious game approach has been adopted for behavioural studies in Disaster Response settings. In particular, this PhD work adopts mixed reality gaming as the approach to study distributed interactions across multiple devices and ubiquitous computing environments ‘in the wild’.

Part II

METHODOLOGY AND APPROACH

4

APPROACH

This PhD work adopts a socio-technical view towards the planning support systems in Disaster Response settings. Introducing a planning support system to Disaster Response ([DR](#)) operations may create a socio-technical gap that needs to be considered by system designers. We argue that the gap can be reduced by an appropriate interaction design supported by a deep understanding of the socio-technical issues surrounding the planning support. In order to gain insight into the socio-technical issues, we adopted an ethnographic approach to explore and unpack interactions between humans and the planning support agent in a disaster setting. A serious Mixed Reality Game ([MRG](#)) approach is adopted to create the game AtomicOrchid ([AO](#)), which is used to simulate [DR](#) operations. We outline two interaction designs that are later implemented in the [AO](#) platform for field studies. In addition, efforts have been made to establish contact with a professional response agency, Rescue Global ([RG](#)), which took part in workshop centered on the [AO](#) game. Feedback on [AO](#) are collected from the discussions in the workshop.

In this chapter, we set out the socio-technical perspective to planning support systems adopted by this PhD work (section [4.1](#)), followed by an introduction to the serious Mixed Reality Game ([MRG](#)) approach (section [4.2](#)) and description of AtomicOrchid platform (section [4.3](#)). The section [4.4](#) outlines the two interaction designs that are studied in the [AO](#) field studies, and the section [4.5](#) gives an introduction to the workshop with [RG](#) for professional feedback.

4.1 ADOPTING THE SOCIO-TECHNICAL PERSPECTIVE

This research aims to inform the design of planning support for organizational work conducted by responder teams. This PhD work adopts a socio-technical view on the responder teams and their technological supporting systems. Integrating new technology support into a human organisation is a well-known challenge for socio-technical system design. In this PhD study, we anticipate the same challenges will be encountered when introducing a planning support system in Disaster Response operations.

The term socio-technical systems is used to describe systems that involve a complex interaction between humans, machines and the environmental aspects of the system (Section 3.1). The term stands for the recognition of both technical and social subsystems and the very complex relationship between the two. Social systems are characterized by phenomena such as communication and cooperation between human individuals, emergence of meaning systems, self-referential development of structures. In contrast, technical systems are characterized by artefacts, control, anticipation, state-transitions, pre-programmed adaptability, learning in respect to purposes which are determined from outside the system [Ropohl, 1999]. Introducing a technology support system into an organization requires the technical system and social aspects to be integrated.

In the context of this study, a multi-agent coordination algorithm is used to build an automated planning agent. From a technical point of view, the planning agent requires a set of rigid inputs and produces plans of task allocations. For a fully autonomous system, plans can be handed over to other components for execution. Re-planning is usually triggered when an unanticipated situation happens. However the plan-execution model may be very different from the way that human make plans and take actions (Figure 5). For human teams, no matter how well the plan is formed, its significance in determining the actual actions is questionable. Section 3.1 showed that actions are

highly dependent upon its material and social circumstances focusing on local interactions between actors, and environments of their action. Computer science is learning how to use techniques such as machine learning and user modelling to build agents with social capabilities (Section 3.2.2). However, such attempts are still based on the approach of plan/intent recognition, which tries to establish logical links between human actions and intention. It is hard to disprove that a technical solution is imminent, but many [HCI](#) researchers have pointed out that it could be extremely difficult as the causal connections between actions and intent is weak (Section 3.1). Arguably, the lack of causal links between intention, plan and actions makes human behaviours hard to predict and generalize from one situation to the next. Therefore, it is important to understand the specific organisational work and social processes in designing human system interactions that support, rather than hinder the social process of human teams. The aim of this thesis is to investigate specific social situations of [DR](#) organisational work to inform interaction design of planning support systems.

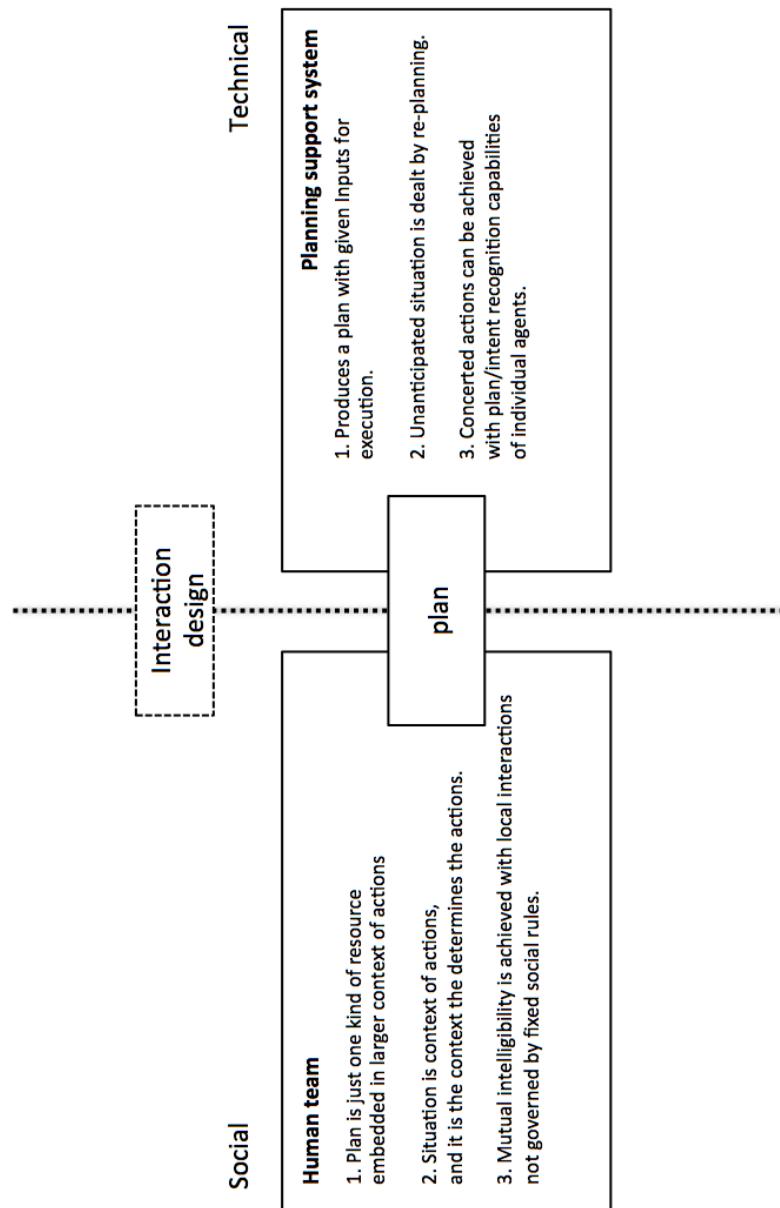


Figure 5: Actions and situated actions

4.2 SERIOUS MIXED REALITY GAMES AS A TESTBED

One of our work's main objectives is to study interaction and coordination situated in rich and 'messy' real-world socio-technical settings. As it is difficult to deploy technological prototypes in real disasters, a serious game approach has been adopted by researchers to study technology interaction in disaster scenarios through game-like simu-

lations (Section 3.3).

This PhD work also adopts serious game approach to simulate a Disaster Response setting in which distributed responder teams coordinate under temporal and spatial constraints (Section 2.1.3). More specifically, we create a Mixed Reality Game ([MRG](#)) as a testbed that enables studying team coordination, interaction and communication in a real-world disaster scenario whilst providing confidence in the efficacy of behavioural observations (Section 3.3).

The [MRG](#) testbed called AtomicOrchid ([AO](#)) simulates a radioactive incident. Participants of the game play both the role of responders ‘on the ground’, and coordinators in the control room. They coordinate with each other through GPS, map sharing and messaging, to achieve game objectives. The [AO](#) game system can be integrated with planning agents to support players on the ground, and the interaction layer between players and agents can be configured in different ways through modifications to the game interface. Through agent integration and interface modifications, we created three ‘probes’ of agent planning support with different interaction designs. The three probes were then used to conduct observational studies, which allow us to unpack human system interaction in the different interaction design settings.

4.3 THE ATOMICORCHID PLATFORM

We designed and implemented the mixed reality game AtomicOrchid ([AO](#)) as a testbed for our field trials of different system interaction designs. The game involves field players on the ground (playing as field responders) and online players in a control room (playing as Headquarters (HQ)).

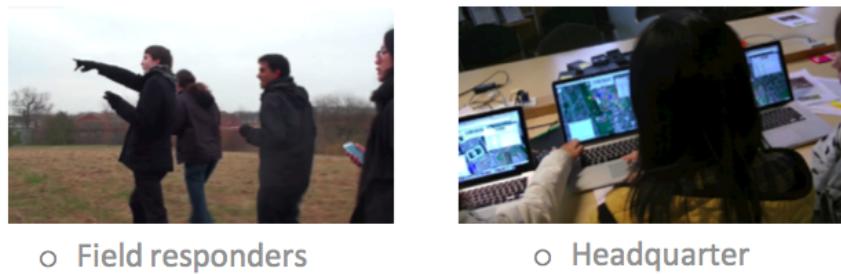


Figure 6: HQ and field players in AO

In the following sections, we give a detailed description of the game design, which covers grounding of the design rationale, the iterative design process, and the system architecture.

4.3.1 Game mechanics

The AtomicOrchid is based on the fictitious scenario of a radioactive explosions creating expanding and moving radioactive clouds that pose a threat to responders on the ground (field responders), and the (virtual) targets to be rescued from around the game area. We chose a radiation scenario because unlike disasters that cause physical devastation it poses an invisible threat, which creates the need to monitor the environment closely with sensing devices, and communicate frequently.

Field responders are supported by a centrally located headquarters (HQ) control room, staffed by HQ players as coordinators who exchange messages with field players through an instant messaging style communication system. The messages are broadcasted, which means they are visible to all players. The core game mechanics are designed to allow us to explore specific aspects of team coordination. In particular, this is inspired by the real coordination challenge of resource and task allocation to coordinate spatially distributed resources and personnel.

Responder roles and targets. Each field responder is assigned one of four roles:

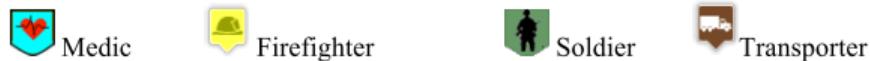


Figure 7: The roles in AO

There are also four types of (virtual) targets:



Figure 8: The targets in AO

The objective of the field responders is to rescue as many targets as possible by ‘carrying’ them to a drop off zone. To pick up and carry one of the target objects, two responders with particular appropriate roles must be in the proximity of the object. For example, a soldier and a transporter are required to pick up and carry fuel, and a medic and a soldier are required to pick up an animal (Figure 9).



Figure 9: Role-target mapping

The role-target mapping mechanic requires players to engage in resource coordination. Field responders have to engage in ‘agile teaming’ forming, disbanding, relocating and re-forming in teams over the course of the game in order to complete the game objective.

The radioactive cloud. The cloud is a danger zone that can incapacitate field responders. It imposes spatial and temporal constraints on task performance and well-being. The cloud is analogous to various spatial phenomena in disasters (e.g. spreading fires, diseases and floods). It requires communication between HQ and field responders, as the spatial position and movement of the cloud is only known to

HQ. The cloud is shown in a heatmap style in the Figure 10. The field player can detect ‘radioactive intensity’ at their current locations with the mobile responder app.



Figure 10: The radioactive cloud

Command-and-control (C₂) structure. The division of responsibility into HQ and field responders simulates a situation where responders are connected to a generic two level C₂ structure proposed by Chen et al. [2005]. This structure highlights the division between remote control room and on-site teams. On-site responders react to the immediate scene without global picture, while the coordination center deals with strategic issues and works with a global picture, leveraging external resources to help on-site response (Section 2.2). Instead of simulating a real command and control model such as Command and control model of Bronze-Silver-Gold (BSG) (Section 2.1.2), we chose the generic structure to simplify the game play, but to still mirror the main characteristics of C₂ in DR.

System interface. The system interface design is closely related to specific interaction designs, and it is refined through three iterations of field trials. Therefore, the details of interface evolution are left to be introduced in the subsequent chapters (Chapters 6, 7, 8) describing field trials.

4.3.2 System architecture

The AtomicOrchid system is based on the open-sourced geo-fencing game MapAttack¹ that has been iteratively developed for a responsive, (relatively) scalable experience. Our mixed-reality game relies especially on real-time data streaming between client and server. The client-server architecture is depicted in figure 11. Client-side requests for less dynamic content use HTTP. Game status and logic is stored and maintained on game server (Appendix A.2). Frequent events, such as location updates and radiation exposure, are streamed to clients using websockets to avoid the overhead of HTTP. In this way, field responders are kept informed in near real-time (Appendix A.1).

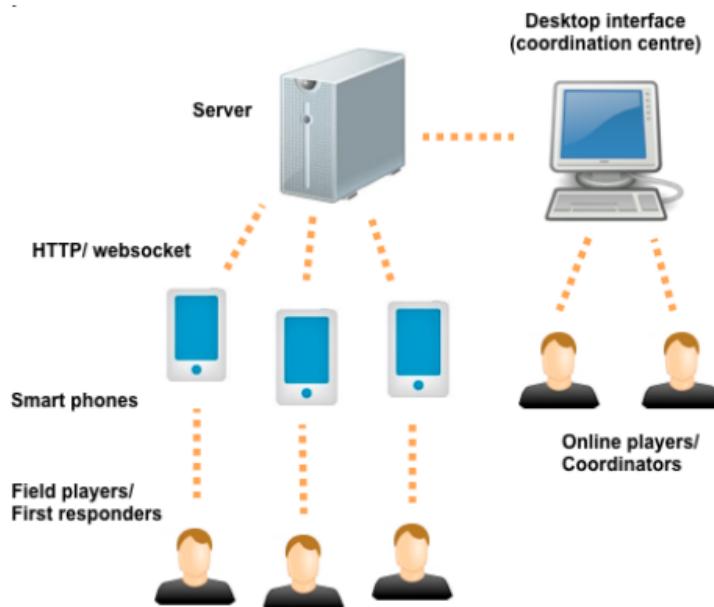


Figure 11: System architecture

The platform is built using Sinatra for Ruby (<http://www.sinatrarb.com/>), and current web technologies such as socket.io, node.js, and AngularJs, and the Google Maps API. Open source mobile client apps exist for iPhone and Android; we adapted an Android app to build the Mobile Responder App.

¹ <http://mapattack.org/>

4.3.3 *The planning agent*

In studies 2 and 3 (chapters 7 and 8), planning agents are integrated into AtomicOrchid to support player's planning activities. The agents are developed by ORCHID research partners from Southampton, more technical details of the planning agent is available in [Ramchurn et al. \[2015b\]](#). In what follows, we briefly describe technical details of the agents and system integration between AO and the agents.

The coordination problem (described in section 4.3.1) is modelled using a Multi-Agent Markov Decision Process (MMDP) that captures the uncertainties of task execution, extending earlier work [[Wu et al., 2015](#)]. The model allows responder actions to be delayed or to fail during the rescue process. The MMDP modelling leads to a large search space, even with a small problem. Hence, an approximate solution is devised to save computation time, which can be executed to support real-time planning. The planning algorithm takes into account both time (cloud and human movement speed) and spatial (path planning for responders) constraints. The planning algorithm run by the planning agent produces high quality task allocations that minimise the travelling distance of first responders, and maximise the number of targets rescued. Before the agent was deployed to support human teams in the game setting, computational simulations were used to benchmark the MMDP algorithm against greedy and myopic methods (Figure 2). The results confirm that the algorithm produces efficient task allocations.

| Metrics | MMDP | myopic | greedy |
|----------------------|------|--------|--------|
| #completed task | 71% | 65% | 41% |
| #responders survived | 100% | 25% | 0% |

Table 2: Result for MMDP, Myopic and Greedy algorithms, Credit [Ramchurn et al. \[2015b\]](#)

For integration, the agent is deployed on a separate server. It communicates with the AO game server through a pre-defined HTTP protocol (Appendix A.3). The agent takes the game status from the game server as input, which includes player's health, road connectivity, locations of players, targets and radioactive clouds. The output of the agent is a set of task assignments such as 'player A and player B, go to target 1' (Figure 12) . The task assignments are sent to the AO game server and presented to the game players. Detailed interaction design between human and the agent will be presented in Chapter 7 and 8. In order to facilitate the different interaction designs, the input of the agent is slightly different in study 2 and study 3, which will be detailed in sections 7.2.1 and 8.2.2.

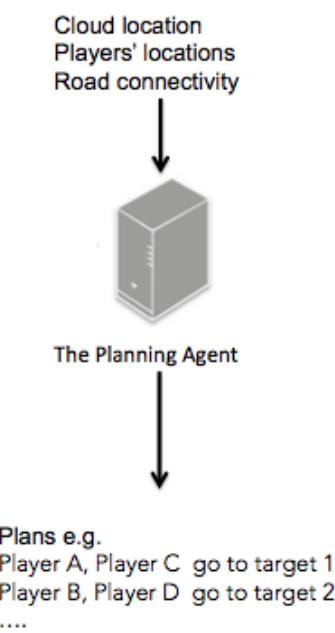


Figure 12: Input and output of the agents

4.3.4 Logging system

A logging system is also built into the AtomicOrchid system. The updates of game status sent between the game server and clients (both HQ and mobile clients) are all recorded in JSON format with system timestamps. The logs also contain the information of cloud status,

task location and status (pick up/ drop off), messages, player location and health. Automated tools were also developed to reconstruct game play and visualise (Section 5.3) the log for data analysis.

4.3.5 *Iterative design and development*

Before the game was deployed for observational studies in this PhD work, the game went through an iterative design and development process to test, refine game concepts and system robustness. We briefly describe three cycles of iterative game design and evaluation that took place before the system was ready for the first formal field study.

In the first iteration, we used a paper-based prototype to test and refine the core game mechanics. We recruited 12 participants, allocated one of four roles to them, and equipped them only with paper maps with locations of targets. They had to form different kinds of teams to retrieve the different kinds of boxes placed in the game area. The paper prototype demonstrated the demand for better support of situation awareness and communication to enable coordination. The resulting technology prototype was first tested with users in the second iteration. Users were equipped with the responder smartphone app to communicate, navigate, locate and pick up targets in teams formed according to role requirements. HQ was staffed by members of the research team. A pilot study was conducted with members of the public that visited an Open Day at a local university. A total of 20 members of the public tested the game in four ad-hoc game trials. The lessons learned in the pilot study revealed problems with user interaction, networking, and game parameter tuning, which we subsequently addressed.

In the third iteration, we improved system stability and interface design. We conducted a pilot study at the campus of another university, to test the system in place. The first full-fledged studies we report on in chapters 6, 7, and 8 were conducted shortly thereafter.

4.4 EXPLORING INTERACTION DESIGNS WITH THREE ATOMICORCHID STUDIES

Based on the serious game approach, three studies have been planned to explore the interactional issues related to the socio-technical integration of the planning agents and the responder team. To build such a socio-technical system, there are various ways to arrange the interaction between responder teams and a planning support system. Inspired by the Level of Automation ([LoA](#)) concept (Section [3.2.1.1](#)) from research on automation design, we outlined 4 paradigms of human system interaction loosely based on the automation level of planning activities: full manual, human in-the-loop, human on-the-loop, and Human-out-of-loop. This PhD work will only consider the former 3 notions of automation as explained before.

In research of automation design, the [LoA](#) model has been developed to categorise systems into a linear spectrum according to degree of automation. Arguably, the model may not fit perfectly into the context of socio-technical system due to some of its limitations identified in section [3.2.1.1](#). However, the terminologies that come with the model can still serve as a reference point for interaction designs to be studied in this PhD work.

1. **Human out-of-the-loop.** Out-of-the-loop represents the highest level of automation. An out-of-the-loop system is supposed to run completely independently. The human is replaced by the machine, therefore no human system interaction is required. This is unlikely to be realized in a socio-technical system, in which organisational work is mainly carried out by humans and supported by technologies.
2. **Human on-the-loop.** In this research we use the term Human on-the-loop to describe a system with high level of automation, which requires a minimum level of human intervention. Compared to out-of-the-loop, the on-the-loop system is designed to

run without human intervention most of the times. However, human supervision and intervention are still required for contingencies.

3. **Human in-the-loop.** In this research human in-the-loop represents a system with medium level of automation. Compared to the on-the-loop system, the in-the-Loop system can not run without human input. Constant human interactions are required to achieve the goal of the socio-technical system.
4. **Full manual.** A full manual system describes a system without automation. In the context of AtomicOrchid, the platform without integration of the planning agent can be seen as a full manual system.

In the context of the AtomicOrchid platform, the notions of in-the-loop and on-the-loop can be used to describe the degree to which the planning agent automates the real-time task planning and the extent to which the human Headquarters need to be involved in the plan-execution loop. Guided by these two notions, we devised two detailed interaction designs for integrating the planning agent into the AO game. In the next two sections, we give detailed description of the two interaction designs, followed up by an overview of the three field studies, which details how a sequence of system prototyping and field trials was organised based on the two interaction designs, and how they are designed to serve the research objectives.

4.4.1 *The on-the-loop interaction*

The on-the-loop interaction is designed to facilitate the division of labour between humans and agent: a planning agent routinely assigns tasks to distributed responder teams, while human coordinators (the HQ) monitor and support the task execution by responding to arising contingencies (see figure 13). In this design, the agent can directly contact field responders to allocate tasks. The responsibility

of the planning agent is to generate and distribute plans for execution. The agent is also responsible for initiating re-planning according to changes in game status. The agent also directly handles feedback from the field player, i.e. the field players can give feedback to the agent by accepting or rejecting the plan, while the agent can generate new plans according to the feedback.

The role of the HQ is to monitor the planning process and provide support when contingencies arise. For example, the HQ may decide to stop some tasks issued by the agent if the threat of radiation increases unexpectedly. It should be noted that the agent can operate without HQ input, and the HQ intervention is supposed to be only occasional.

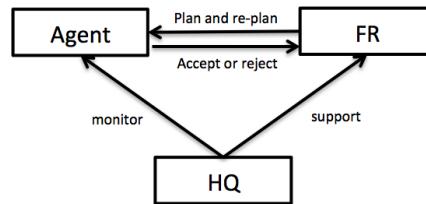


Figure 13: On-the-loop interaction design

4.4.2 *The in-the-loop interaction*

In-the-loop interaction is designed to facilitate a different pattern of labour division between humans and agent: a planning agent proposes the task assignments, and the human HQ needs to approve the tasks before they are sent to the field responders. In this design, the HQ can be seen as a mediator between field responder and the planning agent. If the HQ does not agree with a task allocations from agents, they can intervene by directly editing part of the plan or require the agent to re-plan.

Similarly, the feedback from the field responders (i.e. accept/reject) are delivered to HQ before any actions are taken. The HQ are respon-

sible for reviewing the feedback and deciding what actions to take (e.g. decide to initiate re-plan, or ignore). Compared to on-the-loop interaction, the agent in this design will never directly communicate with field responders and the agent cannot operate without the HQ's input, i.e. the HQ have to make decisions on every agent proposed task, and take actions on the feedback from field responders.

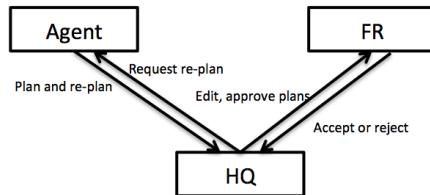


Figure 14: In-the-loop interaction design

4.4.3 *The three AO game studies*

Three studies are conducted to explore the socio-technical issues related to human-agent interaction. The first study focuses on a “manual” version of AtomicOrchid without agent support, while the latter two studies focus on on-the-loop and in-the-loop planning support respectively. For each study, we facilitate several Mixed Reality Game ([MRG](#)) sessions to study the interaction design paradigms.

The first study aims to observe and explore human coordination without planning agent support. The non-agent trial (Chapter 6) supports the two later (Chapter 7, 8) agent-supported system trials by 1) revealing baseline performance of human coordination without agent support 2) and generating design requirements which feed into subsequent prototyping of AtomicOrchid. The purpose of the second and third studies is to investigate socio-technical issues related to the on-the-loop and in-the-loop interactions and derive design implications of interaction designs from the field observations.

4.5 COLLABORATION WITH A PROFESSIONAL DISASTER RESPONSE ORGANISATION

In addition to the three observational studies of AtomicOrchid games, a workshop with Rescue Global (a professional Disaster Response agency, Section 4.5) was organised to get professional feedback about the AtomicOrchid system and the planning support agent. Because the contact with Rescue Global was established at a late stage of this PhD work, the feedback from the RG workshop are not used to drive the development of AO game and interaction design, but to get an insight into the similarities and differences between AO simulation and real world DR operations, which help us to understand limitations and strengths of our observational study. The in-the-loop AO probe was demonstrated to the RG team and a discussion was organised to get feedback from RG. It included a hands-on session for RG members to experience the AO game, and further feedback was collected from discussions during and after the game session.

4.5.1 *Introduction of Rescue Global*

Rescue Global (RG) is a Disaster Response organisation. They are a UK charity and a US not-for-profit headquartered in London, UK. Their remit is to provide “immediate crisis and disaster reconnaissance ability, delivering accurate and timely information and risk data, as well as performing emergency search and rescue operations where needed to save life.” One example of their operation is the deployment of a reconnaissance team in the Philippines for super typhoon Haiyan in 2013. After the typhoon strike, the team conducted disaster reconnaissance on isolated islands from the air and on the ground, assessing needs and delivering aid based on priorities of water, food, medicine and shelter.

RG’s organisational structure represents a typical hierarchy found in emergency services [U.S. Department of Homeland Security, 2008],

termed Gold, Silver and Bronze. Gold denotes the strategic lead, which is associated with RG's senior officers (Section 2.1.2) and the headquarters in London, Silver is the tactical lead, which is 'spun up' for mission planning, both to assess feasibility of deployments and when actually deployed on-site. Bronze refers to the operational level, in which 'Pathfinders' (field responders) carry out operations 'on the ground' supported by Silver command [Rescue Global, 2012]. RG's core staff consists of around 20 highly specialised experts and admin support, many of whom have had prior careers in the military, emergency and first response services.

METHODOLOGY TO INVESTIGATE HUMAN-AGENT INTERACTION

This chapter takes an in-depth look at the methodology that underlies the empirical approach adopted in the presented studies. This PhD focuses on ethnomethodology-oriented field studies based on the AtomicOrchid (AO) platform to generate descriptive results, which contains rich interactions among participants and the planning support system. Ethnographic observations and interaction analysis are central to all three field studies, while group interviews, message classification, and system log analysis supplement the two former in-situ methods.

5.1 ETHNOMETHODOLOGICAL PERSPECTIVE

Observation of participants in the field study is informed by Ethnomethodology (EM). Following the tradition of ethnography, EM seeks to explicate the real-world organisation of work by adopting a naturalistic stance. EM places methodological emphasis on rigorous description of the situated (i.e. local, observable) actions and practices [Suchman, 1987] in and through the contingent accomplishment of daily activities. EM-informed ethnography arguably helps to answer what might be regarded as an essential question in design: what to automate and what to leave to human skill, competence, judgement, experience and expertise [Crabtree et al., 2012]. By producing description of the actions and practices in and through which the work ‘gets done’ time and time again by the members, EM can inform system design by uncovering what actions and activities we should therefore support.

For the purpose of this thesis, the social situation of the interaction with and around the planning support is argued to be a critical factor to understand how the social organisation of work is achieved. Observation of the situated actions and practice employed by the participants is a key method for the field study. The use of the system was observed and filmed for later analysis. Video is widely recognised as an important resource for ethnography around technology use [Crabtree et al., 2012]. The next section goes through the method of video-based interaction analysis for unpacking the interactions observed in the field.

5.2 INTERACTION ANALYSIS

Interaction analysis can be defined as an interdisciplinary method for empirical investigation of interaction of human beings with each other and with objects in their environments [Jordan and Henderson, 1995]. In the context of HCI study, it is a method of analysing naturally occurring talk and activity, with the aim of uncovering and describing something of the order and organisation by which people interpret and interact with each other and with the things around them.

The advantage of interaction analysis lies in its ability to deal with actual details of technologically mediated interactions and allows technology developers to see exactly how technology fits (or doesn't fit) into current working practice. Other methods such as questionnaires and interviews rely upon reports from participants, rather than actual, reasoning and behaviour. The over-reliance on participants' report make those methods vulnerable to the problems of people producing post-hoc rationalisations of actions, forgetting or incorrectly estimating aspects of behaviour, expressing ineffective attitudes, and generally lacking insight into the tacit procedures underlying much of their activity. Instead, interaction analysis can expose the practical reasoning activities of participant's themselves in a way which does

not require them having to remember, justify or even know what they did. This effectively indicates how people think and make sense of technology they are using, in the performance of some task. However, interaction analysis is extremely time-consuming, which means it can only be carried out on small numbers of participants. This limitation makes it unsuitable for answering to very specific design questions and for examining the needs and behaviours of diverse groups of people. Further, the generality of its findings may need to be established by other means [Jordan and Henderson, 1995].

For the purpose of this PhD work, interaction analysis is applied to the observations of the game probes undergoing field trials. In this case, the description generated by interaction analysis could expose information on the sequential organisation of technologically and socially mediated activities, which in turn, reveals how the activities can be supported. The main resource of interaction analysis is video recordings of AO game plays. The video analysis generally consists of three stages [Heath et al., 2010] :

1. Cataloguing the data corpus. This step involves a preliminary review of the corpus. Basic aspects of the activities and events are catalogued at this stage. Preliminary reviews and cataloguing should involve no more than a simple description and classification of the materials without detailed analysis.
2. Selecting Episodes. In light of preliminary review of data, a more focused substantive review of data is carried out in this stage. Repeated analytical searches of the data corpus are also involved to find examples of actions that appear to reflect similar characteristics. Candidate episodes of the particular phenomena, actions or organisation under scrutiny should be gathered and put into collections.
3. Detailed analysis. We begin to look more closely at the selected candidate episodes to unpack the way in which interaction is

accomplished by participants. The process generally involves transcribing and analysing both talk and visible conduct in the candidate episodes.

In the later chapters, results of interaction analysis are presented in detailed episodes of game play. In those episodes, a standard orthographic notation is used [Jordan and Henderson, 1995] and complemented by timestamps [0:00], and system messages from remote players and HQ. Players can be uniquely identified by their initials. Targets are denoted by their unique numeric target id. Task assignments from the planner support system are represented as two initials and one target id connected by a rightward arrow. For example, the notation PC, CR -> 22 means player PC and CR are instructed to team up and go for target 22.

5.3 DATA COLLECTION AND HANDLING

Interaction analysis has been introduced as the main method for investigating socio-technical issues in our studies. This section introduces a number of methods employed for data collection and handling. In particular, the group interview supplements field observation by providing subjective description of game play experience. The message classification method gives quantitative overview of remote communication. It also provides context for interactions in the field and helps to identify interesting game events. The system log analysis produces game event visualisations and replay. When triangulated with the video data, the log data analysis also supports interaction analysis by providing context and helping to identify interesting episodes in videos.

5.3.1 *Video and audio recording*

Audio recorders and video cameras are believed to valuable resources for ethnographic study. Both audio and video recordings offer us a rich resource and enable us to elucidate the methodical ways in which work is organised and accomplished as an interactional matter [Crabtree et al., 2012]. This PhD work uses video/audio recordings to capture distributed activities in the AtomicOrchid (AO) game as it happens, and the subsequent interaction analysis is based on reviewing the video recordings.

For each AO study, multiple researchers were hired to capture activities of distributed teams in the field. The researchers were instructed to follow player teams and film their actions including talking, gestures, and other bodily activities. In some cases, there weren't enough researchers to cover all the player teams in the field. To maximise the number of teams covered, the researchers were instructed to avoid filming the same player teams at the same time. In the control room, one researcher recorded the actions of Headquarters players with two camcorders. One camcorder was fixed on a tripod and the other was held by the researcher.

Audio recordings were used to supplement videos. An audio recording app was installed in the Android phones that were used by the field players in the AO game. The app works in the background, recording player's voices without interrupting players' use of the AtomicOrchid client app. The obvious limitation of audio is that we cannot visually see player's actions with it, while its strength lies in its guaranteed coverage of all player teams at all times. As a great deal of the work of a setting is conducted through talk [Crabtree et al., 2012], audio recordings are useful alternative resources for interaction analysis when video coverage is not sufficient.

5.3.2 Log data handling

Human Computer Interaction ([HCI](#)) researchers often collect rich dataset for investigating interactions. The data set becomes larger and larger as digital record systems increase in availability [[Brundell et al., 2008](#)]. Automated tools are increasingly necessary for managing the organisation, replaying, structured and free coding (and annotation) and analysis of these growing data sets [[Brundell et al., 2008](#)]. For this research, the logging system of AtomicOrchid produces time-stamped system logs. The raw log data is hard to be used directly as a resource for interaction analysis. In order to reveal the information buried in the logs, the data has to be processed so that it can be easily read and triangulated with data in other modalities i.e. video and audio recordings. There are a number of tools already in existence to automate data handling and support the analysis of interactions. However, these tools often have limited or very specific functionalities [[Brundell et al., 2008](#)]. Therefore, we developed our own data visualisation and log replay system which are tailored to handle the raw log data from the AtomicOrchid studies.

To recap (Section [4.3.4](#)), the logging system records players's location, health, targets' location and status (pick up/drop off), task assignments and players' feedback (reject/accept). The data visualisation tool (Figure [15](#)) focus on visualising task assignments in the game play. The game events related to task allocations, including task assignments, player feedback, target pickup and dropoff, are all plotted on a time line with different annotations. The colored dots and squares denote various game events (Figure [15](#)). Detailed information of the event can be displayed when mouse hover on. As can be seen in Figure [15](#), the deep blue dot with a mouse cursor hover on denotes a target pick up event. The two players with initials CY and DS (see lines connecting the pick-up events) picked up target 598 at the 10th minute of game play. This data visualisation tools give an overview of the game-event sequence. It assists interaction analysis by providing

context and guidance in the episode selection process.

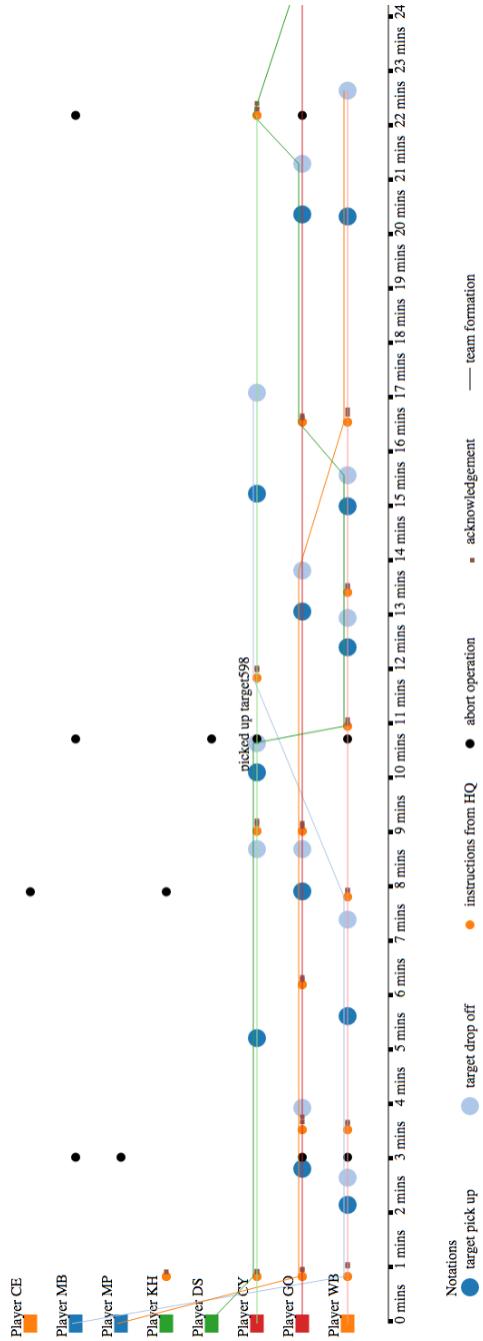


Figure 15: Log visualisation

A replay system was also built to triangulate multiple videos with log data. The main map view on the replay interface (Figure 16) displays game status reconstructed from log data, in a way that is similar to the HQ interface (Section 6.2). By giving a time offset to each video file, the videos are synchronised with the main map view of game status. The replay system is an important tool for interaction analy-

sis, as it presents distributed game play within a single interface in a synchronised way, providing insights into the distributed interaction among participants and the system as it happened in parallel.

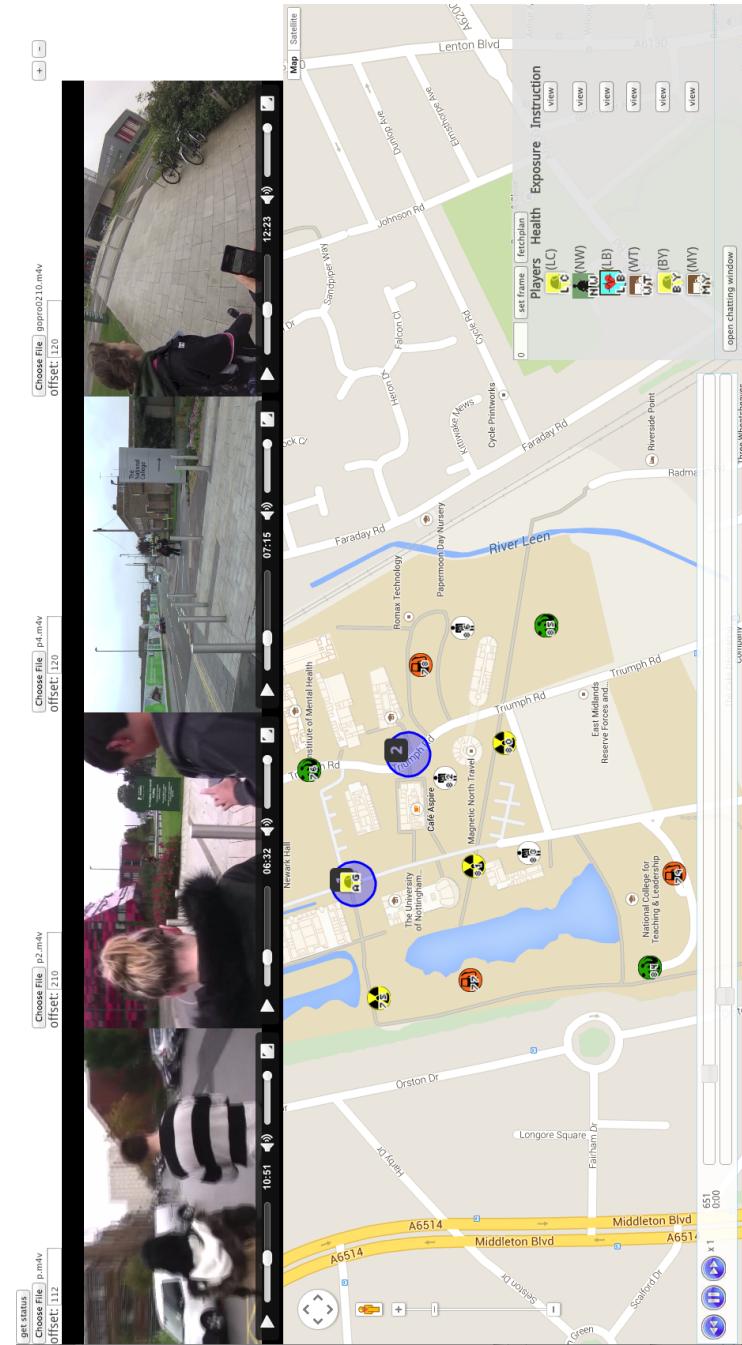


Figure 16: Replay system

5.3.3 *Message classification*

In AtomicOrchid, remote coordination between field and human HQ is achieved through a text messaging channel. The remote messages are recorded as part of system logs. To understand how the team members interact through the remote messages, we devised a message classification method based on speech act theory [Searle, 1976]. We used speech-act theory and the notion of adjacency pairs [Avrahami and Hudson, 2004] to classify messages sent between and among responders and HQ. According to speech act theory, utterances in dialogues can be considered as speech acts from three dimensions [Searle, 1976]. We were primarily concerned with the illocutionary dimension of speech acts.

Searle's classification of illocutionary acts is used to categorize messages in the communication system as follows.

1. Assertives: speech acts that commit a speaker to the truth of the expressed proposition.
2. Directives: speech acts that are meant to cause the hearer to take a particular action, e.g. requests, commands and advice.
3. Commissives: speech acts that commit a speaker to some future action, e.g. promises and oaths.
4. Expressives: speech acts that express the speaker's attitudes and emotions towards the proposition, e.g. congratulations, apologies and thanks.
5. Declarations: speech acts that change the reality in accord with the proposition of the declaration, e.g. pronouncing someone guilty.

The notion of request-response adjacency pairs are also used to gain insights into the reciprocity of communication. In linguistics, adjacency pairs describe conversational turn taking [Avrahami and

Hudson, 2004]. In AtomicOrchid, we expected many actions in remote conversation to be accomplished through pairs of utterances such as request-response, question-answer, or inform-acknowledge. For simplicity, we ignore the typology of adjacency pairs and treat all pairs as request-> response. Any utterance that expects a response is considered as a request.

The purpose of this message classification is to give an overview of the communication in the message channel. Meanwhile, the results of message classification supports interaction analysis, as it helps to identify interesting moments of team interaction in the game such as important decision points.

5.3.4 *Group interview*

For all three field studies, group interviews were conducted with all participants after each game sessions. The interviews consist of open-ended questions with the aim to supplement the field observation and interaction analysis with participants' comments about their experiences of the game. The interview is 'informal and unstructured' in the sense that it is not driven by pre-defined questions, but only by the research scope and interest. It is conducted in the manner of a conversation taking place between the researcher and the participants [Crabtree et al., 2012]. The interview does not stand on its own to provide distinctive results. The primary aim of the interview is to develop an overview of participants' experience of the game. Meanwhile, emergence of unanticipated issues and events was also fostered by asking open-ended questions, which in turn, are used to establish context for the issues for interaction analysis.

5.4 ANALYTIC PROCEDURE

To sum up, interaction analysis is the main analytic method applied in the studies. The data collection and handling processes are supported by methods including log data handling, group interviews and message classification. With this inventory of research methods, we can depict a typical analytic procedure for analysing the interactions in as AtomicOrchid study.

The procedure (Figure 17) begins with field studies after which a set of data are collected from three sources including system logs (1.1), video/audio recording (1.2) and group interviews (1.3). The message logs are then classified according to speech act theory. The resulting classification gives quantitative insight into the remote communication. The further log data handling produces replay and visualisation of game events.

The outputs of the data handling process are then used as resources for interaction analysis. The message classification (2.1) contribute to the catalogue building process (3.1) in the interaction analysis by providing the context of remote communication. The data visualisation (2.2) also helps to identify important episodes of interaction and provide context for further episode analysis. The game replay (2.2) triangulate video recordings with system logs. It is used as the major tool for in-depth data examination in the episode selection and analysis process (3.2,3.3), as it provides a synchronised view of multiple videos and system logs. Additionally, the players' comments from the group interview give us insights into participants' subjective game experience, which provides further context for episode analysis.

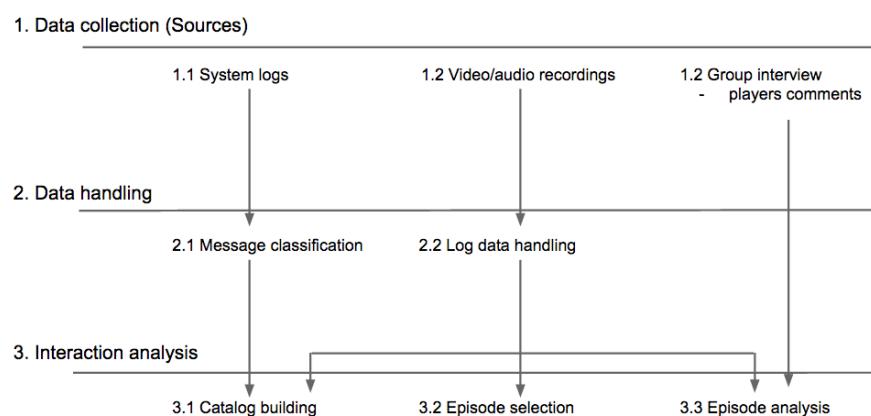


Figure 17: Analytic procedure

Part III
STUDIES

ATOMICORCHID STUDY 1: NON AGENT VERSION

In this study, we analyse team interactions in the AtomicOrchid (AO) game setting which simulates a time-critical distributed task environment in a Disaster Response operation. This study uses the full manual version of AtomicOrchid (AO) without automated planning support. Players are provided with basic coordination support including remote text messaging, GPS/map sharing. Interaction analysis is conducted to examine log data and field observations revealing local and remote coordination within the responder team. We generate design implications for automated planning support and uncover requirements that highlight the role of local coordination, decision-making resources, geo-spatial referencing and message handling.

6.1 INTRODUCTION

In this study, we aim to use the non-agent version of the AtomicOrchid game as a research probe to uncover the requirements and design implication for building an ‘intelligent’ coordination support system. The AtomicOrchid game creates a socio-technical setting in which player teams plan and execute spatially distributed tasks (see section 4.1). Although AI researchers have envisioned that an intelligent agent can support task planning by providing computational optimised task allocations in real-time, we focus on a manual version of AtomicOrchid, which does not involve any computational planning support. The primary objective is to unpack how human teams coordinate in the time and space constrained task setting through interaction analysis of behavioural data collected from field trials. In particular, the interaction analysis focuses on two aspects of coordination in AO, namely ‘remote’ and ‘local’. The remote aspect is con-

cerned with the coordination activities across distributed teams and remote HQ, which are typically mediated by computational systems. The local aspects are about coordination within co-located teams, in which face-to-face conversations play a major role. Drawing on the results of the interaction analysis, we aim to generate design requirements and implications for automated planning support.

Additionally, this study also supports our later system development and trials. As the first of three iterative trials in this PhD work, this non-agent trial supports the agent-integrated trials reported in Chapter 7 and 8 by (1) revealing baseline performance of human coordination without agent support (2) generating design requirements which feed into subsequent refinements of AtomicOrchid. The requirements are critical in that (1) the later studies can use them to recognize non-agent related design factors and (2) it also can inspire the interaction design between agent and responders in later system refinement.

In what follows, we expand on the description of the AO system in chapter 4 with a focus on interfaces and interaction design that are specific to this study. We then present results of the interaction analysis, followed by a discussion of performance implications derived from the results, before we move to design requirements.

6.2 SYSTEM DESCRIPTION

The basic game mechanic and system architecture have been introduced in the chapter 4. This section gives a detailed description of the system interface that support coordination between the field responder(FR) and headquarter(HQ) players.

In this study, HQ is manned by 2-3 coordinators. All of the coordinators are provided with a web-based coordination interface (Figure 18). The interface gives them an overview of the game status and en-

able them to communicate with the field responders.

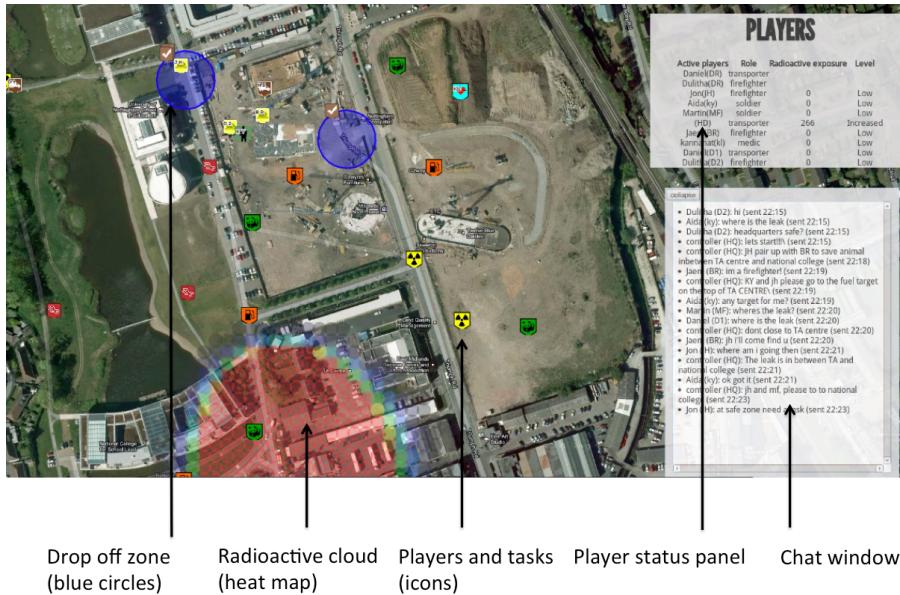


Figure 18: The HQ interface

As can be seen in the figure 18, the majority of the interface is occupied by a map-based presentation of the current game status. Roles and locations of field responders are represented on the map as icons. The field responders can be uniquely identified by their initials shown on the icons. The target types and locations are also shown as icons on the map. Location and intensity of radioactivity is indicated by a heatmap. Health status (health value ranges from 0 to 100) of the field responders is displayed on the right-top panel. A chatbox is placed at the right bottom for HQ to browse and send messages. The messaging system follows a broadcasting model. Everyone can send messages to one public channel, and the messages are visible to every player through the mobile and HQ interface.

Field responders are equipped with a mobile responder app providing them with sensing and awareness capabilities (Figure 19). There are two tabs in the responder app. The “map” tab displays a map showing locations of field responders and targets, which is similar to the map on the HQ interface, except that the radioactivity is not shown. The radiation level of the players’ current location is displayed as a Geiger counter reading (shown as a number on the top left of the

screen), which ranges from 0 to 100. Health status of the field responder is indicated by a health bar on the right side of the Geiger counter. The chatbox (similar to the one on HQ interface) is placed on the "Messages" tab for the field player to receive and send messages.

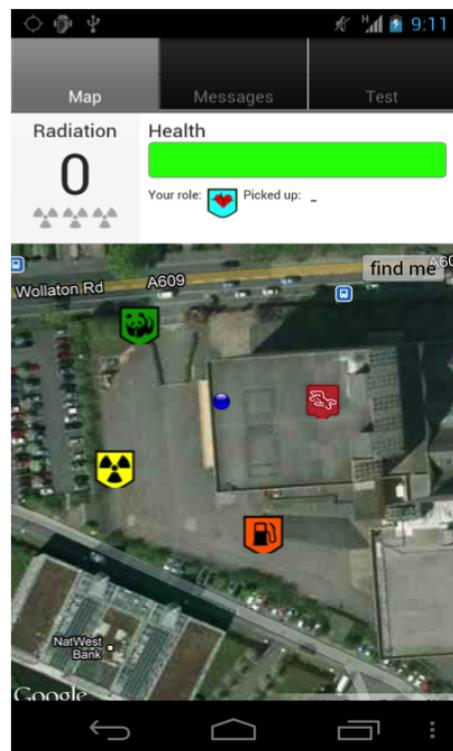


Figure 19: The mobile responder app

6.3 STUDY DESIGN

We ran two AtomicOrchid Game sessions. We describe participants, procedure, session configuration, and methods used to collect and analyse quantitative and qualitative data.

Study participants were recruited through posters and emails. A total of 18 participants were recruited; 7 participated in session A and 11 in session B. All participants were reimbursed with 15 pounds for 1.5 hours of study. In session 1, there was 2 HQ player and 5 field players. In session B, there were 8 field players and 3 HQ players. The session A have fewer participants than session B because fewer participants turned up in the study of session A. Although participants

are fewer than expected, we still believe that session A can provide valid observation of organisational conduct in the AO setting.

The majority of participants were students of the local university (Appendix B.2). Upon arrival in the HQ (set up in a meeting room at the local university), participants were briefed and asked to consent to participate. Roles were randomly assigned to all participants (HQ/field responders: firefighter, medic, transporter, soldier). Field responders were provided with a smartphone; HQ coordinators with a laptop. Game rules and interfaces were introduced, and participants were assisted in setting up their phones and laptop clients. Field responders and HQ coordinators were given 5 minutes to discuss a common game strategy. All field responders were accompanied to the starting point within the designated game area, about 1 minute walk from headquarters.

Once field responders were ready to start, HQ sent a “game start” message. Gameplay commenced for 30 minutes. A “Game over” message by HQ concluded the game. Field responders returned to HQ for the post-game session. A group interview was then conducted, before participants were debriefed and dismissed.

The size of the game area on the local university campus is 400 by 400 meters, without heavy traffic (Appendix B.1). The terrain of the game area includes grassland, a lake, buildings, roads, and footpaths and lawns. There are two drop off zones and 16 targets. The pilot study showed that this was a challenging, yet not overwhelming number of targets to collect in a 30 min game session. There were four targets for each of the four target types. The pattern of cloud movement and expansion was the same for both game sessions.

As described in chapter 4, we took a mixed methods approach to data collection and analysis. Five researchers with camcorders recorded the game play. One researcher recorded action in the HQ, and four other researchers each recorded a field responder team. In addition to

video recordings, a semi-structured group interview was conducted at the end, aiming at eliciting important decision points, strategies and the overall decision-making process. Time stamped system logs were collected that contained a complete record of the game play, including responders' GPS location, their health status and radioactive exposure, messages, cloud location, locations of target objects and task status.

Data handling Firstly, the log data (including remote messages) are handled by a digital replay system to reconstruct the game state (Appendix B.3) that can be triangulated with video data to support interaction analysis (Section 5.3.2). Secondly, to give an overview of how remote messages are used as a coordination resource, we used speech-act theory and the notion of adjacency pairs in linguistics to classify messages sent between and among responders and HQ (Section 5.3.3).

Interaction analysis We focus on the analysis of local field responders' interaction to unpack team coordination, including handling of messages sent by HQ. Video recordings of field action were catalogued to identify sequences (episodes) of interest (5.3.2). Key decision points in teaming and task allocation served to index the episodes. Interesting distinct units of interaction were transcribed and triangulated with log files of relevant game activity for deeper analysis that we present in later this chapter. Results of interaction analysis will be presented by detailed episodes of game play, with standard orthographic notations introduced in section 5.2.

6.4 DATA ANALYSIS AND RESULTS

Here, we present findings from interaction analysis supported by message classification that reveal how team coordination was achieved. Overall, responders rescued 7 and 9 targets in sessions A and B, respectively, out of 16 targets in total per session (Table 7). Two players

were incapacitated in session A, and 1 player was incapacitated in session B. 117 and 70 messages were sent in sessions A and B, respectively.

| | Saved targets | Incapacitated players | Remote messages |
|-----------|---------------|-----------------------|-----------------|
| Session A | 7 (out of 16) | 2 | 117 |
| Session B | 9 (out of 16) | 1 | 70 |

Table 3: Overview of game results

In what follows, results of the message classification will be presented first, followed by detailed analysis of episodes.

6.4.1 *Results of message classification*

We used Searle's classification (Section 5.3.3) of speech acts to categorize messages (Table 4). The table 4 shows that the majority of messages are directives and assertives sent by HQ. The majority of messages from field responders are requests for information, team and tasks. In what follows, we present examples for each category of speech acts, which in turn, provide an overview of remote coordination that is supported by the remote messaging system.

6.4.1.1 *Directives*

Most messages in the category of directives are instructions sent by headquarter (HQ) players. The content of instructions can be related to two themes: task allocation and task execution. The purpose of task allocation instructions is to distribute plans to field teams and require them to execute it. Most instructions in this category follow a common pattern. Taking the following message as example:

| Speech acts | Session A | | Session B | | Example | Total |
|--------------|-----------|----|-----------|----|--|---------|
| | HQ | FR | HQ | FR | | |
| Directives | 57 | 0 | 32 | 0 | JH pair p with BR to save animal in between TA centre and national college | 89(47)% |
| Assertives | 25 | 2 | 8 | 4 | The leak around geospatial is bigger | 39(20)% |
| Expressives | 5 | 0 | 0 | 0 | Good Job, JJ, TV and RL | 5(2%) |
| Declarations | 3 | 0 | 0 | 0 | NOTICE - TEAM B: NS + TD | 3(1.6%) |
| Commissives | 0 | 4 | 0 | 4 | ok got it | 8(4%) |
| Requests | 8 | 6 | 0 | 19 | wheres the leak? | 34(18%) |
| Unclassified | | 7 | | 2 | | 9(5%) |

Table 4: Speech act classification

"HQ : JH1 pair up with BR to save animal inbetween TA centre and national college"

The instruction sent from HQ consists of two parts: (1) Description of Teaming (who are involved) (2) Description of location (Targets) to go to. It is worth mentioning that HQ players use different strategies when they try to describe a location to field players. HQ players in session B frequently referred to landmarks on the map in their description, while HQ in session A used simple directions (north, west, south, east). For example:

"HQ: TEAM A, can you head south to the radiation and animal targets? "

The purpose of task execution instructions is to help players execute their tasks after they have been assigned tasks. Most instructions in this category are related to radioactive cloud. To help field players avoid the radioactive clouds, HQ players frequently send directions to field players or simply urge field players to move quicker. For example:

"TEAM B you need to be quick"

6.4.1.2 *Assertives*

Assertives provide plain information to recipients. Most assertives are sent by HQ because they have access to critical information - the cloud location. Followings are two examples of assertives:

"the leak around nottingham geospatial is bigger"

"HQ:There's another leak by the lake!"

Interaction analysis shows that assertives are important for field players to maintain situational awareness. We will talk more about this shortly in the section [6.4.4](#).

6.4.1.3 *Commissives, expressives and declarations*

We also identified a small number of commissives, expressives and declarations. Commissives are field player's responses to an assertive or directive. It can be an acknowledgement of receiving a piece of information or commitment to execute a plan. (e.g. "ok got it", "I am heading there"). Expressives are typically HQ's congratulations to field players. (e.g. "HQ:Good Job, JJ, TV and RL") In session A, HQ players sometimes declare field players to be in a team (e.g. "NOTICE - TEAM B: NS + TD"). The declarations help HQ to refer to a team more easily.

6.4.1.4 *Requests and Adjacency pairs*

Here we present adjacency pairs identified in the message logs (Section [5.3.3](#)). We found a number of requests sent from HQ and field players (14 in session A and 20 in session B). Those requests can be

related to a number of themes (Table 5).

| | Themes | Example |
|----|-------------------------|-------------------------------|
| FR | Task assignment | Anything for us to do? |
| | Teaming | Firefighter with me for fuel? |
| | Cloud info | Wheres the leak? |
| HQ | Player status | Firefighter who is free now? |
| | Acknowledgement request | Firefighter, respond |

Table 5: Themes of requests

In comparison, only a small number of adjacency pairs are found in both sessions (8 in A and 8 in B), which means not all requests are responded to (Table 6).

| | Total requests | HQ requests/ with no response | FR requests/ with no response |
|-----------|----------------|-------------------------------|-------------------------------|
| Session A | 14 | 8/7 | 6/1 |
| Session B | 20 | 1/0 | 19/14 |

Table 6: Adjacency pairs

It is also worth mentioning that field players didn't send acknowledgements to directives from headquarter. Although we do not classify directives as a request which expects an answer, the headquarter players express their frustration of not having responses to their instructions. A HQ player said in the group interview:

"I guess they did not look at it, they could not respond to it, we were like saying "where are you, respond", but they did not respond. I guess they are busy seeing themselves and the targets"

A field players also commented on the issue:

"I almost would not use the communication system because I was too focused on trying to save the targets."

"Sometimes I check whether the radiation is close to us, but mostly the communication is between us (local team members)"

6.4.1.5 *Summary*

To sum up, the majority of messages are assertives and directives from HQ. Directives are instructions sent by HQs as their attempts to guide/control task planning and execution of the team, while the assertives are mainly informational content about status of the danger zone (i.e. the cloud). A small number of commssives, expressives and declarations are also found in the messages. Both HQ and field responders send requests for various purposes (e.g. request for task, teammate, cloud status, confirmations). Based on analysis of adjacency pairs, response rate to the requests are poor, which aligns with the player's comments about their experience of using the message system.

6.4.2 *Responding to directives from HQ*

Directives index the instances of remote coordination of field responders by HQ. The observed response to messages is critical to understanding the relationships between local and remote coordination.

Here, we examine how field responders deal with messages from HQ that attempt to allocate tasks and manage task execution (i.e., directives). Classification of messages showed that directives were exclusively sent by HQ, and that they were the most frequent kind of message (Table 4). Overall, out of the 43 directives HQ sent for task allocation, the recipient field responders brought up only 15 messages in conversation in the team (Figure 20). The instances in which task allocation messages were addressed reveal the handling and value of HQ directives in the local coordination. Firstly, out of the 15 task allocation messages responders talked about, they decided to ignore the instructions only once. The responders ignored instructions because they were engaged in another task that they did not want to abandon. Secondly, four HQ instructions to rescue a certain target co-

incided with the same plan that had already been made locally by the responders. In 10 cases, field responders chose to follow the instructions. However, due to confusion and misunderstanding they failed to follow them correctly six times. In fact, only 2 instances of directives from the HQ led to task completion. For the remaining 14 saved targets, field responders had locally allocated the tasks without HQ.

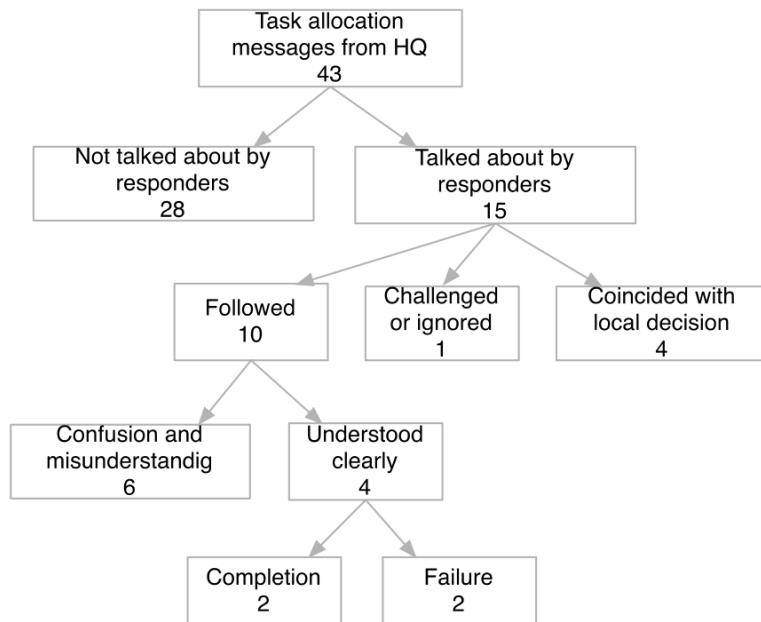


Figure 20: How responders addressed task allocation messages from HQ.

The following episode depicts a team of three on their way to pick up fuel. Their path is blocked by radiation. Without a team, firefighter JH (on the left, Figure 21) has just joined soldier KY (on the right), and firefighter D2 who have just been allocated a task in a message by HQ.

Episode 1.1

KY: ((reading out HQ message)) KY and D2, please walk fast to the junction and quickly return back ((laughs))

D2: Oh is that what we have to do? Ok so we have to run to (2.0) We need to work out where we have to run to first and then get (.) get it back. Which junction is that? If you run to the next (0.5) thing ((points)), and then come back (1.0) that would work (1.0) is it safer to go around?



Figure 21: Episode 1.1, JH (Behind Left), D2 (Middle Front), KY (Right behind)



Figure 22: Episode 1.1, KY (Left) , MF (Right) holding mobile phones

[The team tries to go around the cloud but is stopped by radiation, realising their target is in the cloud. Meanwhile, D2 has left due to increased exposure.]

KY: So we have to run! [through the radiation]

JH: Do we have to run through the (.) through the radiation? ((looking at map)) (Figure 22)

KY: Yah this is what the headquarters told us to do ((looking at messages))

JH: I have a terrible feeling thats gonna kill us.

KY: But its gonna be meaningful ((laughs))

JH: We go around this corner, if it gets to half [referring to

health] we should probably start running back.

[*KY JH begin running into the cloud*] (Figure 23)



Figure 23: Episode 1.1, KY,JH running into cloud

Figure 24: Episode 1.1, KY, JH turning around to escape from the cloud

KY: ((yells)) OH OH! It's a hundred! [refers to radiation level]

JH: We are basically in the middle of it! We are basically in the middle of it!

KY: ((shouts)) I'm going back? Get the fuel first! Get the fuel first! Oh no!

JH: We are not prepared for that! I blame our HQ.

[*They turn around and run back out of the cloud without the fuel.*] (Figure 24)

This episode begins with a message by HQ attempting to help give directions to the target. D2's response to the message is hesitant ('is that what we should do?'). His following question ('which junction is that?') suggests the referent in HQ's message is not understood. They attempt to go around the radiation. They realise their target is in the cloud. They refer back to the message to support their intent to go into the cloud to attempt to save the target ('Yah this is what the headquarters told us to do'). Having run into the cloud, they refer to the Geiger counter and realise the exposure is too high. Meanwhile, their health is decreasing rapidly. They abandon the task and flee to safety, whilst JH expresses his frustration ('We are not prepared for that. I blame our HQ.').

First, the episode shows that geospatial referencing in messages can be problematic. It is unclear to the responders which junction HQ is referencing (and the responders do not ask for clarification), so

they revise the route themselves. At the same time, they draw on the messages to justify their entering of the cloud. It does not occur to the responders that HQ allocated the task at an earlier time, before the cloud had covered the target. HQ does not update the responders on the increased danger, or revise their earlier task allocation. When the responder team fails to complete the task, they place blame instead of thinking self-critically.

6.4.3 Local coordination without HQ

As presented, field responders predominantly coordinated teaming and task allocation without HQ instructions. Recall that 14 targets are saved without HQ's instructions, versus only 2 targets that are saved with HQ's instructions. The following episode illustrates how field responders achieve coordination of teaming and task allocation locally. We join the action as BR and another responder are waiting at the drop-off zone without a compatible teammate, as MF and his teammate join and drop-off their target.



Figure 25: Episode 1.2, MF (right),
BR (middle)



Figure 26: Episode 1.2, MF (right),
BR (left)

Episode 1.2

[MF (on the right, Figure 25) and teammate walking towards BR (center)]

BR: Any soldiers?

MF: I am soldier yeah.

BR: Would you like to pair with me? (2.0) to rescue a fuel?

MF: what are you after?

BR: I am a firefighter.

MF: Soldier and firefighter is fuel isn't it?

BR: yeah.

MF: What can we get? (2.0) ((looks at screen)) this one in the center? ((points at screen))

BR: ((glances MF's screen)) I think there are two people (the team D2,KY) going for that. I think we should go for this one ((points at screen)).

MF: We are going to get killed ((both laugh)).

[The team begins walking to target.]

At the beginning of the episode, MF met BR, who was waiting at the drop-off zone without a compatible teammate. BR requested to team up with MF ("Would you like to pair with me? (2.0) to rescue a fuel?") after MF identified himself as a soldier. BR and MF can then be observed sharing the screen of his device and using the map to identify potential targets (fuels) (Figure ??). They realise one of fuel targets is already being pursued by another team. They agree on another target fuel to pursue. Note that messages do not play a role in this episode. It exemplifies how teaming and task allocation are achieved locally, without consulting HQ.

The next episode is a follow-up of episode 1.2, which demonstrates how two teams resolve the conflict when they approach a same target.

Episode 1.3

The team MF BR met team KY D2 on their way to target.

D2: we are told to get this fuel (target 1) from HQ.((pointing to screen))

MF: you are going to the fuel (target 1) we are aiming for, we thought you are going for this one (target 2).

D2: we were, until we got a message [from HQ] saying not to.

MF: you get that one (target 1), we get that one (target 2).((pointing to the two target locations))

D2: if you want get that one (target 2). It is somewhere in the



Figure 27: Episode 1.3, MF (right), BR (left) met with D2 (middle)

building.

[The two teams split, proceed with the new target allocations]

At the beginning of this episode, the team (BR, MF) has decided to pursue (Episode 1.2) a target other than the one pursued by team D2, KY. However, instructed by HQ, MF and KY changed their target and met BR and MF on the way. The two teams then began to show their intended targets to each other. After they find they are heading to the same target, MF suggested a new allocation of tasks (“you get that one, we get that one.”). D2 then offered some information about the target location to the team MF, BR (“if you want get that one. It is somewhere in the building.”), suggesting he agreed with the new task allocations proposed by MF.

The two previous episodes show how teaming and task allocation are achieved in a seemingly “ad-hoc” manner. By using the word ‘ad hoc’, we stress the actions of field responders are typically not planned ahead to great extent. The players often exchange information through conversations when co-located, and their ‘plan’ is ready to be changed when new information is acquired. Take episode 1.2 as an example, while BR was waiting at the drop-off zone, she requested to team up with MF who happened to pass by. MF agreed to team up and then decided to aim for an available target that had not been

aimed at by others. In episode 1.3, the two teams quickly came up with new task allocations when they found they were actually heading to the same targets. In the interview at the end of study, field responders also confirmed their “ad-hoc” behaviour in the interview:

“Just save the closest target then just pair up and go to the other one”

“We just check, with that group, which target we can get. We see on the map to find the closest one we can get.”

6.4.4 *Remote messages as a resource of situational awareness*

In the AtomicOrchid game, field responders need to be aware of what other responders are doing, where the ‘danger zone’ is (the cloud), and where it is likely to move. Awareness of each other’s actions helps responders avoid conflicts in planning, while awareness of the danger zone is essential to survive. The following episode illustrates how responders use remote messages as a resource to gain situational awareness.

The episode takes place towards the end of game session B. The radioactive cloud has grown so much that navigation in the game area becomes increasingly difficult. MF is with a group of five responders, two of which are carrying an animal. The cloud is blocking their way towards the drop off zone; they stop.

Episode 1.4

MF: ((reads message from HQ out loud)) There is another leak around Geospatial. (1.0) Which is Ah: so there’s a leak sprung up there. ((points)) Geospatial is like (.) that building right there. They say there is another leak. We should go all the way round (0.5) to the top left one, I think. (Figure 28)



Figure 28: Episode 1.4, MF pointing at a building

MF brings up HQ's message about the new leak, and suggests a route around the new cloud. The group ends up following MF's route suggestion as a result. News of the new cloud, provided by HQ, enables the group to change their route to avoid danger. We commonly observed responders sharing information that provides situational awareness through face-to-face conversation. In the previous example, MF shared the message with a group of responders he was with already. The following example takes place between D2 and his teammate, as they are approached by JH, who is currently without a teammate.

Episode 1.5

JH: Where are you guys heading?

D2: To get the fuel.

JH: Okay. The closest one to you?

D2: I believe so.

JH: Ya okay cuz I think the leak is somewhere near the other one and the army. [referring to building of Territory Army] (Figure 29)

D2: Oh (.) which one?

JH: They sent a message saying its between territorial army center.

D2: We are trying to get the one here ((points)).



Figure 29: Episode 1.5, JH (left) met team D2 (male, middle) and KY (female, right)

JH: The closest one. Okay.

Making use of the map as he approaches them, JH asks the others to clarify which fuel they intend to pursue ('the closest one to you?'). He proceeds to inform the team that the "leak is somewhere near the other one". D2's response ('Oh, which one?') suggests they did not know this. In turn, JH elaborates on the location of the cloud, using an anonymous "they" to refer to the source of his information. "They" is likely to refer to HQ as they previously sent a message with the information of the cloud's location. Conversational sharing of important information was a common resource responders employed to achieve and maintain situational awareness. However, requests for information in the messages channel were regularly not reciprocated with a response: out of 14 requests in session A, 8 were not responded to; and in session B, 14 out of 20 requests were not responded to (Table 6).

6.5 DISCUSSION

This section will present broader concerns that emerged from the game for the design of automated planning support.

6.5.1 *Division of labour*

Firstly, the HQ plays an important role in providing situational awareness to the whole team. As the game mechanic provides the HQ exclusive access to the location of the radioactivity, the HQ managed to provide informational messages about the radioactivity to field players. Field players are able to pick up the information and spread it to other field responders through face-to-face conversations (Episode 1.2). Although HQ attempted to organise task allocations directly (through directives), their attempts are often problematic. Although the field responders did not get too much planning support from HQ, they naturally organise themselves into small teams and carry out tasks. As shown in episode 1.4 and 1.5, face-to-face conversation was vital for task and team organisation. We observed that co-located team members collectively make sense of the remote messages and game status shown on mobile screens. The decisions such as choices of team, targets and routes are predominately made through local conversations.

The pattern of division of labour between field responders and HQ indicates the weak role of HQ in terms of task planning. The responder's choices of teams and targets seem to follow an "ad-hoc" manner as they heavily rely on face-to-face conversation, which can only happen when players are co-located. Despite some disruptions from the communication channel (confusions of geo-referencing, and outdated HQ messages), field players seem to be able to find team-mates and generally avoid conflicts in their plans (e.g. avoid pursuing the same target) through local coordination.

6.5.2 Breakdown of remote coordination

The observed division of labour (Section 6.5.1) highlights heavy reliance on local coordination. To some extent, the heavy reliance on “ad-hoc” local coordination can partly be a result of the lack of remote coordination support. In other words, local coordination becomes important when system support for remote coordination is problematic. In a co-located setting, players can naturally make their actions observable and accountable to each other through conversations, body languages, gestures , screen sharing etc, and organise coordination activities reflexively. However, in the remote setting, the natural accountability of their activities become opaque. The game probe provides a set of functionalities supporting remote coordination, including GPS/map sharing, broadcasting. We have observed players utilise these functionalities to make sense of other team members’ actions (Section 6.4.3) and act accordingly. However, coordination with remote players is still problematic which can be evidenced by the lack of response and acknowledgements to requests in the messaging channel (Section 6.4.1.4); and the misunderstanding and confusions observed when field responders try to follow the directives from HQ (Section 6.4.2).

We suggest that future planning support should properly support remote coordination in a way that facilitates accountability among distributed team members. Section 6.6 discuss some detailed requirements of remote coordination support drawn from the field observations. The next section in particular, expands on the design requirements that can enhance remote coordination in a way that supports natural accountability of human activities.

6.5.3 Implications on computational support

To some extent, the observation of “ad-hoc” local coordination is aligned with the view of situated actions. As a whole team, play-

ers form and disband teams without team-wide plans prior to their actions. Players are observed to have conversations about their status, on-going activities when they meet up. Local decisions for next moves are often made during the conversations. Information from mobile interface (e.g. player locations, radiation readings and messages) are often brought to conversations as resources in the context of their situated actions.

The lack of plans may indicate the lack of optimisation of team task allocation. The multi-agent coordination algorithms (e.g. [Ramchurn et al., 2010]) aims to support the team by producing computationally optimised plans for responder teams. However, design of the planning support may not be straightforward.

Firstly, there is a danger of imposing an inappropriate “work model” (assumed by agent support) on the human team. Studies of CSCW systems [Bowers et al., 1994] raise a concern that the work model held by the technological system sometimes comes into tension with the natural human workflow achieved through methods internal to the work. In particular, current division of labour between HQ and field responder suggest that HQ plays a supportive role (providing situational awareness). However, a centralized coordination algorithm may need to coordinate the whole team, requiring every player to follow top-down instructions to reach a global optimum of resource allocation. In that case, the role of the control room may need to change and it is unknown whether the change will disrupt or support human workflow.

Apart from disruption of human workflow, there is also a danger of the supporting system imposing a planning model on human teams. Suchman [1987] suggests that human’s situated action should not be simply treated as a inferior version of scientific planning model. Following the view of situated actions, supporting human’s actions are not as simple as providing an optimised plan to execute. A plan for human teams is only one of the resources that humans can utilize for

their deliberation on their actions. Therefore, the role for the system is to provide plans (as an extra resource) in a way that supports the human's situated actions.

Adopting the view of human agent interaction, we can treat a plan support system as a teamwork agent. Achieving mutual intelligibility between the agent and the human team can also be a major design challenge. Plans generated by a planning agent can only represent a representation of possible actions and effects based on simulations. How the responders make use of plans can be highly dependent on situations in which mutual intelligibility plays an important role. For example, it would be problematic if globally optimised choices conflict with the "ad-hoc" choices that are obvious for human field responders. We cannot assume either the agent or human choice will be always correct, perhaps neither of them can take an authoritative role. Therefore, we need to carefully design the interaction between human and agent to ensure that they maintain mutual intelligibility so that informed collective decisions can be reached.

6.6 DESIGN REQUIREMENTS

Drawing on the problems observed in remote coordination (Section 6.5.2), we now discuss the detailed design requirements that enhance remote coordination. The embodied game probe embedded responders in a challenging setting. They needed to communicate effectively to make time critical decisions on teaming and task allocation, both locally in the field as well as remotely through messaging. Field responders physically engage and navigate the environment to perform tasks while maintaining awareness of risk and danger. The data reveals multiple challenges for team coordination involving communication and decision-making.

Sharing of local decision-making. The study showed that teaming and task allocation were predominantly organised locally among

field responders, in an “ad-hoc” manner. Despite the fact that HQ attempted to coordinate task allocation remotely, few of these directives were brought to conversation locally. Only 2 out of 16 tasks that field responders completed were remotely allocated by HQ (Figure 20). Although players are able to smoothly conduct local coordination, local coordination heavily relies on and is limited by the face-to-face conversations, which means some conflicts of planning can only be found and resolved when players meet each other (e.g. episode 1.3). Therefore, local decision-making can benefit from a shared picture of team-wide planning decisions. We thus argue that local decision-making needs to integrate capabilities to enable team-wide sharing of the local decisions.

Coordinate resources. While field responders made decisions on teaming and task allocation in a seemingly ad hoc fashion, game data reveals how field responders draw on resources to achieve situational awareness in order to coordinate successfully. A common understanding of the location and movement of the radiation cloud was achieved by sharing information from game messages verbally in a local group. Face-to-face talk was an essential resource for relaying information from the Mobile Responder App to teammates, such as radioactive exposure, others’ whereabouts, task status, and other monitoring of the broadcast messages. Future planning support systems need to take into account that such coordinate resources are likely to be comprised of digital as well as embodied human resources.

Geospatial referencing The results show that geospatial referencing was problematic in various ways, particularly in directive messages sent to the field players. Participants had different levels of knowledge of the campus, which made understanding of landmarks references uncertain. Some participants also struggled with making sense of north/south/east/west directions in relation to their current position and orientation. To deal with misunderstandings, players had to ask for clarification via messages or spend valuable time discussing the reference locally in order to understand it. Consistent

with the findings of Troups et al. [2009], designers need to think carefully about how the presentation layer of such systems may be augmented with information that facilitates geospatial referencing (e.g., grids, labelling etc.) to facilitate human in addition to machine readability.

Freshness of messages As demonstrated in one of the episodes (1.1), reading out-dated messages in a dynamically changing environment can contribute to responders taking dangerous actions that they believe to be safe, because they do not realise that the information is out-dated. However, in most cases, recipients managed to identify temporally irrelevant messages, and thus avoided following them. To reduce confusion about message freshness, such systems should address these issues at the UI level, both for responders and for HQ. Develop functionality to flag messages as out-dated, to retract incorrect messages, or highlighting up-to-date messages.

Acknowledgement of messages In most cases, field responders did not acknowledge or respond to messages sent by the HQ. This was particularly problematic for directives (task allocation), as task status and field responder compliance often had to be inferred by observing their location updates on the map. This consumed HQ attention, with negative impact on HQ's overall work on state assessment and task planning. Observations in the field suggest that the physical demands (e.g., co-located team movement through terrain at speed) and cognitive demands to maintain situational awareness (e.g., monitoring of radioactivity and messages) are likely factors that explain lack of acknowledgement.

As a result, user interfaces that enable and encourage field responders to quickly acknowledge HQ messages, with minimum cognitive load, should be considered for messaging in such systems in such high demand settings. For effective team coordination in Disaster Response, interface and workflow designs need to factor in cognitive

load and task demands for effective information distribution.

6.7 SUMMARY

The objective of this study is to unpack how human teams coordinate in the time and space constrained task setting. In particular, we focussed on a scenario in which responders coordinate role-based teaming and spatially distributed task allocation and execution using a real-time location and messaging system, with no automated planning support.

We presented the design and study of the AtomicOrchid game as a mixed-reality game probe to investigate challenges for team coordination in a setting in which participants experience both physical strain through bodily activity, and cognitive challenge through time pressure and task complexity. We use mixed-reality game probes as a platform for investigation of concomitant socio-technical issues: handling of mobile devices to communicate and maintain situational awareness (messaging, sensing, interaction, and display) intersect with face-to-face interaction, whilst the physio-cognitive challenges created through game mechanics and environment induce stress. We created a setting that allows exploring requirements to support team coordination of relevance to time-critical coordination domains such as real Disaster Response.

Findings from interaction analysis of field observations, triangulated with log files, reveal how field responders achieved coordination by drawing on local face-to-face conversation with fellow responders, and situational information provided by the interactive map, the Geiger counter, and the messages sent by HQ. Drawing on these findings, we highlight challenges of introducing automated planning support into the socio-technical setting of AO. We also generate requirements for supporting team coordination, emphasising the roles of local coordination, decision-making resources, geospatial referenc-

ing and message handling. Most of the requirements are fed into the refinements of the AO system in later studies, which help take into account the non-agent related design factors.

ATOMICORCHID STUDY 2: THE HUMAN ON-THE-LOOP DESIGN

This chapter presents the second iteration of AtomicOrchid field trials. In this study, a planning agent is integrated into the system by following a human on-the-loop interaction design, in which the HQs only monitor the planning agent and occasionally intervene. The purpose of the trials is to investigate socio-technical issues related to human system interaction. Interaction analysis is conducted to examine log data and field observations, revealing how human system interaction plays out, in turn, revealing the process by which players interpret and negotiate the agent's guidance as well as how these are intertwined with social dynamics of the teams.

7.1 INTRODUCTION

In order to support effective Human agent collaboration, two patterns of interaction design (human on-the-loop and in-the-loop) are presented in section 4.4. In this study, the planning agent is integrated into the AtomicOrchid system with a straightforward human on-the-loop design (Figure 30). This design assumes minimal Headquarters intervention, that is, the agent can directly interact with field responders to generate and implement plans without constant involvement of HQ. The interaction design is aimed to facilitate a pattern of division of labour, in which a planning agent routinely assigns tasks to distributed responder teams, while human coordinators (the HQ) monitor and support the task execution by responding to arising contingencies. The agent is designed in a way to take into account simple human feedback, i.e., a field responder can either reject or accept their task assignment. The agent will consider the feedback for the next it-

eration of task assignment.

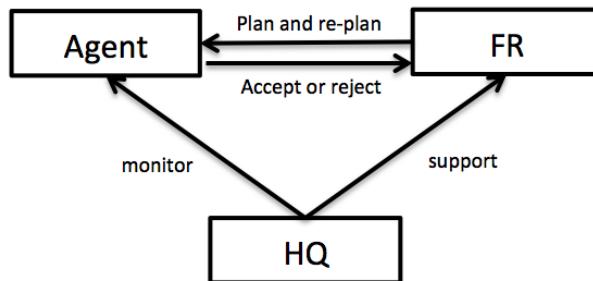


Figure 30: On-the-loop interaction design

This study uses the agent integrated version of AtomicOrchid as a probe to unfold socio-technical issues in human-agent interaction, with focus on the implications of the human on-the-loop design on human team performance. More specifically, this chapter addresses the following research questions on how agent guidance affects the social organisation of team performance:

1. How do human teams respond to being instructed by an agent, particularly on switching teams and tasks?
2. The planning agent makes decisions based on limited assumptions about human behaviour, but what are the hidden costs of human behaviour that the agent does not take into account?

Findings from the study highlight the social processes in which members interpret, negotiate, and manage the agent guidance within the social dynamics of teams. We discuss the division of labour between humans and teams and the hidden costs of instructions that suggest team reformation and interrupt on-going tasks. We conclude this chapter with a number of emerging interaction design recommendations to consider when building planning support systems for human teams, which emphasises the need for common ground between humans and the agent, facilitate accountability between team

members, and balance responsibilities between humans and the planning agent appropriately.

7.2 SYSTEM EVOLUTION

Compared to study 1 (Chapter 6), the system has evolved to provide agent planning support with an on-the-loop interaction pattern. This section gives a description of the changes of the system, which covers integration of a planning agent, implementation of a quick feedback system, and improvement in both HQ and mobile interface.

7.2.1 *The planning agent*

One major change of the system is the integration of a planning agent into the AtomicOrchid platform. The planning agent is developed by ORCHID research partner Wu Feng, Sarvapali Ramchurn, and detailed in section 4.3.3.

The planner, implemented in Java, is wrapped in a PHP server framework and deployed on an independent server separate from AtomicOrchid. The planner server exposes a HTTP interface for AtomicOrchid to request plan. Each plan request issued by AtomicOrchid is appended with updated game status, which includes players' health, distribution of radioactive cloud and locations of players, and targets. Based on the updated game status, the agent will produce an optimised task allocation and return it to AtomicOrchid. The plan requests are triggered frequently in game sessions so that the task allocation can be frequently adjusted according to task execution status. In this study, plan requests (and thus re-planning) is triggered by two kinds of game events:

1. Completion of task. On successful rescue of a target, a new plan (i.e., allocation of tasks to each responder) is requested from the

agent.

2. Explicit reject. On rejection of a task allocation by any of the first responders, a new plan is requested. The feature of rejection is part of a feedback loop between human and agent, and will be introduced in next section.

7.2.2 *A feedback loop*

The feedback system is part of the on-the-loop interaction design, which enables the agent to take into account simple human feedback. It is also partly inspired by a requirement generated in study 1, which highlights the importance of quick acknowledgements from field responders. The feedback system can be seen as a system mechanism for the field team to provide quick responses to the planning agent. This section goes through the implementation details of the feedback loop.

Once a plan is received from the agent, the AtomicOrchid game engine splits the plan for a given team into individual task allocations and sends these to each responder's mobile app (Figure 31). The app displays the task allocation in a pop-up and details it in the new task tab, including: i) the responder to team up with, ii) the allocated target (using target id), and iii) the approximate direction of the target (e.g., north, east) (Figure 32).

On receiving an instruction from agent, the field responder can choose to either reject or accept the instruction (Figure 32). In the case of rejection, a new plan will be requested and the agent will consider the feedback for the next iteration of task assignment. More importantly, the rejected allocation is used as a constraint within the optimisation run by the planner agent. For example, if two responders (a medic and a soldier) were allocated a task and the soldier

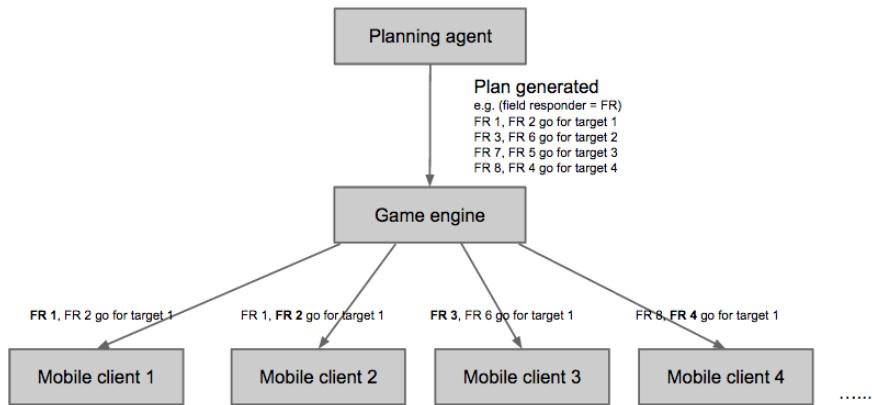


Figure 31: Game engine handling plans from agent

rejected it, the planning agent would return a new task allocation with the constraint that this soldier should not be allocated this task. The instructions sent to field responders are also displayed in the HQ interface for monitoring purposes. The task allocations are represented as yellow lines connecting players and their targets (Figure 33). Only one task allocation is displayed at one time, HQ player can click on the ‘show’ task button on the player status panel (top right) is to chose whose task to be shown.

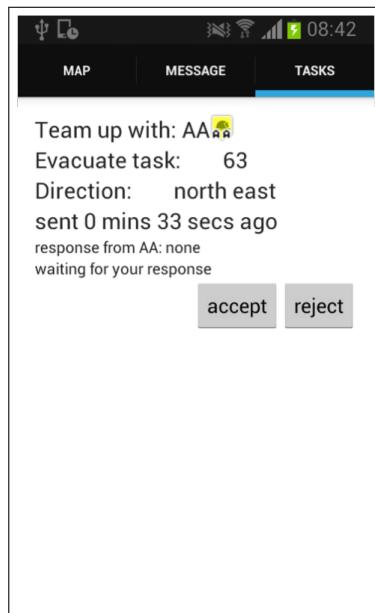


Figure 32: Mobile task interface in study 2



Figure 33: HQ and mobile interfaces in study 2

7.2.3 Interface improvements

Apart from the agent integration and feedback system, two small modification of the interface were inspired by the requirements generated in the previous study (Section 6.6). Firstly, all icons of targets are now marked by a unique target number for HQ and field responders to cross-reference. Secondly, all the messages are labelled by timestamps for players to help identify outdated messages (Figure 34).

7.3 STUDY DESIGN

Study participants were recruited through posters and emails. A total of 18 participants were recruited for 2 sessions of Game play (Appendix C.2). For each session, there was 1 HQ player and 8 field players. All participants were reimbursed with 15 pounds for 1.5 hours of study. The majority of participants were researchers and students of the local university. The procedure of study is the same as the one described in study 1 (Section 6.3)

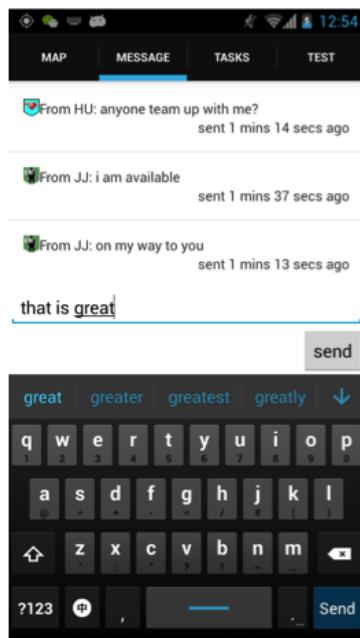


Figure 34: Mobile message interface in study 2

The game area was the same as for study 1 on the local university campus (Appendix C.1), 400 by 400 meters, without heavy traffic. The terrain of the game area includes grassland, a lake, buildings, roads, and footpaths and lawns. There are two drop off zones and 16 targets. There were four targets for each of the four target types. The pattern of cloud movement and expansion was the same for both game sessions, and it is the same as for study 1.

7.4 DATA ANALYSIS AND RESULTS

Here, we present findings from interaction analysis that reveal how team coordination was achieved. Overall, responders rescued 12 and 11 targets in sessions A and B respectively, out of 16 targets in total per session. No player was incapacitated in the two sessions. The planning agent sent a total of 51 instructions (Figure 35), 24 of which are accepted and 11 of them are rejected. The remaining 16 instructions do not receive complete responses (at least one of the player did not reply). A total of 21 instructions were finished successfully, versus only 2 of the targets are saved without agent instructions. Of the the 19 instructions that were unsuccessful, some were ignored or violated

by players and some were overridden by the agent in the replanning process due to change of circumstance.

| | Saved targets | Incapacitated players | Average Health |
|-----------|----------------|-----------------------|----------------|
| Session A | 12 (out of 16) | 0 | 80/100 |
| Session B | 11 (out of 16) | 0 | 82/100 |

Table 7: Overview of game results

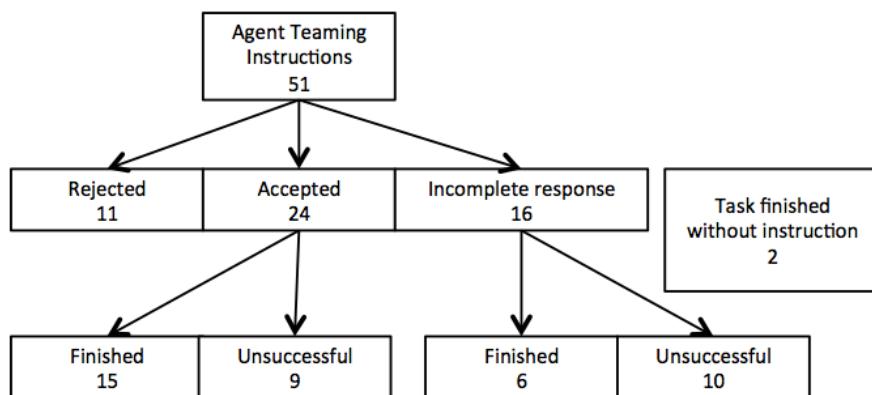


Figure 35: Overview of agent instructions

In what follows, we presents selected episodes of game play to reveal how teams accomplish the tasks in the rescue mission, particularly focusing on the social organisation of interaction with and around the agent instructions (Appendix C.3). The order in which we present the episodes follows the common practice of moving from exhibits of typical/unproblematic instances, via more complex/difficult instances, to exhibits that display problematic interaction or even complete breakdowns. The episodes in later sections will be presented with a standard orthographic notation detailed in section 5.2.

7.4.1 Assigning task assignments to existing teams

The following episode depicts a team of two dropping off a target and planning the next step.

Episode 2.1

[0:00] *The team dropped off a target.*

PC: I think we dropped off now. Ok.

[0:07] *The team receives a new agent instruction: PC, CR -> 22*

PC: I have a task now (3.0) ((studying screen)), I need to go with CR to 22. Are you CR? (Figure 36)

CR: Yes.

PC: go 22.

CR: We have done 22.

PC: Oh (1.0), no (2.0) 22 is there ((pointing to direction of 22)), Let's go ((PC leads the way, they start walking to 22))

PC: Right this way. (Figure 37)

[0:28] *The team finishes the task assigned by the agent.*



Figure 36: Episode 2.1, CR (Left) and PC (Right) studying screen together after drop off.



Figure 37: Episode 2.1, PC (Right) leading the way to new target.

At the beginning of this episode, the team (PC, CR) drops off a target at a drop off zone. Player PC vocalises that they have finished

the task (“I think we dropped off now. OK”). After about 7 seconds, PC says she received a new task allocation from the agent (“I have a task now”). PC confirms the initials of the other player (CR), and suggests CR to join her to go for target 22. The action is consistent with the agent instruction (PC, CR -> 22), suggesting that PC has read through the instruction and decided to follow it. CR said that they have already finished target 22 (“We have done 22”), which indicates he is confused about the current task allocation. PC resolves the confusion by pointing in the direction of 22 and repeating to go for it. Later, the team successfully drop off target 22 as instructed by the agent.

The episode shows how an agent instruction is brought up and followed by a team in relative straightforward manner. The instruction was delivered immediately after the drop off of a previous target (7 seconds after). PC successfully locates the new target in the instruction and leads the team to pick it up. Although CR is confused at first, PC manages to rectify CR’s mistake and they finish the task successfully.

This episode is a typical case of task assignment to existing teams, i.e. the agent sent a new task to a team immediately after they finished their previous task. Out of a total of 51 agent instructions, 23 fall into this category. The rate of compliance is high for these cases of task assignment to existing teams (21 out of 23; 91%).

7.4.2 Task assignments involving team reformation

Unlike episode 2.1, sometimes the agent instruction implies players need to disband and form new teams after finishing their previous task, in order to enact the computationally optimal plan. 10 out of 51 agent instructions fall into this category (Table 8). The compliance rate of instructions that require reteaming (50 percent) is substantially lower than compliance of instructions where players can stay in the

same teams (91 percent). The following episode depicts a typical case in which team reformation fails.



Figure 38: Episode 2.2, players from left to right: LT, SS, CR, PC. LT walking around the team, her body orientation suggesting attempts to leave the group.

Episode 2.2

[0:00] After a target drop off, LT and SS joined PC and CR at drop off zone.

[0:24] HQ sent message A: LT, if you think you have the stamina to run to 10 around the north of the lake do so now with a firefighter.

[0:28] Agent instruction received: NK, LT -> 16

LT: They said ((reads out aloud HQ message A))

[0:35] CR ((facing LT)): Shall we go get 10

LT: Mine is 16.

[0:38] HQ sent message B: Avoid 17 at all costs (...) I'd avoid 10, too.

CR: ((read out HQ message B)) avoid 10 now.

[0:55] New agent instruction received: NW, LT -> 15

LT: 15!

[Fig. 3] LT keeps walking and turning back and forth from others.

PC and SS discuss next steps, LT does not engage in the discussion with them.

[1:12] SS ((facing PC)): Shall we go get 19? ((turning towards LC and CR)) are you going to 10 or something?

CR: Eh::, HQ said no. [referring to message B]

[1:24] SS and PC decide to go for target 19, and leave.

[1:29] NW sent message: LT where you

CR: ((facing LC)) Are you LT?

LT: Yes.

CR: NW is looking for you.

LT: Yah thanks. ((turning away from CR)) Ah... I will go towards them. ((starts walking))

CR: Okay. Do you want company?

LT: ((turning back towards CR)) Yeah.

CR and LT leave drop off zone together to find NW.

The episode begins with a recommendation by HQ to LT to go for 10 (message A). The message is topicalised by LT, but it is soon overridden by an agent instruction (NK, LT -> 16). When CR proposes to team up with LT to go for target 10, LT declined ("mine is 16"). HQ then withdraws its previous suggestion to go for 10 in message B. Shortly after; a new instruction (NW, LT-> 15) prompts LT to read out the target number (15), but she fails to raise the other players attention. While other group members engaged in planning next steps, LT does not engage and keeps looking around. She can be seen turning and walking back and forth (Figure 38). Perhaps LT is trying to locate the player NW who she had been instructed to team up with. LT does not take any action until prompted by CR ("are you LT? NW is looking for you"). Then, LT begins to walk to find her teammate. However, when she finally manages to meet up with NW two minutes later, NW has already been assigned another task.

On the one hand, LT seems to feel obliged to follow the agent instructions. She turns down other teaming invitations and appears to try to look for NW in her immediate vicinity, indicating difficulty with locating teammates out of sight (despite the real-time location map). On the other hand, her body orientation displays a sense of attachment to the existing group. Her indecisive walking and turning back and forth suggests she is struggling to leave. She does not leave the group to follow the instructions until prompted by some-

one. When CR points out NW's messages, LT does not answer the message either. The episode illustrates a combination of interactional troubles as a result of which the reteaming fails: being attached to the local group, struggling to locate teammates out of sight, and failing to reciprocate messages.

Further, we found the distance between instructed players to be a key factor in successful reteaming. That is to say, if instructed players are not within line of sight, the rate of non-compliance with the agent instruction is high. Take episode 2.2 as an example, player LT was instructed to team up with a distant player twice. Neither one of the instructions was successfully implemented. Overall, there were 17 agent instructions that implied teaming with distant players; only 1 of them were actually followed by players. Players explicitly rejected 11 of them by pressing the rejection button; the other 5 were not followed without an interface action (i.e. neither accepted nor rejected).

7.4.3 Task assignments involving task interruption

In some other cases, the agent also sent new instructions to teams that had already commenced their task; that is, teams were interrupted by the new instructions. The following two episodes 2.3 and 2.4 describe how players handled task interruptions caused by the agent.



Figure 39: Episode 2.3, AW (right) leads the way, heading to target 44 as instructed.

Episode 2.3

[00:00] HB, AW at drop-off zone, new instruction received: AW,



Figure 40: Episode 2.3, After the team received an instruction to disband, AW (right) and HB (left) simultaneously turn back and start walking back to the drop off zone, displaying bodily alignment.

HB->44

HB: Alright, who is AW?

AW: Me.

HB: let's go southeast (the direction of target 44).

[00:07] *AW, HB looking at their screens.*

[00:26] **HB:** There is no 44.

AW: down there.

HB: Ok, yea, yea, yea (0.5), I can't see, Oh, there, yea, let's go.

[00:35] [Fig.4] *Team begins moving towards 44.*

[00:48] HQ sent message: Target 42 and 44 is not reachable.

AW: ((reads out the message))

AW and HB stopped walking.

[00:52] *New instructions received: AW, KD -> 44, HB, AR->31*

AW: I got a new instruction.

[Figure 39] *AW and HB simultaneously turn and start walking back towards the drop off zone.*

HB: I need to team up with AR.

AW: I need to team up with KD! Oh, it is 44 again.

[01:01] *AW, HB arrived at drop off zone, met AR, KD.*

HB: AR?

KD: AW? We have got (1.0), 44, right?

AW: It said 44 is not reachable, but I got it again, so, let's try.

KD: Alright.

[01:14] AW, KD begin walking to 44, AR, HB team up as well.

This episode begins with an instruction (AW, HB -> 44) from the agent. At that moment, there were 5 players at the drop off zone (AR, KD, LC, HB, AW). Immediately after the instruction, HB starts looking for AW in the local group. Shortly after, AR and HB team up to go for 44 as instructed. However, 13 seconds later the team is interrupted with a HQ message telling them not to go for 44 (Target 42 and 44 is not reachable). Four seconds later, a conflicting agent instruction was delivered, implying they disband the team (AW, KD -> 44, HB, AR->31) but still pursue the target 44. At first, AW stops walking and topicalises the instruction ("I got a new instruction"), followed by both teammates simultaneously turning towards each other (Figure 39, 40). The bodily alignment in the action suggests agreement to follow the new instruction. On their way back to drop off zone, HB and AW confirm their intentions ("I need to team up with AR", "I need to team up with KD!"). In this case, the teammates respond to the interruption by mutually agreeing to abandon the current team and task in favour of following the new assignment.

It should be noted that the interruption was received only 17 seconds after the team commenced the task, probably contributing to a low perceived cost of abandoning the current task. Further, all players involved in the subsequent reteaming were not far away from each other. AW and HB had not walked far from the drop off zone; so everyone was still within line of sight, further facilitating successful reformation.

7.4.4 Disagreement on task interruption

Episode 2.4

[Following on from Episode 2.3]

AW, KD on their way to target 44.

[01:39] New instruction received again, AW, HB -> 44, AR, KD

->31

AW: new instruction, HB and 44 again, haha.

AW turns back towards drop off zone immediately.

KD: AR and 31 ((Reading his new instruction)) ehh, have they gone?

Because we can just decline and carry on.

AW: Ok, I rejected it.

AW turns back towards KD, who also rejects the new instruction.

They resume their walk to 44.

[01:54] *New instruction delivered to AW (AW, YF ->46)*

AW: new instruction 46, yeah! ((team stop walking))

KD: Do they know we are already on the task?

[02:00] *New instruction delivered to AW (AW, LC ->37)*

AW: yea, but I think, Oh, no, got new instruction again, (team up with) LC.

[02:13] *AW starts walking to LC, who is at drop off zone within line of sight, leaving behind KD.*

KD: ((reads out HQ message)) AW and KD you won't reach 44. Alright, Let's go to 46.

AW ((turning back towards KD)): I don't know, I got a new task with LC.

KD: Ahh, I do not have a task.

AW turns and walks towards LC again. KD follows.

In this fragment, we can observe disagreement and negotiation of team reformation. Following episode 2.3, player AW disbands his team with HB and teams up with KD. However, 20 seconds after the reformation, AW is instructed to abandon the on-going task again. AW laughs, but turns back to find player HB again. Before AW sets off, KD disagrees with the new instruction and proposes to reject it ("Ehh, have they gone? Because we can just decline and carry on"). AW accepts KDs suggestion and turns back to KD.

After the rejection, AW receives 2 consecutive reteaming instructions from the agent, finally teaming them up with LC, while KD does not receive another instruction. KDs question ("Do they know

we are already on the task?”) suggests that he might think the agent is unaware of their situation, and that he disagrees with disbanding the existing team. In spite of KD’s disagreement, AW declares his intention to follow the new instruction (“got new instruction again, [team up with] LC”) and he turns to find LC. However, KD ignores this (“Alright, Lets go to 46”), indicating he does not agree with AW’s intention to disband the team. AW interjects (“I dont know, I got a new task with LC”), and continues to walk towards LC, denying KD. As KD realizes he is without assignment (“Ah, I do not have a task”), he follows AW to find LC.

In this episode, teammates agree to reject the first task assignment. We found task interruption could be a major reason to reject new instructions. 10 out of 11 rejected instructions are associated with task interruption. In an extreme case (not pictured), one team reached an agreement to ignore any agent instructions after the agent tried to interrupt the teams’ on-going task.

In the end, the player that received the new instruction disagrees with his teammates suggestion to ignore the instruction and decides to leave the current team. The team is disbanded in disagreement, in contrast to episode 2.3 where both teammates agree to leave the team after both received new instructions at the same time. Here, the teammates spend a fair amount of time arguing whether to follow or ignore instructions, hinting at the hidden social cost of agent coordination algorithms when applied to human teams.

Overall, the majority of new instructions that interrupted on-going tasks required team reformation. When tasks were interrupted, the rate of compliance (22 percent) is substantially lower than when teams were required to reform after a task was completed (50 percent). Task interruptions were also much more likely to lead to rejection of the new assignment: 10 out of 11 assignments that interrupted tasks were rejected.

7.4.5 *The headquarters*

HQ sent a total of 147 messages in the two sessions. We identified 50 assertives and 68 directives in two sessions through speech act analysis 5.3.3. The majority of assertives were focused on providing situational awareness and safely routing the responders to avoid exposing them to radiation. E.g. “NK and JL approach drop off 6 by navigating via 10 and 09.” or “Radiation cloud is at the east of the National College.”

16 out of 68 directives were directly related to task allocations and teaming, which is substantially less than the number of agent instructions (51). Among the 16 directives, HQ sent 11 direct instructions to the field players (e.g. “SS and LT retrieve 09”), while the remaining 5 are related to forward planning, (e.g., “DP and SS, as soon as you can head to 20 before the radiation cloud gets there first”). 6 of the HQ instructions are consistent with agent instructions, while 5 other HQ instructions override the agent instructions. It is worth mentioning that field players implemented only 5 out of 16 HQ instructions. In the interview, HQ reported that they felt they supported the agent rather than taking control.

7.5 DISCUSSION

In the previous sections, we described how the agent guidance is interleaved with social interaction, in which teammates organise the task planning and execution. We found that while the agent supported division of labour, the agent guidance had various social implications. We now reflect on (1) how division of labour is achieved; (2) the social implications and hidden cost incurred by team reformation and task interruption; and (3) the limited feedback mechanism.

7.5.1 *Division of labour between the agent and the human teams*

Overall, players followed 30 out of 51 agent instructions, out of which 21 tasks were completed according to the instruction (success rate of 70 percent). Only 2 targets were evacuated without agent instruction (Figure 35), which indicates that, to a large extent, the agent successfully supported routine task planning activities. Episode 2.1 demonstrates a typical case of division of labour: the agent handles planning of teaming and task assignment, freeing the team to focus on other issues such as navigation (identifying the target on the interactive map and finding directions) and organising team meet up. The following of agent instructions speaks of players trust in the agents decisions. In the 30 cases where instructions were followed, we can observe similar patterns of division of labour.

The distribution of HQ messages may also indicate a division of labour between HQ and the agent. Only a small proportion (16 out of 147) is directly related to task assignment, indicating routine task allocations were delegated to the agent. A relatively large proportion (118 out of 147) of messages are used to provide situational awareness and safely routing the responders to avoid radiation exposure. However, the fact that only 5 (out of 16) HQ instructions are implemented suggests that HQ was unable to effectively override the agent when they wanted to. This fact highlights that the planning agent plays a strong role in the control loop, compared to the human coordinators in the HQ. The planning agent can directly instruct field responders without consent of the HQ, and the HQ does not have an effective way of overriding the agents decision.

7.5.2 *Hidden costs of team reformation and task interruption*

With this division of labour introduced in the previous section, it appears that system can successfully generate plans for humans to “execute”. However, Following the view of situated actions [Suchman,

1987], The plan is actually a resource in the situation that human can leverage or not. In what follows, we reveal the significance of the plan in determining humans' actions varies based on the social situations the players are in.

First, while the compliance rate with agent instructions was high when no reteaming was required (91 percent), we found that the rate of compliance with agent instructions is much lower when team reformation is involved (50 percent), and even lower when in addition an on-going task is interrupted (22 percent) (Table 8). Our interaction analysis shows the ways in which team reformation and task interruption are associated with hidden costs in the social organisation of team performance.

| Context | Instructions | Followed by FR | Compliance rate |
|---------------------------|--------------|----------------|-----------------|
| Instructing existing team | 23 | 21 | 91% |
| Require team reformation | 10 | 5 | 50% |
| Interrupting tasks | 18 | 4 | 22% |
| Total | 51 | 30 | 59% |

Table 8: Compliance with agent instructions by context

We found that team disbanding can be difficult. Players have to make their actions accountable to gracefully disengage from an existing team to avoid breaching social norms (e.g., politeness). Members have displayed a sense of attachment to a local group (episode 2.2), which delayed the task substantially until the team reformation failed. Despite interrupting an on-going task, new instructions for both teammates can facilitate smooth, mutually agreed disbanding (Episode 2.3), while instructions for only one member have coincided with interactional trouble, disagreement and delays (episode 2.4).

The impact of attachment between co-located teammates was further amplified by distance between proposed teammates. While they frequently accounted for actions with co-located players, they did not make their actions equally accountable to remote team members. For example in episode 2.4, the agent interrupted the local teams task

and instructed them to team up with distant players. The co-located team decided to reject the instruction without contacting the potential teammates they rejected. The system lacked support for accountability between remote members.

A further observation is that players were unwilling to give up on-going tasks after a certain time. In episode 2.4, the teammates first agree to ignore new instructions. This preference to stick with on-going tasks may also explain the high rejection rate for instructions involving task interruptions.

The social organisation of coordination reveals implications for the simple model of interaction held by the agent. The agents algorithm re-plans and reshuffles teams, in order to optimise group performance by minimising the travel distance to the targets. However, the study has revealed the ways in which social norms and the accountability of social conduct get in the way. This raises questions about the effectiveness of approaches that treat coalition formation of humans as unproblematic. The agent does not consider the social cost of team reformation and task interruption. The study has shown that the social process to disengage from groups and on-going tasks can be costly. The tension between the social process and the model held by the agent echoes the notion of workflow from within and without [Bowers et al., 1994]. The authors point out that models imposed by technology (from without) may come into tension with the actual workflow achieved through methods internal to the work (from within).

7.5.3 *Feedback to the agent*

To recap, a feedback mechanism was included in the interaction design to give responders some control over the task assignment (Section 7.2.2). On receiving an instruction, players can either accept or reject the instruction. On rejection of a task allocation, a new plan is requested. The rejected allocation is, in turn, used as a constraint within the optimisation run by the planning agent, which means that

the rejected target will not be assigned to the rejecting player for a time (1 minute).

Our observations show there may be a significant cost associated with rejection. Overall, 6 out of 25 re-plans were triggered by rejections. In turn, tasks were re-assigned to all players (not just for the player who rejected). Frequent new instructions may cause extra coordination overhead (time spent on interpreting new instructions, more team reformation and task interruptions) and over-constrain the planning task. Players did not seem to be aware of the implications that their rejections had on others.

We also found that players' expectations of the rejection were not always aligned with its actual effect. Instructions involving reformation and interruption were more likely to be rejected. Players' statements indicated that they perceived the rejection as a way to reverse to previous states (Episode 2.4). Other statements indicated rejections were expected to pair them with a new teammate instead of a new target. The mismatch between expected and actual effect highlights a lack of intelligibility in the current interaction design. We aimed at simplicity (by providing only accept/reject options), which might be important for interaction in time-critical task settings, but it comes at the cost of intelligibility. Therefore, we argue that intelligibility and simplicity need to be carefully balanced according to details of the setting.

7.6 DESIGN IMPLICATIONS

Our observations reveal the tension between agent planning support and the social organisation of teamwork. The tension does not simply mean that the model held by the agent is incorrect; it highlights potential trade-offs we need to consider in system design [Bowers et al., 1994; Sukthankar et al., 2009]. Providing a detailed design solution is beyond the scope of this thesis. Instead, we propose three design implications to scaffold the division of labour when building agent-

based planning support for human teams.

7.6.1 *Achieve common ground*

Two main issues arose that challenged the basis for collaboration [Bradshaw et al., 2011]. Firstly, a notion of the social cost associated with instructing teams should be taken into account when designing planning agents. For example, disbanding teams can be difficult and time-consuming as it is governed by rules of social conduct and etiquette, particularly where the new teammates are out of sight or only one of the teammates receives a new instruction. Secondly, a mismatch between the expected and actual function of rejections further shows intelligibility needs to be improved. Therefore, we suggest that the design of agent support should a) take social factors into consideration (e.g., ensuring team disbanding is facilitated by reteaming both teammates at the same time; avoiding task interruptions etc.), and that b) agent functionality is appropriately surfaced to help achieve common ground (e.g., by providing explanations of agent action at the interface level).

7.6.2 *Facilitate accountability*

While the rules of social conduct ensured accountability of action among co-located teammates, we found the impact of rejections on remote players was not properly appreciated; nor did the interaction design support making these rejections accountable. Therefore, we believe the interaction design should reveal the hidden cost of certain actions (e.g., rejections) to facilitate the accountability of local decision making accountable to remote team members, ensuring consequences of local decisions for the welfare of all teams are understood.

7.6.3 Balance responsibilities between humans and agent

The social implications and other situational contingencies are likely difficult to be modelled computationally. Alternative approaches argue for mixed-initiative control and flexible autonomy between humans and agents [Bradshaw et al., 2011]. The ways in which the HQ used messages to provide situational information that complemented the agent instructions show that humans are able to deal with arising situational contingencies. The division of labour between humans and the agent appeared generally effective in that the agent took on routine and repetitive jobs (task assignment), which freed the responders to focus on the situated rescue mission. In this on-the-loop interaction design, the role of the human HQ was relatively weak. For example, the HQ struggled to override the agents' instructions through the messaging channel. In the next study we explore an in-the-loop design which seeks to allow the HQ to play a stronger role in the control loop to enable more direct mediation and amendment of agent instructions (e.g., by directly modifying the task assignments, or by adding information relating to the assignments, such as safe routing).

7.7 SUMMARY

In this chapter, we examined how the guidance from a planning agent is handled socially in the human on-the-loop setting. To support our field trial we integrated a planner agent with AtomicOrchid and modified both mobile and HQ interface to facilitate the on-the-loop interaction pattern between human and agent. Findings from interaction analysis of field observations, triangulated with log files, reveal how the on-the-loop interactions played out. The results of the analysis show a division of labour in which the agent takes over the majority of planning activities while field responders only focus on other issues such as finding routes and targets. However, field observations also reveal significant costs associated with instructions that require members to reform new teams, and that interrupt on-going tasks.

In addition, some confusions and misunderstanding are also discovered in the human system feedback loop. Based on the findings, we presented three design implications to consider when creating agent-based planning support systems for human teams, including establishing ‘common ground’, facilitating accountability and balancing responsibilities between human and agent.

ATOMICORCHID STUDY 3: AGENT-SUPPORTED IN-THE-LOOP DESIGN

This chapter presents the third iteration of the AtomicOrchid ([AO](#)) field trials. The purpose of this iteration is to investigate socio-technical issues relating to agent planning support with a human in-the-loop interaction design. Based on the on-the-loop version of AtomicOrchid (Section [7.2](#)), the system has gone through another development iteration to facilitate an in-the-loop interaction pattern. Through interaction analysis of video recordings and game log data, we reveal how this in-the-loop design unfolds and, through a number of critical incidents, how it breaks down. A workshop with a professional Disaster Response team is conducted to reflect on the realism of the AtomicOrchid scenario.

8.1 INTRODUCTION

In this study, we refine the [AO](#) interface and place a human coordinator in-the-loop between the planning algorithm and the physical world. The in-the-loop design pattern assumes the constant HQ supervision and intervention is required to ensure the agent works properly (section [4.4](#)). The AtomicOrchid system has evolved from the second game probe in study 2 (chapter [7](#)), to facilitate the in-the-loop interaction. The in-the-loop design enables HQ human to involve in the planning process by:

1. allowing HQ to review, edit and approve every instruction generated by the agent. (Figure [41](#) (1)). In extreme cases, HQ can override all agent instructions, i.e. manually allocate all tasks;

2. allowing HQ to decide when to initiate re-planning (Figure 41 (2)); and
3. allowing HQ to review feedback from field players and decide how to act on them. (Figure 41 (3))

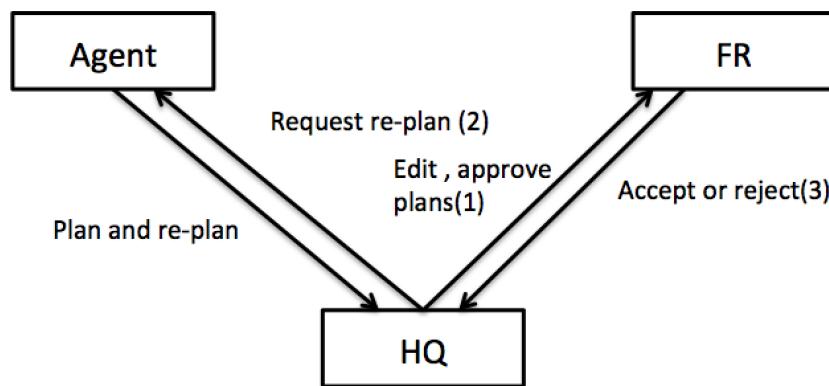


Figure 41: In-the-loop interaction design

Compared to the human on-the-loop design in the previous study (Section 4.4, chapter 6), the major change is the rebalancing of responsibilities in the control room (between HQ and agent). The interaction between field responder and the control room is mostly unchanged. Therefore, this study has a strong focus on control room interaction compared to study 2. More specifically, the objective of this study is to unpack the human system interaction in the human in-the-loop paradigm, particularly in the control room, revealing division of labour, interactional issues and design implications that can be drawn from the interactions. Findings from the study reveal the processes in which the agent and HQ players collectively generate task assignments for field players. Occasional failures in the process highlight a set of issues including complacency; silent, missing or invisible information; and limited support for human planning. In addition, a workshop with a professional Disaster Response team is reported in this chapter to reflect on the realism of the AtomicOrchid scenario.

8.2 SYSTEM EVOLUTION

The in-the-loop version of AtomicOrchid is not designed from scratch, but evolved from the on-the-loop version introduced in the previous study (Section 7.2). In study 2, we observed HQ struggling to get involved in the planning loop even when they wanted to intervene. Combining the observations from study 2 and the definition of the in-the-loop interaction pattern (4.4), we further generate several system requirements for realizing the in-the-loop interaction.

1. HQ should be able to review, edit and approve every instruction generated by the agent.
2. HQ should be able to decide when the agent should re-plan.
3. HQ should be able to plan for part of the team, leaving the agent to plan for the rest of the team.
4. HQ should be able to communicate their assignments (or task cancellation) to FRs in a structured way.

The purpose of requirement 1-2 is to give HQ more control over the planning loop, by delegating to them the responsibility for the final decision in planning. Requirement 3 enables HQ to modify the agent planning without having to take full manual control of plan generation. Requirement 4 is derived from the observations (study 1 and 2) that HQ struggled to override agent planning through unstructured text messages. Therefore, we suggest that HQ should be able to deliver and cancel their assignments in the same structured way that the agent does. Two interfaces are designed for the 2 HQ players in the control room. The task assignment interface provides a set of interface functionalities supporting task allocation activities, while the situational awareness interface provides game status and a broadcast message channel 42.

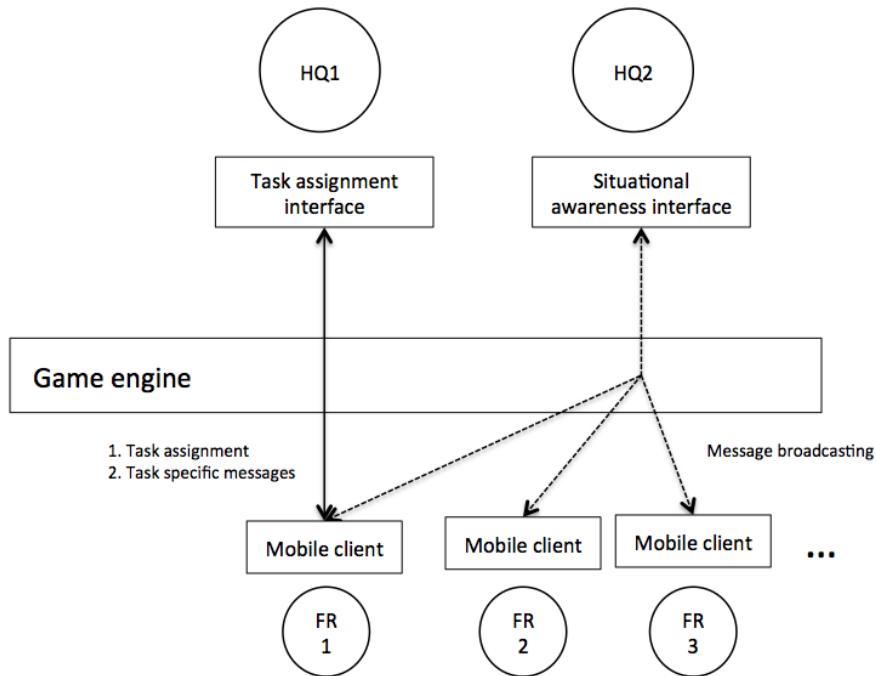


Figure 42: Messaging system design

8.2.1 Interfaces

This section describes the three game interfaces used by players, which are the mobile responder interface for field responders, the task assignment interface and the situational awareness interface for HQ players.

Compared to the on-the-loop version of AtomicOrchid, the mobile interface is largely unchanged except for the HQ task tab (Figure 43). The task tab now displays a task with text description and map visualisation of the task at the top. The bottom half of the interface is a message box showing task-specific information from HQ. It should be noted that the HQ can still send broadcast information (visible to everyone), which will be displayed in the chat tab.

The situational awareness interface is the same as the HQ interface used in study 2 (Section 7.2.3). It provides information about game status for monitoring, and a broadcast message channel for communicating with field players (Figure 44).

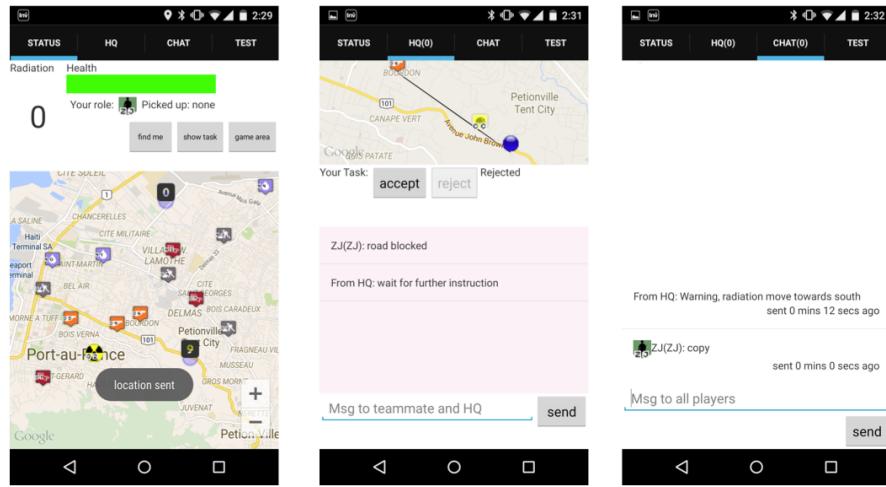


Figure 43: Mobile interface in study 2



Figure 44: Situational awareness interface operated by HQ2

The task assignment interface is created to support in-the-loop interaction with the agent. As an overview, the interface has a map on the left. Figure 45 shows a screenshot of task assignment interface. Components are numbered in the figure to illustrate functionalities of the interface. Player/target locations and assignments are presented on the map. At the right side of the interface is a task assignment panel. The left (1) column of the panel shows pending assignments while the right column (2) shows the existing task status. Figure 45 (5) shows an example of a proposed task assignment: player MD and GO are assigned to target 07. Within each confirmed task assignment (6) a feedback indicator indicates the field player's response to this assignments (no response, reject, accept).

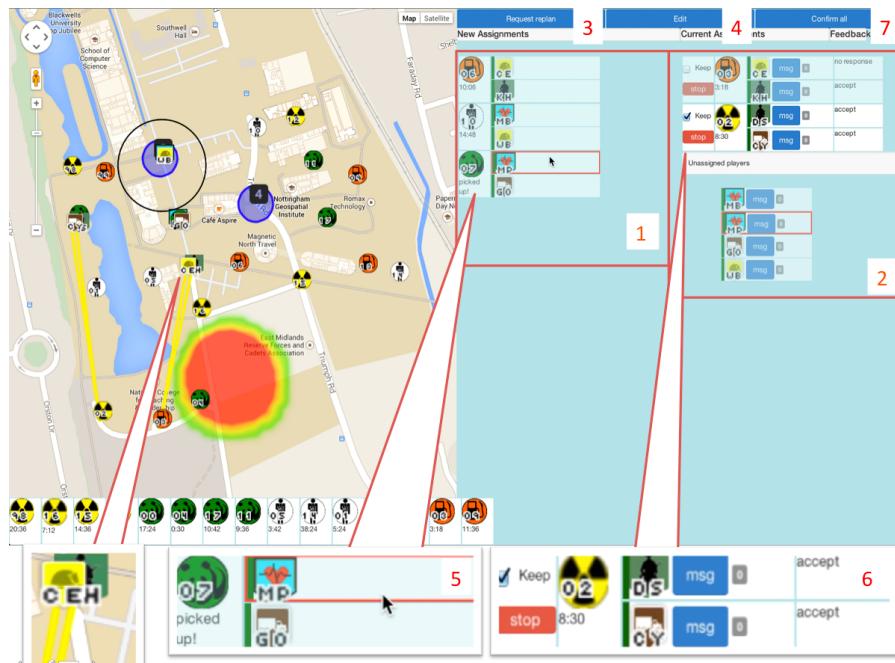


Figure 45: Interfaces in study 3

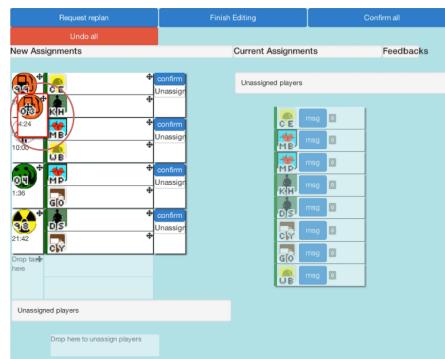


Figure 46: Edit mode of task allocation interface

- Plan request button:

This button triggers agent re-planning (Figure 45 (3)). The agent will calculate an optimized plan based on task status, and present it to the HQ on the pending panel. This button allows HQ to decide when to initiate a re-plan.

- Plan keeping checkbox:

These checkboxes are attached to every task assignment in the confirmed panel (Figure 45 (6)). If the checkbox is ticked, the planner will keep the corresponding assignment in the next re-planning. In other words, the planner will keep the assignment fixed, performing partial planning for the rest of the team.

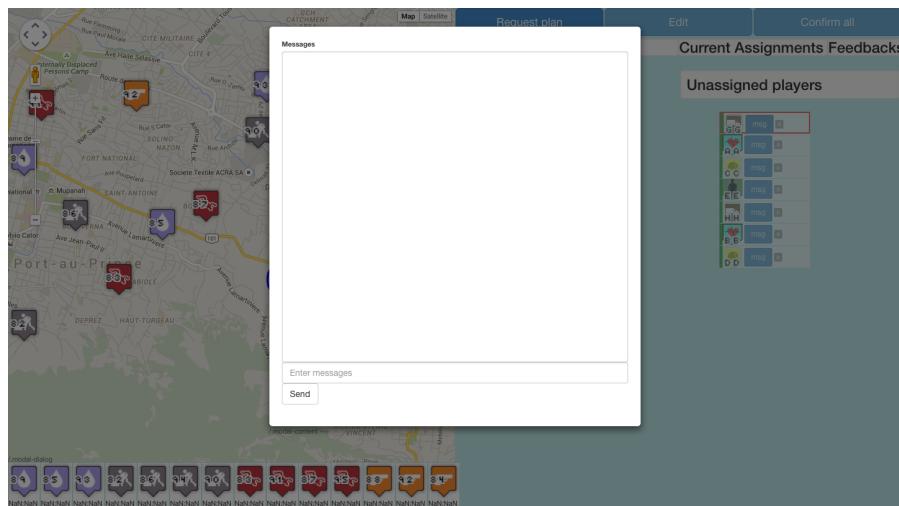


Figure 47: Pop up messaging panel

- Plan edit panel:

Manual plan edits will be activated by clicking the ‘edit’ button (Figure 45 (4)). The assignments in pending area will change to edit mode. Assignment can be created, modified and deleted through drag and drop interaction (Figure 46).

- Plan approval button:

This button approves all pending assignments (Figure 45 (7)). All pending assignments will move to the confirmed area. Alternatively, assignments can be approved individually by clicking individual confirm button on the pending assignment when the edit mode is activated.

- Text messaging panel:

The messaging panel (Figure 47) can be toggled by clicking msg button on the confirmed assignments (Figure 45 (4)). The panel is supposed to be used for assignment-specific information. Therefore, the messages in this panel are only visible to the two involved players and HQ.

- The Feedback indicator:

The feedback indicators are attached to the right hand side of

the confirmed assignments (Figure 45 (6)). The field players can easily provide feedback on their assignment through the mobile responder interface (introduced later). There are three possible values for the indicator (no response, reject, accept). Because rejections typically indicate issues that needs to be followed up by HQ, the rejection will be highlighted in red, and flashing.

When both involved players accept the assignment, the keep checkbox will be ticked automatically. This is a mechanism to avoid accidental interruption of the accepted assignments in subsequent re-plans.

- Stop button:

The stop button can be used to indicate an emergency termination of an assignment (Figure 45 (6)). If this button is clicked, the assignment will be dismissed both in the mobile and HQ interface.

8.2.2 *The planning agent*

One big change to the planner compared to study 2 is the addition of a partial planning feature. The agent can take a list of fixed assignments as input. It then plan for the players and targets that are not involved in the fixed assignment list. This functionality has two potential uses:

1. It allows human operators to contribute part of a plan and ask the agent optimize the rest.
2. It allows human operators to annotate some on-going tasks, so that they could not be changed in dynamic re-planning.

Apart from the partial planning feature, the input/output of the agent is not changed.

8.3 STUDY DESIGN

Participants were recruited through posters and emails. A total of 20 participants were recruited. 10 participated in session A and 10 in session B (Appendix D.2). All participants were reimbursed with 15 pounds for 1.5 hours of study. For each game session, there are 2 HQ players and 8 field players. The majority of participants were students of the local university. The HQ players were recruited from researchers in the computer science department.

Because the HQ interface is a lot more complicated than that of the study 1 and study 2, we add an extra 0.5 hour training session before the formal study for HQ players to get familiar with the new task assignment interface. We anticipated that the workload of operating the human in-the-loop interface would be a lot more than that of operating the on-the-loop interface. Therefore, there are two HQ players recruited in each session to split work in the control room. One of the two HQ player operates the new task allocation interface (described in section 8.2), while the other player operates the situational awareness interface (described in study 2, section 7.2.3) to assist the other HQ player by providing situation awareness and sending broadcasting information.

Upon arrival in the HQ (set up in a meeting room at the local university), participants were briefed and asked to consent to participate. Roles were randomly assigned to field players (firefighter, medic, transporter, soldier). Field responders were provided with a smartphone; HQ coordinators with a laptop. Game rules and interfaces were introduced, and participants were assisted in setting up their phones and laptop clients. Field responders and HQ coordinators were given 5 minutes to discuss a common game strategy. All

field responders were accompanied to the starting point within the designated game area, about 1 minute walk from headquarters.

Before the formal session begins, there was a training session for field players to get familiar with the mobile interface. The training session has a very simple game setting with only four targets nearby the starting point. The training session ends when field responders collect all four targets nearby. Once field responders were ready to start the formal session, one researcher started the game engine, triggering a “game start” message to be sent to the mobile interfaces. Gameplay commenced for 30 minutes. A “Game over” message by HQ concluded the game. Field responders returned to HQ for the post-game session.

The game area was the same as for study 1 and 2, on the local university campus (Appendix D.1), 400 by 400 meters, without heavy traffic. The terrain of the game area includes grassland, a lake, buildings, roads, and footpaths and lawns. There are two drop off zones. Due to extra HQ training sessions and improved interface, the number of targets is increased to 20 (from 16 in study 2) to increase the game challenge. There were four targets for each of the four target types. The pattern of cloud movement and expansion was the same for both game sessions and the same as for study 1 and 2.

We recorded both system logs and video of interaction in the field for analysis. To capture the distributed, concurrent nature of the interaction, four researchers with camcorders shadowed the field player teams, and one researcher recorded the action in the HQ. The replay tool was used to synchronise and analyse triangulated game events, player positions, and concurrent video recordings (Appendix D.3).

Our interest in this chapter is how socio-technical interaction is organised around the computational planning support, hence our focus is on the control room first, but then we trace information flow and decision making into the field. In practice, video recordings of the

control room were catalogued to identify key decision points in teaming and task allocation, which served to index sequences (episodes) of interest (Section 5.2). Interesting distinct units of interaction were then transcribed and triangulated with log files and field video for deeper analysis; the results of which we present in this chapter.

8.4 DATA ANALYSIS

This section starts with overview of game results, messaging system usage and task assignments, as they served to index episodes of interest. Selected episodes of game play are then presented in order to unpack the interactions surrounding the task assignment activities in the control room. We provide these episodes as vivid exhibits of how members accountably organise their team coordination *in situ* [Crabtree et al., 2012]. The order in which we present the episodes follows the common practice of moving from exhibits of typical/unproblematic instances, via more complex/difficult instances, to exhibits that display problematic interaction or even complete breakdowns [Heath et al., 2010].

Overall, 28 of the 40 targets were evacuated in two sessions (16 in Session A and 12 in Session B). The player's health status in session 1 is better (Avg 90, Sd 9.3) then that in session 2 (Avg 48, Sd 41). Two deaths occurred at the beginning of session 2. More details of death will be presented as episodes later in this section.

| | Target saved | Health max | Health min | Health avg | Health Sd | Death |
|-----------|--------------|------------|------------|------------|-----------|-------|
| Session A | 16 | 99 | 75 | 90.75 | 9.337 | 0 |
| Session B | 12 | 97 | 0 | 48.12 | 41 | 2 |

Table 9: Result overview of study 3

8.4.1 Messaging system

One change of the messaging system made for this iteration is the separated channels for assignment-specific and broadcast messages. The HQ₁, with the task assignment interface, is responsible for sending message in assignment-specific channel, while the HQ₂, with the situation awareness interface, is responsible for sending messages in the general message channel. This section will reveal how this design plays out in the field trials.

| Session 1 | Sent by HQ | Sent by FR | Total |
|-------------------------|------------------|-------------------------|-------|
| | Broadcasting msg | Assignment specific msg | Total |
| Total | 35 | 3 | 38 |
| | 22 | 15 | 37 |
| | 57 | 18 | 75 |
| Session 2 | Sent by HQ | Sent by FR | Total |
| Broadcasting msg | 31 | 13 | 44 |
| Assignment specific msg | 19 | 11 | 30 |
| Total | 50 | 24 | 74 |

Table 10: Task assignment overview

We found HQ players frequently send messages to update FRs about the locations of the radiation cloud, (e.g. “Radiation Status- 38 39 37 and Drop Point 7 all out of bounds”) and provide navigational guidance (e.g. “go north and west around the water”). HQ is also observed to send messages to repeat and enhance the task assignment (e.g. “turn to 49”).

On the other side of the message channel, field responders send messages to request tasks (e.g. “please advise”) and cloud status (e.g. “Which way is it moving?”). Field responders also occasionally send acknowledgments to HQ’s messages (e.g. “Copy that.”).

Most messages in the general message channel are general information about the clouds. However, we also found 11 messages in the general message channel (out of 82) are clearly addressed to in-

dividual teams. The specific player initials are mentioned in those messages. (E.g. “NG and YI approach quicker to 41 drop off to 8”)

8.4.2 Overview of task assignments

In the following tree diagrams, plans are broken down into individual task assignments. Each individual assignment may go through 4 stages in the planning process (creation, approval, feedback, and execution). The assignment status for each stage is summarized in the following diagrams(Figure 48,49).

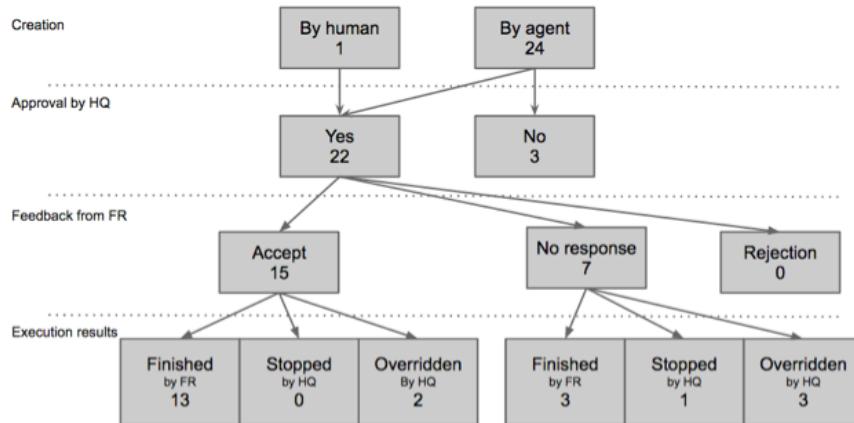


Figure 48: Task assignment in session 1

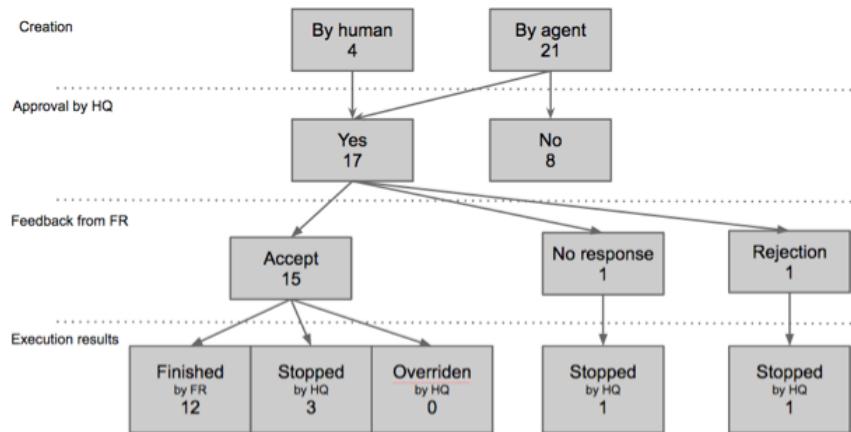


Figure 49: Task assignment in session 2

In summary, HQ and the planning agent contribute to the task planning activities in the control room. The planning agent created a total of 45 task assignments with an additional 5 assignments created manually by HQ. HQ approved a total of 39 assignments. Field responders accepted most of the approved assignments (30 out of 39). 9 assignments were rejected or not responded to. During task execution, occasional HQ interventions resulted in 5 task cancellations and 5 assignments being override. In the end, players managed to evacuate a total of 28 (out of 40) targets in 2 sessions.

Although the diagrams (figure 48, 49) suggest a sequential planning-execution process (creation -> approval -> feedback -> execution), we actually found that various planning activities (e.g. assignment creation, approval, intervention, communication) are highly intertwined in the control room setting. In the rest of the section, episodes of game play will be used to unpack the interactions surrounding the task assignment activities in the control room. The episodes in later sections will be presented with a standard orthographic notation detailed in section 5.2.

8.4.3 Confirming the plan

As summarised above, a majority of task assignments are generated by the planning agent and approved by the HQ players. The episode 3.1 demonstrates a typical case of the routine task planning process in control room.

Episode 3.1 Context: HQ paid attention to a team (MV XW) who was carrying target 43 back to drop off zone 7.

16:45, HQ2: XW and MV.

16:50, HQ1: (taking) 43.

16:51, HQ2: They should be going to drop off (zone) 7 and get 36 (Figure 50)

16:58, HQ1: why don't they go this way?

16:49, HQ2: tell them to go 36 afterwards.

17:04, HQ1: ok this one (refer to target 36). Do you tell them? (Figure 51)

17:05, HQ2: Should I tell them? (Typing)

17:07, HQ1: Yeah, go for 36. Maybe After the drop off, I think then (will) get confused.

17:13, HQ2: I will tell them to go after drop off.

17:14, HQ1: Yeah, Yeah.

18.10, (The team dropped off target 43)

18:22, HQ1: (click re-plan)

18:26, (new assignments, the team is assigned to MV,XW -> 36)

18.28, HQ1 36, yes (Click confirm)



Figure 50: Episode 3.1, HQ1 (left)

HQ2 (right)



Figure 51: Episode 3.2, HQ1 (left)

HQ2 (right)

At the beginning of this episode, the team MV, XW are carrying a target (43), approaching drop off zone. At [16:45], the HQ2 noticed the team was going complete its current task and began to consider a new target for them [16:51]. HQ2 proposes that 36 should be prioritized (Figure 50). HQ1 agreed with the suggestion and decided to send the assignment after the team completes existing task (Figure 51) [17:07]. At 18:10, the team dropped off target. After the drop-off, the HQ requested a re-plan. The agent assigns the team MV, XW to target 36, which is consistent with decision of the HQ players. At the end, HQ approved the assignment [18:28]. In this typical case of task assignment, HQ can be seen to be monitoring the task execution, and making timely requests for new task assignments. HQ's discussion suggested that the agent task assignment was approved after a careful review of player, target and radiation status. In addition, much of HQ1 and HQ2's discussion happens before the team drops off the tar-

get; this kind of forward planning was observed on several occasions.

8.4.4 Just following the Plan

Episode 3.2 gives a more problematic example of accepting the planning agent's task assignments. Episode 3.2-1 describes the interaction in the control room; while episode 3.2-2 gives the perspective of the field players.

08:32, Episode 3.2-1

NG and YI have just completed a task together; DI is some distance away.

08:32, (HQ1 request assignment for Idle player NG, YI))

08:37, (New assignment NG, YI -> 50)

08:40, (HQ1 confirmed plan)

09:09, (HQ1 request assignment for DI who has just become idle)

(NG have not confirmed the previous assignment; DI is closer to target 50)

09:16, (New assignment arrived NG, DI -> 50)

09:17, (HQ1 confirm NG, DI -> 50)

In this episode, the HQ1 requests assignment twice for idle players, which is a routine activity for HQ1 (As discussed in episode 1). Firstly, HQ1 request and approve new assignment for NG, YI [08:32]. 30 seconds later, DI become idle [09:09], so HQ1 request plan again. Because DI is closer to target 50 and NG YI have not accepted their assignment, the agent replaces DI with NG in the new assignment to minimize travelling distance of field players. Although HQ1 does not verbalize his reasoning process the quick approval suggests HQ1 may not have inspected the new assignment carefully and may not have noticed that the previous assignment will be overridden, potentially interrupting a task in progress. The episode 3.2-2 picks up the activity of NG and YI as the first new task assignment is approved by HQ.

Episode 3.2-2

NG YI received a task update after they finished their previous task

(Team NG YI received new task NG, YI -> 50)

08:39, NG: Task changed, to what?

08:43, NG: Oh sh*t, there is a radiation zone sh*t.

08:53, NG: It (target 50) is close to the lake. It is triumph road we have to go that way.

08:55, YI: Yea, we need to. No this is the lake.

09:00, NG: Oh that is the river, so it is that way, Can we go? Yes we can go, this road?

09:10, YI: Yes it should be.

(Their task is interrupted by new assignment YI, DI -> 50) 09:18,

YI: Task changed.

09:19, NG: Task changed? What?

09:23, YI: No, it is the same.

09:26, NG: I do not see anything.

09:28, YI: You are DI, right?

09:29, NG: No.

09:30, YI: En? You are no DI? DI, wait.

09:37, NG: I can not see my Task.

09:45, YI: You are NG, Oh no, I need DI.

09:48, NG: You should go back then. Or send a message, Oh I can not see my task.

10:02, NG: I think the page is not loading.

NG and YI have already received the task to rescue target 50 but have not accepted or rejected it. While they are discussing how to reach the target the second task assignment arrives [09:10], the agent requires YI to team up with the remote player DI to pursue the same target, 50. The team is confused after the change of task. YI thinks that the task did not change because the target is still 50 [09:23] (YI: "No, it is the same."), and NG thought the interface is no longer working [10:02](NG: "I think the page is not loading."). After YI confirms

NG's initials, she realized she needs to switch teammate (to DI).

Overall, the task interrupt is resulted in a problematic sense making process in the field. Meanwhile the players in the control room appear unaware that anything untoward has occurred. The observation also reveals that several factors could have led to the task interruption including the absence of response from field responders; computational planning performed by agent without timely field feedback; and HQ's failure to discover the task interruption during assignment approval.

8.4.5 *Correcting the plan*

HQ players occasionally chose to change the task assignments generated by the planning agent; episode 3.3 presents one such example.

Episode 3.3

CE and KH are currently assigned target 03 but have not accepted; other players are free 04:18, (HQ1 click request plan)

(Assignments arrived: CE KH -> 06, MP GO ->07, MB ,WB -> 10)

04:24, HQ1: What? Why I am getting? Ahh, one of these guys does not accept. (Figure 52)

(Referring to the team of CE, KH.)

04:29, (HQ1 clicked keep on assignment CE KH -> 03)

04:33, (HQ1 request re-plan)

04:40, (Assignment arrived: MB, WB -> 10, MP, GO -> 07)

04:42, (HQ1 click confirm)

04:56, (MP, GO accepted)

After HQ requests a new plan, the agent proposes a set of assignments, one of which (CE KH -> 06) interrupts the existing task of a team (CE KH -> 03). HQ1 queries this change ("What? Why I am getting?") realises that they have not explicitly accepted the previous task [04:24]. It should be noted that this kind of task interruption only happens when field players did not accept the tasks (Section 8.2),

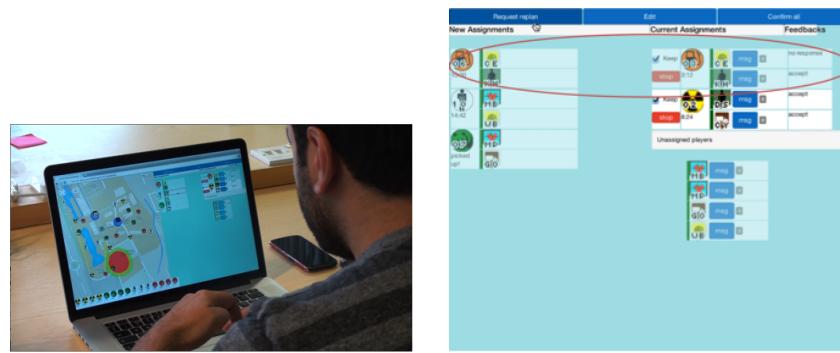


Figure 52: Episode 3.3, HQ identifies a task interruption, highlighted on right figure

making the agent think they are idle at that moment. After finding this problem, HQ then requires the agent to “keep” existing assignment [04:29] and requests a new plan from the agent, which is then approved. In contrast to episode 3.2, HQ1 notices and compensates for the field players’ failure to explicitly accept the task, and the field players are able to continue with the previously allocated task without interruption.

8.4.6 *Changing the plan*

At the start of session B there is an extreme example of the HQ players overriding the planning agent.

Episode 3.4

Context: At start of the session 1, all players were idle waiting for initial plan.

01:25, (HQ1 requested initial plan)

01:28, (4 initial assignments arrived)

01:29, HQ1: why, it is stupid.

01:33, (HQ1 click edit)

01:53, HQ1: I want this one (HQ1 drag targets to replace agent planning) this one and this one.

HQ1 replaced 3 out of 4 targets in the task assignment. The three prioritized targets very are close to the original cloud

02:03, (HQ1 clicked confirm)

02:07, (HQ1 talk to HQ2) HQ1: I think we should get the far ones first.

The episode begins with HQ requesting initial task assignments for the whole team. When the agent gives HQ a set task assignments for approval, HQ complained about it [01:29, HQ: why, it is stupid.], indicating he is not satisfied with the plan. HQ then click edit button to switch to edit mode. Under the edit mode, HQ dragged 3 targets to replace the targets in agent assignments. The three prioritized targets are the ones that are closest to the radiation cloud. After HQ confirmed his modification [02:03], he said to HQ2 that the far away targets should be rescued first [02:07]. The execution result of this heavily edited plan is not ideal. Among the 3 three modified assignments, 1 of them is finished successfully, 1 assignment is cancelled later, and 1 assignment leads to player “death”.

8.4.7 *Coping with the unexpected*

Episode 3.5 picks up shortly after episode 3.4, and exemplifies how HQ adapts task assignments to changing task status.

Episoe 3.5 Context: after plans are confirmed in episode 3.4, HQs are monitoring FR's progress

02:09, HQ2: they cannot walk there straight. (Figure 53)

Referring to team MB, GO, HQ2 point out that the straight path to one of the target is blocked by radiation

02:14, HQ1: So who is that. (Target) 04 (HQ1 open the message panel)

02:20, (HQ 1 Types message) You are heading to an area affected by cloud, You need to be very fast. [Fig 6]

02:45, HQ2: 04 is now in the cloud.

02:45, HQ2: Oh god, that is so fast.

03:20, (HQ1 Types message) Stay to the lake as possible!

03:33, HQ1: Shall we cancel (assignment 04) that or shall we wait for report? (HQ1 opens msg panel to talk to MP G0)

03:43, (HQ1 opens msg panel to talk to MP G0)

03:50, HQ2: I think (Target) 04 is only an animal, screw the animal.

03:52, (HQ1 Types message) Abort, target compromised, proceed to target 07.

04:06, (HQ1 sent message)

04:11, (HQ1 click stop on the assignment of MB, G0->04)



Figure 53: Episode 3.5, HQ2 points to a team

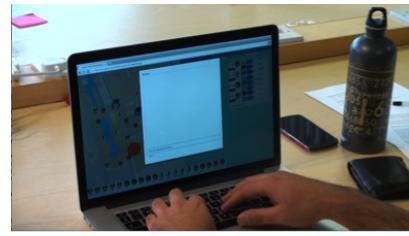


Figure 54: Episode 3.5, HQ sends msg for task cancellation

After confirming assignments in episode 3.1, the HQ2 points out that one assignment may now be impractical because the route to target has already been blocked by a radiation cloud [02:09] (Figure 53). HQ1 immediately opens message panel to send warnings and urge the team to move quickly (Figure 54). However the cloud expansion seems to be faster than HQ originally expected [02:51]. Apart from sending route guidance to the team, HQ1 starts to consider cancellation of the assignment. After HQ2 agree with the cancellation, HQ1 sends a message to the team to inform them that the assignment is going to be cancelled and instructs them to go to new target 07. The assignment is formally cancelled in [04:11]. After HQ1 cancels the task, he starts to request new assignments from planner and allocates target 07 to the team later. In this case the HQ players realise the risk and are able to abort the task and redirect these field players out of danger.

8.4.8 When it all breaks down

In episode 3.4, in which the HQ modified 3 agent assignments. In the modified plan, the team CE, KM is instructed to go for a target very close to a radiation cloud. We look first at the HQ perspective (episode 3.6-1), and then at the field players' perspective (episode 3.6-2).

Episode 3.6-1

(Context: One team (CE, KM) is end up heading to radiation cloud.)

4:52, HQ2 They (CE KM) are walking right into their death.

(CE and KM were heading to target 03 which is now at opposite side of a radiation cloud)

5:14, (HQ1 messaging) Careful, you are approaching the cloud.

Move extremely fast to the cloud. Go around the lake to return.

5:32, HQ1: God, too slow.

(The team CE KM is completely in the mid of radiation)

5.46, (HQ1 typing message) Move closer to the lake.

5:50, (The team went through the cloud, their health is below 20, the target 03 is now at the edge of the cloud)

6.39, HQ2: What are they trying to pick up?

6:43, HQ1: The fuel.

6.48, HQ2: I am not sure whether they can survive picking up the fuel. No, now is more concern of surviving.

7:05, (HQ1 messaging) Abort 03, return top to (target) 99.

9.04, (HQ1 stopped task of CE KM)

9.05-10:08, (HQ1 requested re-plan 4 times, no new plan available for CE KM)

9.50, HQ1: No targets for them? There are lots of them.

(Team CE, KM stay in the cloud all the time, and finally dead)

At the beginning of the task, the radiation cloud is already between target and the field team CE and KM. The cloud is expanding quickly, which catches the HQ1 by surprise. As a result, 2 messages are sent by HQ1 to guide the team to avoid the cloud [05:14,5:46]. However

the team is still excessively exposed to the radiation when they reach the side of the target, so HQ1 decides to abort the task [07:05]. After HQ1 cancels the assignment, the team is still exposed to radiation. HQ1 then tries to assign the players to other targets by requesting the agent for new assignments [09:05-10:08]. However, the agent does not assign the team to any task because the team's health is too low. HQ1 seems to be confused about why the agent refused to assign targets. He repeatedly requests plans three times and says "No targets for them? There are lots of them (refer to targets)." [9:50]. During this process, the team are still standing in the cloud. They finally lose all their health points and become incapacitated. Field players' perspective is given below.

Episode 3.6-2

CE and KM are heading towards target 03

05:15, CE: (Reading out the message) You are approaching the cloud, move extremely fast to the cloud, Ok!

05:21, (CE grabbed KM and started running) (Figure 55)

06:12, (The team CW, KM ran all the way across the cloud)

(After ran through the cloud, they are checking the health value.)

07:02, CE: how dead are you?

07:04, KM: Pretty much dead.

07:06, CE: I am pretty dead as well.

(The team is trying to locate the target, but it is unsuccessful)

(Figure 56)

07:28, CE: Do you know what is funny? We have to get back.

08:24, CE: It should be somewhere around here! We are getting close.

(Assignment cancelled)

08:52, CE: there is a new task, No task at the moment?

[...] *(After assignment was cancelled, the team was still trying to locate the target)*

09:44, CE: I am pretty much dead and radiation is 12. Where is it! It says it is on this street but it is not.

10:06, CE: Oh I am dead.



Figure 55: Episode 3.6, CE (left) grabs KM (right), start running into cloud.



Figure 56: Episode 3.6, CE (left) grabs KM (right) walks around, looking into mobile screen, trying to locate target

The part 2 begins with player CE reading the HQ message ("You are approaching the cloud, move extremely fast"). After this message, CE grabs teammate KM and starts to run through the cloud 55. After they run through the cloud, the team check their remaining health value [07:02] and start searching for the target at the edge of the radiation cloud. About half a minute later, the assignment is cancelled by HQ [08:52] and no further task is allocated (as seen above). Although the task has been cancelled, the HQ does not assign a new task to the team [3.6-1, 9:50]. The field players note but then appear to ignore HQ1's instruction to abort [08:24] and continue to search for the target. The field players seem to interpret the lack of a new task as license to remain where they are and are eventually overwhelmed

by the radiation.

8.4.9 Missing feedback from field

Lack of feedback leads to false assumptions by the agent (i.e. players are still available), which in turn compromises its optimization (e.g. section 8.4.4). We found that HQ may also be distracted by the lack of response, in that they have to guess a player's intention without their feedback.

Episode 3.7

(New assignment MV XW -> 36, XW accepted, MV had no response)

13:05, HQ1: I think they have not received any task probably

13:22, HQ1: Just do not have response, I do not know why.

13:32, HQ2: MV and XW, I suppose they go to 36.

13:35, HQ: They are moving, but they did not accept that.

13:48, (HQ click request, new assignment to MV XW -> 43, NO response from MV XW)?

13:51, HQ2: I am not sure they are still going south; I think they (MV XW) are going to 43 instead.

13:59, HQ1: HQ: they are going to 43.

14:05, HQ1: HQ: they just do not have any response, no reply.

But they are coming for it, they are coming.

In this episode, the team MV, XW was assigned target twice [12:30,13:48], but the team neither rejects nor accepts those two assignments. The HQs are observed to guess intentions of the field responders 3 times [see 13:32, 13:51, 13:59] and complain about the non-response twice [see 13:22, 13:48]

8.4.10 *Division of labour in HQ*

In session 1, the responsibility for HQ1 and HQ2 to send messages are not fixed as originally designed, but dynamically negotiated between the two HQ operators. The following fragment is a typical case of negotiation.

Episode 3.8 17:04, HQ1: Ok this one (refer to target 36). Do you tell them?

17:05, HQ2: Should I tell them?

17:07, HQ1: Yeah, (tell them to) go for 36.

(HQ2 sent message) MV and XW move north of the water and go round to 36

In this fragment, HQ1 asks HQ2 to send a message to instruct a team to go for a target 36. The message is a task-specific, but it is sent by HQ2 through the broadcasting message channel.

In session B, we do not find similar negotiation. However, in the post game discussion, the HQ2 expressed the view that he had too little to do (other than sending general messages), while HQ1 had too much to do:

“The interface I use was completely useless, because I can only message everyone at the same time, basically spamming everybody, so the only thing that I can do is, Oh , the cloud moves there and there. At the same time, HQ1 had to message to everybody privately. While he was doing that, he couldn’t do new plans. So there is a lot stuff for him be I coundn’t do anything.”

8.5 DISCUSSION

In previous section, episodes were presented to illustrate how the in-the-loop interaction design plays out in the field. The episodes reveal

that, to a large extent, the HQ players are successfully involved in the control loop. However, we also found some confusions and misunderstandings occur between agent and the human which lead to issues and implications for interaction design.

8.5.1 *How does the in-the-loop design play out?*

The interface functionalities are designed to enable HQ to engage in a range of interface interactions such as plan requesting, editing, approval and cancellation. This section examines how these functionalities are utilized and to what extent they help HQ to stay in the control loop. In what follows, we firstly examine the usage of individual interface functionalities. Combining the episodes presented in the previous section, we then reveal how the division of labour plays out with the support of these interface functionalities.

1. **Plan request and approval.** The plan request function is used by HQ to trigger agent re-planning. The design of the plan approval stage gives HQ an opportunity moment to review and influence the final plans before they are sent to field responders. Both functions are essential for the human and agent to collectively produce task assignments, constituting the routine task planning work in the control room. The two functions are the two most frequently used interface functionalities (HQ requested plans for 45 times; approved 39 task assignments). Deciding the appropriate moment to request plans and approve desired assignments requires HQ to closely monitor the task status. The uses of the two interface functionalities are usually observed together with discussion of task execution status, which indicates that HQ players are engaged in the planning-execution loop with a supervisory role.

2. **Plan edits.** The “plan edits” enable HQs to directly intervene in the planning. We observed HQ modify undesirable agent plans twice throughout the field study (Episode 3.3, 3.4). It should be noted that the function is designed to be used infrequently, because the planning agent is supposed to take over the majority of the computational intensive planning activities.

3. **Plan cancellation.** Further, the task cancellation functionality allows HQ to influence task execution after plan approval. In episode 3.5, we observed an assignment being cancelled and teams being reassigned due to unexpected cloud activities. This may suggest that the combination of cancellation and re-planning can be a useful tool for HQ to respond to contingencies in the task execution.

4. **Partial planning** The partial planning functionality allows HQs to indirectly influence the planner. For example in episode 3.3, HQ identified a task interruption in the proposed plan. He then required the agent to keep an existing assignment and perform re-planning again. This episode can be seen as a case in which HQ is able to make sense of the task status and utilize the partial planning functionality to influence the assumptions of the planner agent.

Apart from the above mentioned functionalities to support HQ intervention, the task assignment interface appears in many cases to provide an effective shared representation of the current state of the game. As well as showing current player and target locations and player health it also makes visible the currently approved task allocations, field player responses and any new plan that has been requested or is being edited. This shared information forms the common ground between the HQ players and the planning agent.

HQ players are observed to closely monitor this view and its representation of plan execution. For example episodes 3.1, 3.2, 3.3, 3.5 and 3.6 all reveal HQ players' awareness of field player progress and current tasks, episodes 3.5 and 3.6 show awareness of the cloud's location in relation to players, and episodes 3.1, 3.3 and 3.4 show HQ players engaging actively with proposed (rather than current) task assignments. We observe that the HQ players are quite capable of modifying the agent's plans when they wish to, for better (episode 3.3) or worse (episode 3.4). HQ is also able to intervene in current task allocations, which is successful in resolving the situation in episode 3.5 (but not in episode 3.6).

On the other hand, the agent is found to take over the computationally complicated task optimization, freeing up the HQ to play their supervisory roles. Throughout the two sessions, the HQ requested the planner to re-plan 49 times. The agent generated 45 task assignments, 34 out of which are approved by Headquarters. In comparison with the agent planner, the human only created 5 task assignments. The fact that agent creates a large proportion of assignments suggests the agent successfully takes over the routine planning to a large extent.

In many cases the communication between HQ and the field players is unproblematic, and most targets are successfully evacuated according to plan. This situation seems to be considerably better than that reported in chapter 6 and chapter 7, and we conjecture that this is due at least in part to differences in the mobile interface in the trials reported here. Specifically, unlike in the studies reported there, the current task allocation is shown as a graphical overlay on the mobile map, not just as a textual instruction (given by the HQ player in chapter 6 or the planning agent in chapter 7). This seems to significantly reduce the field players' confusion about their current target and team-mate and where to find them. The seemingly better situation of task communication is consistent with the requirements (Section 8.2) outlined for interface improvement that stressed the im-

portance of structured, domain-specific task representation for both HQ and field responders.

A pattern of division of labour is also observed between the two HQ players in the control room. To recap, the two HQ players in control room are split into two roles. The role of HQ1 is to handle task assignments and send assignment-specific messages to field responders. The role of HQ2 is to support monitor task status and send general chat messages. To support them, the messaging system is split in to two channels, general information and assignment-specific (Section 8.2). The messaging interfaces allow each HQ player to take control of one of the messaging channel. Participants were briefed on the designed role division in training session before the game play. However, the system does not play out exactly as designed in the field studies. 11 messages in the general message channel are task specific, which indicate the HQ players may occasionally violate the designed division of responsibility. Specifically, we have seen in session 2 the HQ players accepted this division of capabilities but were frustrated by it. However in session 1 the HQ players developed their own workarounds for this based on verbal negotiation, so that at times HQ2 would use the general broadcast interface to send task-specific messages for HQ1, relying on the identification of the intended field players by their codes within those messages.

8.5.2 *Responsibility and complacency*

In the game probe, the planning can be seen as partially automated by the planner agent. The task optimization performed by the agent can be considered to be ‘imperfect’ because it fails to consider any organizational efforts/overhead required execute the plans and plan changes, that is, the social cost of implementing plans. The result of this imperfection has been illustrated in previous study (Chapter 7), and also in episode 3.2 and 3.3. In the episodes 3.2 and 3.3, the agent proposes new plans that will disrupt ongoing activities, albeit because the respective field players have not (yet) explicitly signalled

their acceptance of these tasks.

This imperfections can be thought of an “natural” for agent planning support, as computationally “optimized” assignments from the planner may not be socially optimized for the responder teams (Section 7.5.2). This highlights the importance of human involvement in the planning activities. In principle, the human in-the-loop interaction design allows the human, in this case the HQ players, to take active responsibility for system. For example, in episode 3.3 we see HQ1 successfully ‘correcting’ the plan to allow for missing information (responses) from the field players. However in episode 3.2 the HQ player failed to prevent the unnecessary task interruption in the approval stage, resulting in extra coordination work and sense making for field responders. Considering this case in more detail we see that it arose from the combination of: lack of field player feedback; failure of HQ monitoring; and the particular computational optimization performed by agent (which assumes that there can be no overhead if a task has not been formally accepted).

The failure case in the episode is similar to the well-observed phenomena discovered by automation researchers called complacency [Kaber and Endsley, 1995]. The complacency phenomenon refers to the human failure to detect occasional automation failures. Studies of automation suggests that complacency can be classified as an attention-based monitoring failure, which is likely to happen when human is engaged in multi tasking. In the case of episode 3.2, the quick approval of problematic assignments suggests that the plan is not properly reviewed. The HQ’s attention could be a factor of the monitoring failure. As the HQ player is trying to allocate tasks for idle players, they may exclusively focus on the new assignments for the idle players rather than other potentially conflicting assignments. If the interface failed to attract HQ’s attention to the conflict, HQ may approve the plan as long as idle players are properly assigned. Although the failed in episode 3.2, the interface highlight does grab the attention of the HQ in episode 3.3. Some studies shows that com-

placency is found in both naive and expert participants and cannot be overcome with simple practice [Parasuraman and Manzey, 2010]. Therefore other mechanisms to counter the complacency effect may be needed, such as a richer involvement of the human in the planning, which we pick up further in Section 8.5.4.

8.5.3 *Tacit confusions*

While many interactions were essentially unproblematic we do observe a number of areas where silent, missing or invisible information led to confusion and break down.

For field players, task cancellation is presented on a notification (“task changed”) followed by a blank task page with only the text “No task assigned at the moment”. The field study suggests that the presentation could be problematic for the field players. Firstly, being assigned with no task does not successfully convey the meaning of task cancellation (i.e. “Don’t carry on”). The episode 3.6-2 shows a case in which the players completely ignores the task cancellation and carry on doing a risky task. In episode 3.2-2, the player NG even thinks that the blank task interface implies a malfunction of the mobile interface. Even if the players understand task cancellation, its implications can vary. Without a task, the players can chose to, for example, (1) stand by doing nothing, (2) find targets by themselves, (3) escape for their lives. Therefore, the field players may be confused about what to do next. In another example (not presented) of cancellation, the field players were stuck in a radiation cloud when their task was cancelled. With a high radiation reading and loss of a target, they don’t know what to do next so they send messages to HQ for clarification. They finally leave the radiation cloud after they receive and accept new assignments from HQ. This example shows the implicit instruction of “Don’t carry on” is confusing. Instead, the explicit instructions, which directly point out what to do next may be desirable. For example, in the case of AtomicOrchid, frequent instructions

might be “stand by”, “go back” or “escape”. We therefore argue that the interface should encourage HQ consider such “explicit” instructions and support quickly sending such instructions.

Second, episode 3.4 in which the HQ player drastically modifies the agent’s proposed plan illustrates vividly that the current interface makes the agent’s proposed plan available to the HQ player, but does not reveal the agent’s priorities or reasoning. The HQ player enacts a policy of evacuating the targets nearest to the radiation cloud first, presumably in the belief that the planning agent has (a) used a different priority and/or (b) not given due consideration to evacuating those targets. In fact the planning agent will have considered and rejected those targets, based on its model of field player movement, radiation cloud spread and ‘permitted’ radiation exposure (none). None of the agent’s models are likely to be perfect but at present they are not open to inspection. It soon transpires (episodes 3.5 and 3.6) that the HQ player’s initial expectation of how quickly the radiation cloud will spread is very wrong. In an ideal world we would want this information to be available to the HQ players, so that the common ground can be expanded to include not only the plan but the reasoning behind the plan. However, as in the previous point, there is also the challenge that we may overwhelm the HQ player with information in a time-critical decision-making situation.

Apart from the issues of task cancellation, issues of field player feedback also emerged. The field observations showed that the feedback system is crucial for both the agent and HQ players to keep track of the task status. Overall, most of the assignments (30 out of 39) have been responded to by the field players. However, there are still a number of un-responded-to assignments that causes issues in coordination. For the agent planner, the lack of field response may lead to false assumption of the availability of the players, which in turn, compromises the subsequent re-planning (Episode 3.2 and 3.3). For the HQ players, the lack of field response adds to the uncertainties in the task status as well. HQ players have to take efforts to guess the

intentions (Episode 3.7). We might wish that the field players would always respond in a timely manner. However there may be good reasons why they have not or cannot at present, e.g. if they are incapacitated or there are temporary problems with communication. Also, in the scenario of AtomicOrchid, no matter how fast the responders accept assignments agent planning may happen concurrently.

One strategy is to treat the lack of response as an uncertainty in the task execution, and it might be better for the system and agent to view this lack of information more positively and concretely as a ‘known unknown’. For example, the planning agent might create a plan (or multiple plans) that take into account the possibility that these assignments may transpire to have been accepted or rejected (at least in the player’s thoughts). Similarly, the interface might make visible the uncertainty, and perhaps also highlight the other information which could be impacted as a result, such as contingent plans.

Interactive system researchers have outlined a list of strategies to tackle the inevitable uncertainties, one of which is to reveal the uncertainties to operators so that they can make informed decisions [Benford et al., 2006; Skeels et al., 2008]. One straightforward approach for visualizing uncertainties is to highlight the affected data and various methods for highlighting have been proposed in literature [Conti et al., 2006]. However, in a real-time control scenario like AtomicOrchid, the highlighting approach has to be applied with caution. The operators have limited resources (attention, limited time and cognitive workload capacity) to tackle various issues in a multi-threading, time constrained task setting. Abuse of interface highlighting may contend for the limited attentional resources. There are evidences that visualization of uncertainties is likely to overload decision makers in time limited decision-making, which in turn, degrades their ability to respond in a timely manner [Zuk and Carpendale, 2007]. Take episode 3.5 as example, although the system highlighted potential task interruptions as a result of lack of response, the operator still

ignored it.

In episode 3.6, the HQ become confused when the planner stops giving assignments. In this case, the HQ requests plans four times and complains about the lack of response from the planner. However, the planner does not produce more assignments simply because the idle players have too little health to take any more assignments. This policy and reasoning of the planning agent is not visible to the HQ player (although in this case the reasoning leads to not doing something, which may have its own representational challenges).

8.5.4 *Support human planning*

As seen in all episodes, HQ players are observed to utilise the task interface to assess current game status, while in episodes 3.3, 3.4, 3.5 and 3.6 we have also seen how they can modify the agent’s plans. This suggests that the interface is sufficient in providing basic situational awareness for HQ players to make their own plans. The drag-and-drop based plan editing interface also enforces various constraints on task assignment so that all plans are at least valid, i.e. well-formed. For example, each player and each target can be assigned to at most one task, and each task can only have players with the correct combinations of game roles for the target. The interface also highlights players and targets on the map when they are manipulated so that the HQ player can readily assess location and proximity when editing task assignments. In the view of situated action [Suchman, 1987], the plan for the team is itself a projective representation of actions. What the system achieved here is to enhance the representation on the interface level and open an opportunity for human teams to sketch on the representation with situational awareness support. However, the observations also reveal some potential for improving support for human planning.

Returning to episode 3.4, the HQ player massively revises the plan. The modified plan turned out to be undesirable as it leads to 2 assignment cancellations and 2 player deaths (Episode 3.6). We observe that the planning agent is silent with regard to the problems in the player's proposed plan. As the planner has ruled out the risky plan in the first place, the reasoning behind the original plan may be important information that should have been exposed to HQ to inform their decisions. While making visible the planning agent's reasoning might have discouraged the player from changing the plan so dramatically, there will still surely be situations in which plans could or should be changed. And in this case the player is currently left to "do their best"; in short, the current system will make plans for a person to change or approve, but will not help that person to plan beyond ensuring that their plan is well-formed. If, rather, the planning agent were to simulate (and perhaps extend) the proposed modified plan then it could provide the HQ player with at least one view of the possible outworking of their plan.

In the current system the agent performs forward planning, i.e. it considers what field players might do in the future, not just in the current/next task assignments. However this information is also not made available to the HQ players. In episode 3.1 we also saw one of several examples of the HQ players also planning for future task assignments. However they had no way to record this or feed it into the system; they simply had to make a note or remember what they were thinking when the current task was completed and they had the chance to check and intervene. For at least some situations it might offer benefits if the agent's future plans could also be viewed, and if the HQ players also had some way to support their own future thinking.

However, again we are considering attempting to visualise and interact with significantly more data and more complex data which is liable to further complicate the interface. For example, for the agent to share information about its reasoning behind a particular task assign-

ment, it may need to present its prediction of the future game state and its assessment of current game status. It may well be problematic for the HQ to digest this extra information in a time-constrained task environment. Various studies have identified the problem of information overload in system interaction design in the DR domain [Carver and Turoff, 2007; Turoff and Chumer, 2004]. The problem occurs when information is presented at a rate too fast for a person to process [Hiltz and Turoff, 1985], and it may lead to a list of human performance consequence [Hiltz and Turoff, 1985] including failure/inaccurate response, systematically ignoring input, and input recoding (to more compact form).

Various techniques have been explored to alleviate the problem on information overload in different application domains such as security monitoring [Conti et al., 2006], teleoperation [Kadous et al., 2006], and communication systems [Hiltz and Turoff, 1985]. However, the specific research of information overload for task allocation support in DR domain is still rare. There is a danger that the extra forward planning support may introduce the risk of overloading human operators. Therefore, more studies are required to evaluate the trade-off between its benefits and consequences for human performance.

8.6 LESSONS FROM INTERACTION DESIGN

Herein, we offer a brief summary of the lessons learnt from this third study that may benefit the designers of planning support systems. These may be particularly relevant for settings in which timely human decision-making is critical. We furthermore also set our lessons into context with related work.

8.6.1 Common ground

Common ground is a critical requirement to making collaborative decisions in an effective and timely manner. Through our field trials, we have identified the following features as constitutive to common ground through providing a mutual situation awareness for the participating parties (HQ, field responders, and the agent).

- **Shared representation.** The task interface enabled common ground between HQ players and the planning agent – the basis for enabling HQ players to read, modify, and confirm the agent's task assignments. Shared representation was also critical to align field players and HQ view of the environment, e.g., through the messaging channel.
- **Awareness of current state.** Episodes evidenced HQ displayed awareness of task progress, threats, and field responders' locations and proximity as critical in their planning.
- **Domain-specific task representation.** Providing a consistent representation that made use of domain-specific visual cues was an important feature to enable alignment and consistency across views and between mobile (field) and stationary (HQ) representations.
- **Awareness of future actions.** HQ engaged with *proposed* tasks – reading proposals in conjunction with awareness of current state, modifying, and confirming was seen to be an essential aspect of the situated planning work.

Our recommendations also align with the theoretical framework model of situation awareness proposed by Endsley [2001], which argues that the situation awareness needs to be supported by three levels including 1) Perception of the elements in the environment 2) Comprehension of the current situation and 3) Projection of future status.

8.6.2 Supporting Mixed-Initiative planning

Some interactional challenges have become evident in the field trial that can be attributed to the interaction with the computational agent.

- **Constraining task assignments** was seen as a clear advantage of the planning support in this setting, to guarantee only well-formed task assignments.
- **Making reasoning visible.** Modifications of plans as a result of which responders were sent in harm's way may have been avoided if the grounds upon which the agent computed task allocations had been available for inspection. The challenge is how to implement this without inundating the operator with information.
- **Feedback on modifications.** As a further step to making visible, the agent could also provide feedback on human modifications of its task assignments, e.g., in terms of safety and risk.
- **Forward planning** of the agent could be made visible on demand to enable planning ahead.

Our recommendations echo seminal work on human considerations in context-aware systems, which provided principles to support intelligibility and accountability [Bellotti and Edwards, 2001]; similarly we stress that the goal for planning support systems should be to be accountable for their actions, therefore, 'what they know, how they know it, and what they are doing about it' [ibid., p. 201] needs to be legible by the operator. Furthermore, as planning is oriented towards the future, yet produced as a contingent, situated activity [Suchman, 1987], the interface needs to support revision and revoking of plans in situ, and furthermore provide the situational awareness essential to do so.

8.6.3 Interactional trouble

We have encountered the following interactional challenges that likely generalise more broadly to related settings.

- **Complacency** describes the phenomenon whereby occasional failures of automation may be difficult to detect [Kaber and Endsley, 1995]. Particularly, when the operator has learnt to trust the computational component. Mechanisms to counter this may turn towards more human involvement in the planning, however, there will be a fine line so as not to create information overload.
- **Non-responsiveness** has been observed to create uncertain situations. It is important to realise this could equally be caused by technical communication outages, as well as by human non-response. Designers may attempt to incorporate this as 'known unknowns' in the system, for example planning could be done with an estimated *probability* of a positive response.

The wealth of computational approaches to the problem of 'planning under uncertainty' [Chang et al., 2007] suggests uncertainty is a key recurring problem, particularly in time critical settings. In this context, it may also be important to look towards interface design guidelines to complement computational approaches in order to avoid or minimise complacency and non-responsiveness at the interface level [Norman, 2013].

8.7 PROFESSIONAL FEEDBACK

This section will present the results of a workshop we conducted with a professional Disaster Response charity called Rescue Global (RG) (Section 4.5). The purpose of this workshop was to get professional feedback about the AtomicOrchid(AO) platform to understand the realism of the disaster simulation based on AO. Based on the feedback, we reflect on the strength, limitations and potential improvements of

AO platform.

The workshop was centred on a demonstration of the HAC-ER system for AtomicOrchid (AO) (Section 2.3.3). The AO is a component of the HAC-ER demo system. Other components include information crowdsourcing, multi-UAV control interface for DR, and DR operation provenance store (Section 2.3.3). Each component was allocated one session (about 45 mins). The AO session was structured as follows:

1. A presentation introducing the AO platform and HQ interface.
2. A hands-on session for RG members to operate the HQ interface. For the purpose of the demo, the field players in AO were simulated.
3. A discussion of the AO platform for Rescue Global to reflect on some themes including: Command and Control structure, Division of Labour/Task and team planning.

We also encouraged free discussion throughout the AO session, which led to some emerging themes including: data connection, time-scale of planning in real DR operations and other interface improvement suggestions.

8.7.1 *The Rescue Global feedback*

Based the feedback from RG, the AO setting mirrors some aspects of RG's disaster operations. Firstly, the RG members quickly recognise the task-team matching mechanism is very similar to aspects of resource-needs matching in their operation. At the beginning of the introduction presentation of AO, one of the responders commented:

"This is actually very similar to what we do when we do reconnaissance now. We identify needs, and we are aware

of the resources that we have and match the needs to resources."

Further, the team formation mechanism could also be mapped to aspects of their operations by RG members. RG confirmed that members from different organisations and with different skills often teamed together for tasks. However, cross-organisation at teaming activities are mostly done in an informal way, as one RG member described:

"What we did in the Philippines is that , say we found a school, and we talked to the MSF medics ¹ we have bumped into. They have medics, can run some clinics there. Then they need some security, they are comfortable with Simon and Bren (two path finder in RG), we can help them out, we did that, but almost in an informal way"

The statements of RG indicate that their collaboration with the MFS medic teams are based on an informal relationship with them and the formal organisational collaboration is often missing. RG also stressed that the paradigm of cross-organisation team formation in AO is desirable, but difficult to achieve in a formal way. They highlighted that the major challenge for realising the AO paradigm is organisational. As one RG member stated:

"The challenge with that is that UN clustering system, makes separation of different tasks, they then delegate the tasks down to say Health cluster, and UNISAF, all the way to little agencies, they won't work with anyone else, because they do not know them, they do not have relationship, so UN structure stops that... does not mean you can not change it, but the current paradigm would not allow it "

¹ Medecins sans frontieres, rescue charity, <http://www.msf.org.uk/>

Based on RG's feedback, we also identify some unrealistic aspects of the AO setting. Firstly, RG member expressed concerns related to data connection. They highlighted that the activity in AO requires good connections between field and HQ, which is not always the case in a disaster situation. Given that, RG members can still imagine a version of AO being used in situations with poor data connection. They suggested that the information in the field can be collected and input into the planning support system in a slow and manual way (e.g. through a satellite phone call). In that situation, they still think planning support aspects of AO would be useful to them, though the pace of planning and information collection would be much slower and manual.

“say OK, there is not live link between HQ and field responder, could you have that, in the field people can start sending reports, which could be done like satllete phone, then HQ can load on to this (AO), come up with suggestions, because can not task directly to the field , there is another call, it puts in lots of gaps and pauses, but still be able to have the system giving suggestions based on manual input, which do not require broadband. ”

In terms of Command and Control (C₂) structure, RG frequently refer to two frameworks of C₂, the Bronze, Silver, Gold model (Section 2.1.2) and the National Incident Command Framework ([HM Government, 2008]). The assumption of Command structure in AO seems to be detached from both NICF and BSG in that both NICF and BSG have several hierarchical command levels (Bronze, Silver, Gold) and ways to sectorise units (geographical, functional). In the AO setting, both the command hierarchy and unit sectorisation are missing. At the interface level, the information presented to each command level may need to be tailored to their specific paradigms (operational, tactical, strategic):

“For operation, we would like have interface with 2-3 levels of sophistication, that takes you into different paradigm.

For example, you do not what simple transactional barrier for your section commander,... you do not section commander know that. Simple data (at the bronze level), enhanced data (at silver level), at gold, stripping away all details."

From feedback, we also noticed that the time-scale of RG's operation is much longer than AO. As one RG member described how tasks are allocated in a multi-agency operations:

"The amount of work we do here , is probably a month of work from what we saw in Philippine, like 15-20 job orders, and that take about a month.... The problem is how they are all disconnected, you got tightly coupled system with lots of moving part, with massive gap between each one, so say example, that cluster of systems, you have that in , that in ..., that in another, they then meet once a week, can be once a week and then get information, disseminate and give out jobs, it is just unbelievably slow."

Although the time-scale of AO is unrealistic for RG's typical activities, they think that AO can be seen as an ideal, speeded-up version of multi-agency operation, because it automates lots of manual process in terms of information capture and presentation and promotes commonly recognised information picture across agencies:

"What is good about this, it is getting towards automating and visually representing what we manually constructing, with posts, white board and flip charts. It is not because what we are crap, it because everyone in the world does that ... That is why commonly recognised information picture is so important, you essentially got, here is what happen, here is our resource, then you can make decision."

Apart from information capture and presentation, the RG team also saw the value in AO as a tool to balance work between field and HQ:

"I think it ([AO](#)) could be useful that you can use to help people in the field to gather information, file it back and forget it, which is really helpful, because otherwise what we have to do is we do the plan do the resource allocation in the field which is really time-consuming, if we say field get information and send it back, HQ receive and plan based on them."

The statement indicates that planning of resource allocation does happen in the field and it is desirable to transfer the workload to HQ. The statement also identifies [AO](#) as a (potential) tool to balance division of labour.

8.7.2 Conclusion and reflection

In the previous section, we have reflected on both realistic and unrealistic aspects of the [AO](#) game based on feedback from the [RG](#) workshop. The resource target mapping and teaming aspects of [AO](#) setting are found to be similar to activities in [DR](#) operations, though the teaming in [DR](#) operations is likely to be conducted in an informal way. On the other hand, both the time scale and command structure in the [AO](#) do not match what currently happens in [RG's](#) [DR](#) operations. The time scale of [RG](#)'s operation is much longer than [AO](#); The Command and Control framework in [RG](#)'s [DR](#) work has hierarchical levels and sectorisation for some levels, both of which are missing from [AO](#). Further, the assumption of good data connections in the field are also found unrealistic. However the [RG](#) team thought that the core planning support aspects of [AO](#) would still be useful even when data connection is poor.

Although [AO](#) was not built as a tool to support real disaster operation (it is designed to be a simulation game), the [RG](#) team still found some aspects of [AO](#) demonstrates elements of a future planning support system that would be desirable to them. Firstly, [AO](#) portrayed a picture of automated information capture and visualisation which is

appreciated by RG. Secondly, RG also appreciate that the division of labour facilitated by AO, in which the HQ (with the planning agent) handles the task planning while the field responders just need to “gather information, file it back, and forget it”

Elsewhere in this discussion (Section 8.5) we have touched on several areas where more information might in principle be shared between the planning agent and the HQ players, such as agent priorities and reasoning, uncertainty and future plans. We note that would almost certainly make the interfaces and visualisations more complex and increase the risk of information overload. However, if the system were operating at the kind of pace that Rescue Global use to characterise their own activities then there would be a lot more time to work with than in the current version of AO. So a challenge for future work is to understand more about the interactional challenges of this kind of time-constrained but relatively slow-paced situation.

8.8 SUMMARY

The objective of this field study here was to explore the design issues surrounding an in-the-loop interaction pattern. The exploration was conducted through a field trial of an AtomicOrchid (AO) probe with in-the-loop support, with particular focus in the planning activities in the control room. The Game probe is evolved from the on-the-loop version of AO that was trialled in study 2. To realise a human in-the-loop interaction, the AO system allows the human HQ to approve, edit and cancel the agent generated plans.

Findings from interaction analysis of field observations, triangulated with log files, reveal how the in-the-loop interactions played out. Our observations indicate that the human coordinator and the automated planning support can successfully work together in most cases. Supported by a task assignment interface, human coordinators take active responsibility in the system by inspecting and ‘correcting’ the

agent proposed plans. However, occasional failures of planning are also observed due to a number of issues including 1) complacency 2) silent, missing or invisible information 3) and limited system support for human planning. Several of these observations suggest additions to the information to be conveyed by the interface, but at the risk of information overload for the people in the HQ.

Based on the findings, we also summarised a list of lessons for the in-the-loop interaction design. Firstly, the importance of common ground for human system coordination is highlighted and we suggest that common ground should be supported by domain-specific information models, appropriate visual representations of the models, and articulation of future actions. Secondly, several interface features are highlighted to support mixed reality interactions, including constrained task assignments, visible plan reasoning, feedback on task modifications and forward planning support. Third, our observations reveal several interactional troubles including complacency and non-responsiveness that need careful consideration.

In addition, a workshop with a professional organisation has revealed strengths and weaknesses of the **AO** simulation. In more detail, the resource target mapping and teaming aspects of **AO** setting are found to be similar to activities in their **DR** operations. However, the time-scale, command control structure and the assumption of data connectivities do not match their experience of current **DR** operations. Although several aspects of the **AO** simulation does not match **RG**'s particular operations, the member of the organisation still appreciate the core planning support aspects of the **AO** system.

Part IV
CONCLUSION

CONCLUSION

To sum up, this PhD work uses a mixed reality game, AtomicOrchid, as a testbed to probe interactional issues, and to develop implications for agent planning support system applied to disaster operations. Field trials were conducted for three versions of the AtomicOrchid game with different interaction design patterns (Figure 57). The first study was a non agent version of the game. The second and third studies probe two interaction design patterns, on-the-loop and in-the-loop). Video and system logs of the field trials are collected and interaction analysis is conducted to reveal interactional issues, generate design requirements and interaction design implications.

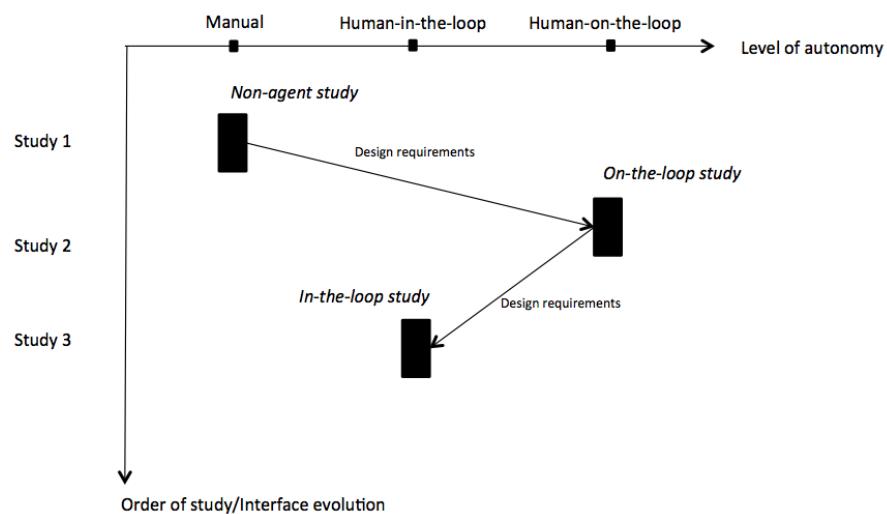


Figure 57: Overview of studies

The following sections summarise the contributions of the thesis, and conclude the thesis with limitations and an outlook on future work that this thesis might inspire.

9.1 SUMMARY OF CONTRIBUTIONS

Overall, the thesis has three key contributions to the interaction design of planning support for Disaster Response teams:

- A **The system prototypes.** The research produces real-world interactive prototypes (i.e. AtomicOrchid) for investigating human-system interaction in a Disaster Response settings. An iterative prototyping process is conducted throughout the three AtomicOrchid studies, which results in three interactive prototypes of planning support systems. On the one hand, the prototyping process is guided by the the interaction patterns generalized from LoA models (3.2.1.1). On the other hand, the system evolution is also inspired by key observations from iterative field trials.
- B **The interactional issues.** The field observation of the AO trials generate thick descriptions of human system interaction in a DR scenario. The key observations reveal both interactional issues which inspire system evolution across the three trials (Figure 57).
- C **The design implications.** For each study, field observations are further analysed to generate design implications which contribute to future development of automated planning support systems in the complex collaborative work setting of DR.

The section 9.1.1 gives details about the system evolution and the interface design rationale, followed by a summary of key observations in section 9.1.2. The interactional issues that emerged from field observations are detailed in section 9.1.3.

9.1.1 Interface evolution

This section presents the interface evolutions of AtomicOrchid (AO) across three field studies and the design rationales behind them.

The initial interface of AtomicOrchid (used in 1st study) is designed to provide basic coordination support by displaying game status and providing a broadcasting channel for communication. The control room is manned by 2-3 HQ players. Every HQ player uses the same interface, which shows the game status on a map. A chat box is provided to send and receive broadcast messages. A mobile interface is used by field responders. The responders share the same map with HQ, which displays game status with the exception of the radioactive cloud. The messaging interface is placed in another tab (Figure 58).

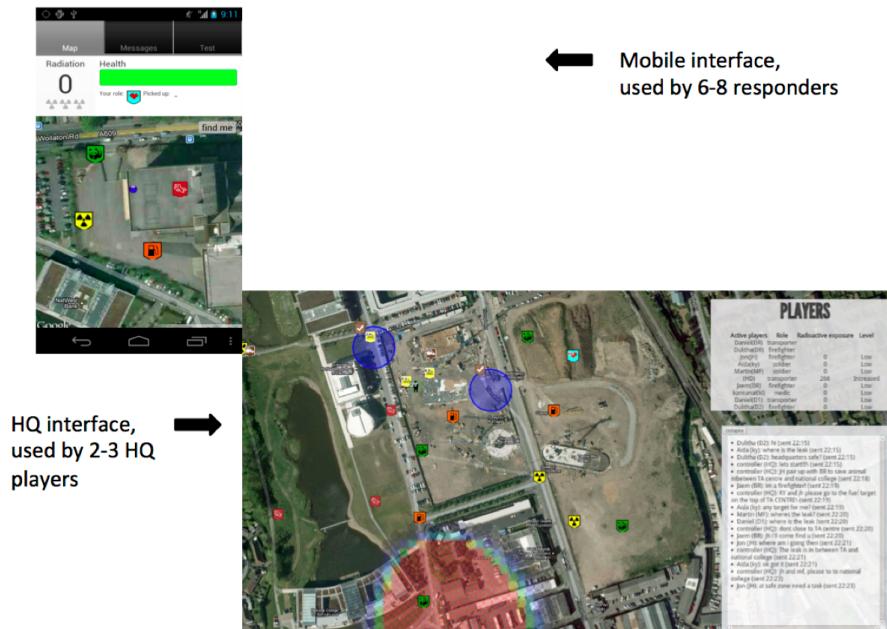


Figure 58: Interface of study 1

The 1st study discovered several flaws in the initial interface, which led to a number of design requirements:

1. The interface should support quick acknowledgement and feedback in communication channel.

2. The interface should show freshness of messages in the communication channel.
3. The interface should provide support for geo-referencing.

The second prototype is designed to integrate a planning support agent with an on-the-loop design pattern (Section 4.4), in which HQ only monitors task allocation and execution with infrequent intervention. A set of changes in the second version (used in study 2) of AtomicOrchid are partly inspired by the requirements from the study (Figure 59). In this version, all targets are identified by unique task IDs to support referencing. Text messages in the broadcasting channel are labelled by time stamps to flag potentially outdated information. With the introduction of the planning agent, the interface also supports quick feedback to the agent with a single press. A new tab (“task”) in the mobile app also shows a text-based description of the agent task-assignment. In the HQ interface, the agent assignments can be revealed on request by the HQ clicking on “show task” button).

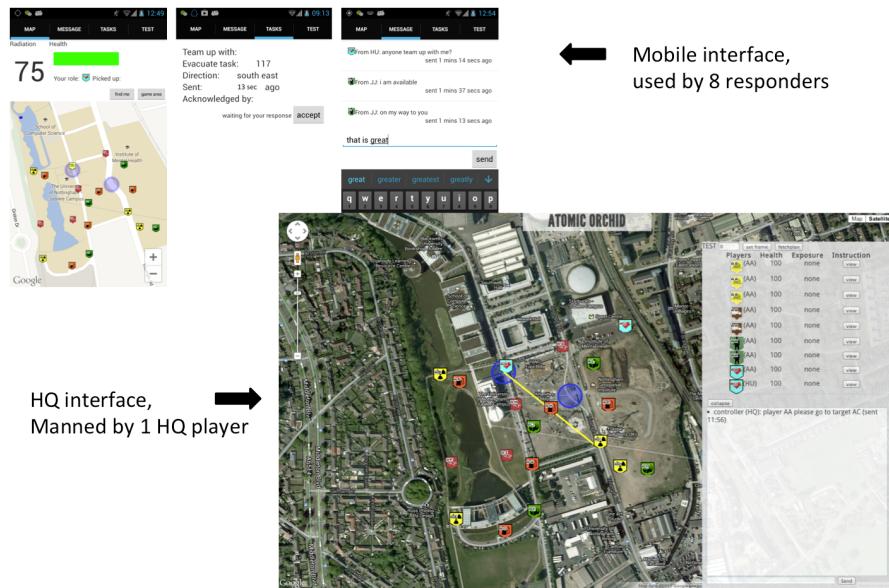
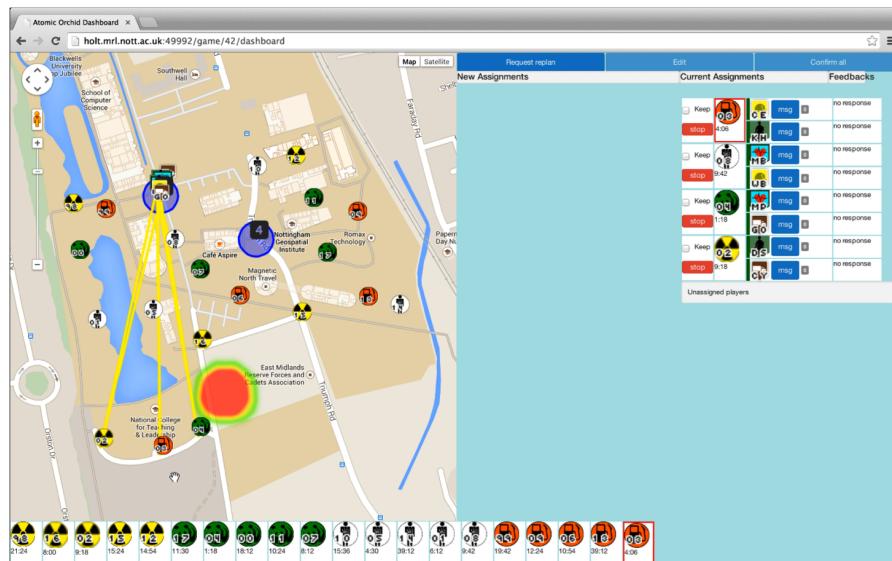


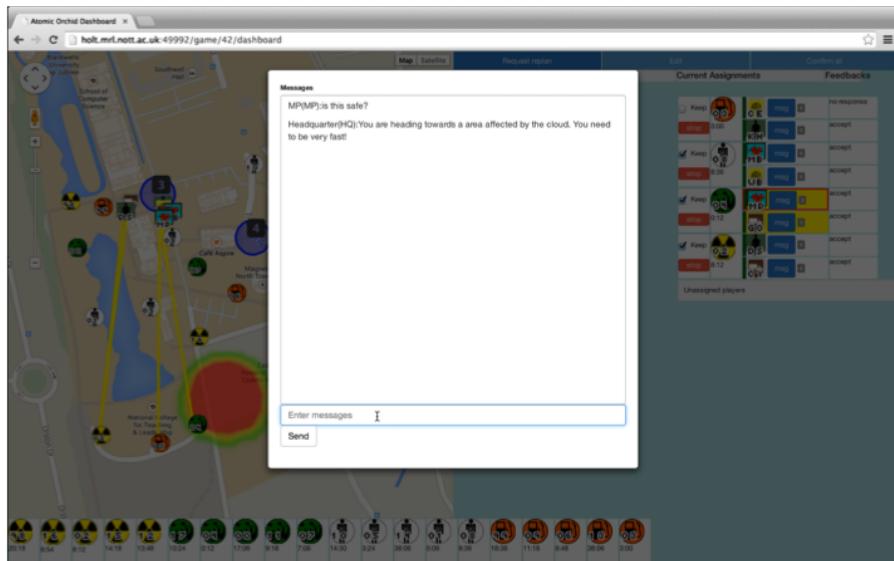
Figure 59: Interface of study 2

The results of the second study revealed that interface level support for HQ intervention is very limited in the on-the-loop version of

AtomicOrchid.

This kind of intervention support is also essential for the in-the-loop study (the third). Therefore, a task assignment (Figure 6o) interface is introduced for the third study to enhance the HQ's ability to intervene in the task planning. Task-specific communication channels are also introduced to avoid information overload in the broadcasting channel. Compared to previous HQ interfaces, the operations on the task assignment interface are more complicated. Therefore, one HQ player is dedicated to operate this interface and another player is provided with the same HQ interface as the one in the on-the-loop version to provide situational awareness and handle broadcast messages.





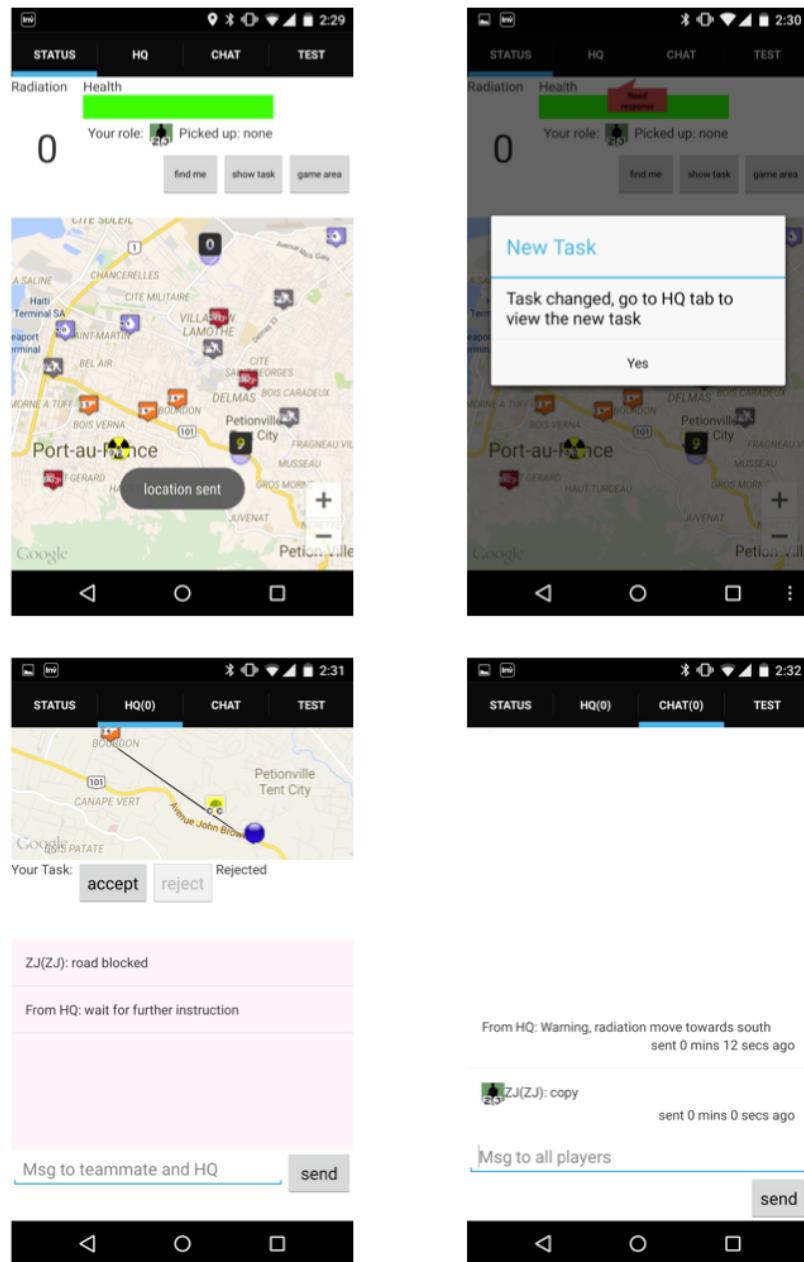


Figure 62: Mobile Interface of study 3

9.1.2 Key observations

This section gives a brief overview of the key observations from the three studies. First, the non-agent AtomicOrchid trial (Study 1, chapter 6) aims to give insight into how humans conduct planning (without agent support) in the time and spatially constrained disaster settings and to generate general requirements for a coordination support system which can be applicable to the next two agent-integrated ver-

sions of AtomicOrchid. The result of interaction analysis showed that team planning was dominated by local coordination between field players in a situated manner. The field teams managed to organise their team and task allocations by utilising available resources such as local conversation, mobile interface, and remote messages. The HQ was observed to successfully provide awareness of the “danger zone” to the field teams through remote messages. However, HQ had little direct influences on the planning and actions of field teams. One potential reason could be the breakdown of communication between HQ and field responders. The communication breakdown is thought to be caused by a set of factors including communication modalities (text vs voice), training level of players and HQ and system interface designs.

In study 2, a planning agent is integrated to the system with the human on-the-loop interaction design. Through interaction analysis, we gain insight into the division of labour between human and agent (Section 7.1) in which the agent takes over routine planning activities while the human focuses on other issues such as finding teammates, targets and choosing the best routes. However, there is also evidence showing the agent planning occasionally interrupts the workflow of the human team potentially because it fails to consider social cost of task changes. We also observed HQ players struggle to influence the plan because of the lack of interface level support. Further, a set of misconceptions in feedback loops are also observed (Section 7.5.3).

In study 3, the system is evolved to facilitate in-the-loop interaction with the feedback from study 2. The main changes are a number of interface functionalities which enable HQ to approve and edit agent planning and monitor player feedback. Through observing the usage of these new functionalities in the control room, we observe an expected division of labour in which:

1. The HQ decide when to perform re-plan.

2. The HQ review every agent instruction for routine task planning.
3. The HQ deals with player feedback.
4. The Agent proposes task allocation.

Field observation show that division of labour is at least partially facilitated by the task assignment interface deployed in study 3 (Section 8.2). The task assignment interface appears in many cases to provide an effective shared representation of the current state of the game. As well as showing current player and target locations and player health it also makes visible the currently approved task allocations, field player responses and any new plan that has been requested or is being edited. This shared information forms the common ground between the HQ players and the planning agent. HQ players are also observed to closely monitor this view and its representation of plan execution to gain awareness of the cloud's location in relation to players and current task progress. HQ players also engage actively with proposed (as opposed to current) task assignments. HQ players are quite capable of modifying the agent's plans when they wish to, for better (Episode 3.3) or worse (Episode 3.4).

Analysis of some failed cases of coordination also points out a number of interactional issues. The details of the issues, possible solutions and design implications are discussed in the following section 9.1.3.

9.1.3 *Interactional issues and lessons*

There are a number of interactional issues that emerged from the field observations. This section reflects on interactional issues related to 4 themes including division of labour, common ground, accountability and support of situated actions.

9.1.3.1 Division of labour

The on-the-loop and in-the-loop interaction designs have been trialled in two studies. To recap, the main distinction between the two interaction designs is the extent to which the human HQ is involved in routine task planning. For the on-the-loop design, the system is supposed to run without human input, while the in-the-loop system can not operate without constant human interaction. Guided by these two patterns, detailed system design has been implemented.

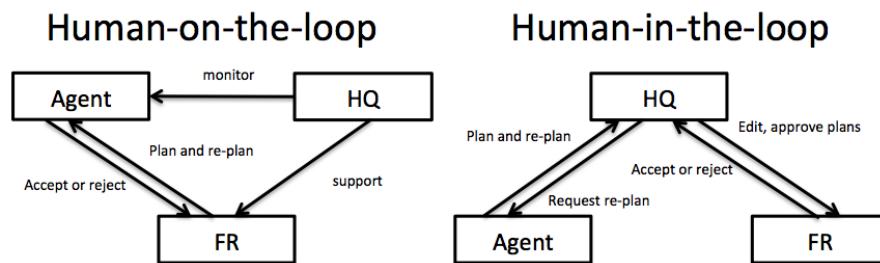


Figure 63: On-the-loop vs in-the-loop

There are several performance differences between the in-the-loop and on-the-loop studies. Direct comparison of performance between in-the-loop and on-the-loop is not possible because the study 2 introduces a number of improved interface features (Table 11) that influence team performance as well. However, the difference of performance can still suggest some implications for interaction design.

- 1. In-the-loop design is an option to reduce unnecessary task interruptions**

As we have summarised in section 8.5.2, there is evidence that the On-the-loop design is more likely to cause extra re-teaming and task interruption when compared to the In-the-loop design. In study 2, HQ has been observed to deliberately avoid unnecessary task interruptions and team reformations caused by agent planning.

It can be argued that the difference is simply caused by a lack of reliability of the agent. Some AI researchers have attempted to enhance the reliability of the planning agent in handling so-

| Division of labour | | |
|--------------------|--|---|
| FR | 1. Interpret instructions 2. Organise team meet-ups 3. Deals with navigational issues (Locate target, find routs) 4. Provide feedbacks (accept/reject) | 1. Interpret instructions 2. Organise team meet-ups 3. Deals with navigational issues (Locate target, find routs) 4. Provide feedbacks (accept/reject) |
| HQ | 1. Monitor and chose to intervene (Edit when necessary) 2. Sending other information in broadcast channel | 1. Initiate re-plan 2. Review every agent instruction 3. Send information in broadcast channel 4. Deal with feedbacks from field responders |
| Agent | 1. Send task allocations directly to field responders 2. Initiate re-plan 3. Deal with rejections from field responders | 1. Send task allocations to HQ for approval 2. Perform partial task planning based on constraints specified by HQ |
| Interface support | | |
| | | 1. Representation of game status 2. Broadcasting message channel 3. Task-specific message channel 4. Field responder feedback highlighting 5. Support for plan edits (Drag and Drop interface) |

Table 11: On-the-loop vs in-the-loop

cial situations (Section 3.2.2.1) with the approach of plan/intent recognition. For example, some may believe that a planning agent with better user/environmental modelling techniques can factor in the social process involved in task changes to inform plan generation. However, the work of Suchman [1987] acknowledges the extreme difficulty of plan/intent recognition because of the weak logical inferential connections between a human's intent and actions in situ. Therefore, we believe that the in-the-loop design may be a useful approach to overcome the agent's limited reliability and utilise its capability at the same time.

However, the in-the-loop design may also have its own cost. On the one hand, it increases the workload of HQ players because they need to check each instruction proposed by the agent. HQ players sometimes failed to recognise task interruptions even with interface highlighting support, which suggest that HQ may not be able to eliminate every unnecessary interruptions. Effectiveness of the HQ may also depend on a balanced workload. In a multi-tasking control room environment, it is critical to balance the workload of the HQ players. Therefore, we need to carefully consider the trade-off between the increase of workload and the reduction of unnecessary task-interruptions

2. HQ's ability to intervene in agent planning is required for both in-the-loop and on-the-loop design pattern.

Compared to the in-the-loop study, the HQ players in the on-the-loop study are observed to struggle to intervene in the planning process, and we believe the performance differences are at least partially linked to the distinction between in-the-loop and on-the-loop. In the on-the-loop study, the only way for HQ to intervene in the planning is to send unstructured text messages in the broadcast channel. HQ's ability to intervene has been greatly enhanced by a set of interface support introduced in the in-the-loop study. Some of the interface support is inspired by

the implications from the on-the-loop study. It highlights the need for interface support for HQ intervention because both in-the-loop and on-the-loop appear to require HQ get involved when necessary.

3. Human involvement may also incur system planning failure.

Compared to the on-the-loop design, the in-the-loop design supports greater human involvement in plan generation. However, some planning failures appear to be directly linked to human involvement in the in-the-loop design. For example, there is a case of player "death", which occurred when field players were trying to follow a human generated plan in the in-the-loop trial. Although there are lots of factors contributing to the 'death' case (e.g. insufficient training for field player, communication breakdown), the task assignment is thought to be too risky and not recommended by the agent in the first place. It is believed that the chance of human error can be reduced by appropriate interaction design, which helps to establish the 'common ground' between the human and the agent. However, such a design would not be an easy task as it involves multiple design trade-offs (Section 9.1.3.2).

To sum up, task interruptions are more frequent with the on-the-loop design (compared to the in-the-loop study). This fact highlights the agents' inability to appropriately take into account the social process in human teams. Following the view of Suchman [1987], we believe that human's intentions and actions *in situ* are extremely hard to be modelled appropriately. Thus, the in-the-loop interaction design maybe a much more feasible approach to overcome the limited reliability of the agent, however, at the expense of higher HQ workload. Secondly, the interface support for HQ intervention has been proved to be important in both in-the-loop and on-the-loop settings. Finally, a human involvement in planning (in-the-loop) may incur some planning failures. This issue highlights the need for effective information sharing between human and agent, which helps to establish the 'com-

mon ground' for coordination.

9.1.3.2 Establishing common ground

The "common ground" (or as Suchman [1987] suggested "mutual intelligibility") is the basis for human system interaction [Bradshaw et al., 2011]. While, "common ground" may be improved with plan/intent recognition technologies (e.g. user modelling), this approach, if possible, can be extremely difficult (Section 4.1). The author believes that appropriate interaction design can enable human operators to understand and influence the planning agent. This section summarises the interaction design implications concerning the establishment of "Common Ground".

In both study 2 and 3, the agent behaves like a "black box" which simply provides the planning results (task allocation) for next immediate step. The interface only exposes the results of the agent for HQ to monitoring. Through our observations in chapters 7 and 8, we find there could be some extra information shared between human and agent to improve planning:

1. The HQ players are observed (section 8.5) to occasionally make some forward planning for field players. Because the forward planning is also performed by the agent to derive the current plan, this information could be shared so that the HQ players can be provided with the agent suggestion when doing forward planning.
2. The reasoning behind current task allocations would also be usefully shared. In chapter 8 , HQ is found to override agent plans which sometimes lead to undesirable results. The HQ is also found to be confused when the agent stop assigning tasks to players with low health. Exposing internal reasoning of the

agent may help the HQ to make informed decisions.

3. Misunderstanding between agent and human is also observed in the feedback loop in study 7. Firstly, human responders do not know how the agent is going to handle the rejection. They try to use rejection to revert to previous tasks, while the agent will give them more new instructions. Secondly, it is not apparent to field responders that their rejection will cause replanning for the whole team which can lead to costly task interruptions. Therefore, information indicating consequence of interface interactions should also be made available to human to facilitate accountability and ensure informed decisions.

Although we have identified a range of information which may be missing for establishing true "common ground", presenting this information could also be challenging. The information should be delivered in the right form (e.g text, visualisation, dialogs) and at the right time (e.g. pop up, on request?) [Carver and Turoff, 2007]. Especially in the multi-tasking, time-critical settings such as AtomicOrchid, multiple sources of information can compete for the attention of the human operator. Information overload could be a real danger of interaction design in this setting [Lieberman and Selker, 2003]. Therefore, ways to expose the agent's internal information and its human performance consequence need to be further studied.

9.1.3.3 Facilitating accountability

The observations shows that both the spatial divide and the existence of planning support can be the factors that affect the natural accountability of member's activities.

In the 1st study, players heavily rely on local coordination for task planning. In a co-located setting, players can naturally make their actions observable and accountable to each other through conversa-

tions, body languages, gestures etc. In the distributed work setting, the spatial divide makes accountability of member's activities opaque to each other. The technology support may need to play a role in supporting member's accountabilities. However, the reliance on local coordination indicates a lack of remote coordination support in the first version of the AO system. A set of functionalities including GPS/map sharing and broadcasting has been provided. We have observed players utilise the functionalities to make sense of other team members' actions (see section 6.4.3) and act accordingly. However, coordination with remote players is still problematic, which can be evidenced by frustrations related to remote communication (Section 6.5.2). Therefore, we suggest remote coordination support should be built in a way that facilitates accountability across the distributed team by making player's activities observable and reportable.

In the second study, the issues of members' accountability are further complicated by the existence of planning support. Individual player's interactions with the planner agent (reject/accept plans) have impact on the planning of the whole team. However, while the rules of social conduct ensured accountability of action among co-located teammates, we found the impact of rejections on remote players was not properly appreciated; nor did the interaction design support making these individual human-agent interactions accountable to the whole team. We believe the interaction design should reveal the hidden cost of certain actions (e.g., rejections) help to make local decision making accountable to remote team members, ensuring consequences of local decisions for the welfare of all teams are understood.

9.1.3.4 Supporting situated action

The thesis takes the view of situated actions proposed by Suchman [1987]. The problem solving agent necessarily produces plans in an attempt to guide human teams, while humans act in a situated manner. There is a need to design interactions that bridge the gap between the planning model of the agent and situated actions of the human. The

observed human actions in the studies are well aligned with the view of situated actions. In the non-agent version of AO players are able to organise their activities with all their available resources (local conversations, messages, game status from mobile interface) without reaching a “plan” for the whole team. In the second (on-the-loop) study, players are observed to utilise the agent plans as an extra resources for taking action in various social situations (Section 7.5.2). Based on our observations, we summarises some recommendations for supporting situated action.

1. **Presenting agent plans as a resource for situated action.** The agent “plan” itself is a representation of actions and their projective effects [Suchman, 1987]. One key task of the system is to provide interface support for presenting the plans so that they can be easily utilised by people as a resource for deliberations on their actions. Through the three studies, we incrementally enhance plan representation. The final interface presents the agent plan in a structured way, alongside the current task status (player feedback and existing plan under execution) and map-based location visualisation. This representation turns out to be effective for HQ players to make sense of the situation and act on the task allocation and monitoring.
2. **Facilitating the situated actions.** We observed that the human operators are faced with various social situations and contingent circumstance (Section 7.5.2) in the field trials. The system may need to keep the flexibility to allow the human to proceed with their actions, no matter whether they choose to follow the agent plan or not. With the on-the-loop version of AO, the agent plans sometimes “gets in the way” rather than supporting players. The agent support system follows a strict planning model (plan-act-replan) without sufficient support for human involvement. Consequently, although the human can chose to “disobey” the agent (by rejecting, or ignoring), the result can

be the complete abandonment of the use of the agent or even more problematic confusions (e.g. Episode 2.4 in Section 7.4). On the other hand, the in-the-loop version of the interface provides flexibility for human action by providing interface functionalities for plan editing and the opportunity in the workflow for intervention. In this way, frequent human inputs are synthesized with the agent plan as a resource that support their situated actions.

3. **Supporting human-system mutual intelligibility.** By adopting the view of human agent interaction (Section 3.2.2), the support system can be treated as an equal team player, with whom the mutual intelligibility should also be maintained to achieve concerted action [Suchman, 1987]. Here we consider “mutual intelligence” and “common ground” as interchangeable concepts. Therefore we refer to section 9.1.3.2 for suggestions on improving mutual intelligibility.

9.2 LIMITATIONS

This section outlines some of the limitations of this PhD work.

By following a serious mixed reality game approach, the game AtomicOrchid is used to simulate some key factors of a distributed, time-critical task setting. Of course, fictitious game scenario can not completely mirror the setting of disaster operations. Firstly, the time-scale of planning is generally much longer in a rescue and reconnaissance mission (from hours up to days, Section 8.7), while AtomicOrchid game sessions only have a 30 minutes time-scale. Secondly, Disaster Response teams always prioritise critical tasks according to some specific code of conduct (e.g. tasks with human injuries may always be prioritised.). Therefore, the game objective of maximising targets saved may not match the real goal of a DR team. According

to professional feedback (Section 8.7), the communication modality (text, rather than radio) and the assumption of ‘constant connectivity’ are also not aligned with common real world environments.

Further, the lacking seriousness of the simulation may also be a limitation of the game approach. Participants are frequently observed to laugh and make jokes about their health values (life), which indicates that they see the trials as a recreational experience. Therefore, the participants may be less concerned about risks and life threats, when compared to a real disaster setting.

Untrained participants were recruited for the field trials. We anticipate that the behaviours of players in the game may change with increased training and experience. Further, the professional responders may also behave differently because they have high a level of training, real experience of DR operations, a different code of conduct and a distinct organisational structure.

In terms of data analysis, the result do not have quantitative robustness because only 6 game trials were conducted. This limitation is actually a limitation of the data analysis method we apply, namely interaction analysis. The method is extremely time consuming, thus limiting the amount of data that we can analyse [Crabtree et al., 2012]. With limited trials and participants, individual difference (of participants and teams) can significantly affect the generalisability of the study results. Consequently, the generality of the results may have to be established with other means. However, this limitation does not affect the descriptive power of the qualitative approach of the analysis.

Overall, the PhD work employed a game approach, which has a number of limitations in terms of the game scenario, participants’ attitude, participants’ selection and data analysis. We cannot claim that observations of human behaviours based on a game approach would be exactly the same as what would happen in a real disaster. However, we argue that the game-based approach reproduces to

some degree some critical factors of DR operations, including time pressure, distributed teams, and mental/physical stress. Compared to computational simulation, the field trials can reveal rich human-system interactions under time and spatially constrained task conditions. Therefore, we argue that the observations from game trials are still valid and can be used to generate design implications for future automated support systems.

9.3 FUTURE WORK

Following the limitations and contributions of the thesis, its impact can be extended with future work in several possible ways. Firstly, the planning support itself might be enhanced with other AI technologies. The agent in this study does not consider the social cost of re-teaming and task interruption. Other AI technologies (e.g., user modelling techniques, section 2.3.3) can be explored for modelling the human social behaviour and factoring it in to the plan generation process. However, when the agent is enhanced by new AI technologies, it is unknown whether the new capabilities would hinder or improve team coordination and how the interaction design should be adapted to support the new capabilities. Therefore, one future direction of this PhD study is to incrementally enhance agent capabilities and conduct observational studies to gain insight into the implications of the new enhancement on the interaction design.

Secondly, the game can be re-engineered so that the game setting can be better grounded in the real practices of Disaster Response teams. In order to do that, the close collaboration with Disaster Response teams may be required. The collaboration with Global Rescue happened at a late stage of this PhD work, and helped us to identify some potential improvement of the game setting. Firstly, the targets are not always equally important. For example, human injury will typically be prioritised. Considering different weighting on targets may make the game setting more realistic. Secondly, radio rather

then text messages is the communication modality normally used by operation teams, which may impact our behavioural observations. Thirdly, AtomicOrchid assumes constant connectivity between HQ and field responders, but disaster operations are often characterised by intermittent communication. Therefore, it would also be useful to simulate the intermittent communication in AtomicOrchid.

In a similar vein, the subjects of study might be changed to professional responders in order to better elucidate human behaviour in real disaster operations. As we believe that the trained professional responders may behave differently to the general public, it would be desirable to recruit professionals in the future work to compare with the findings of this PhD study.

In terms of the methodology, more field trials might be conducted in order to carry out statistically robust quantitative analysis. If there are more field trials, quantitative analysis can compliment the qualitative interaction analysis in some aspects. For example, the statistical differences in terms of game results (including players death, targets saved) can be used to infer effectiveness of interaction designs. Also, the social cost and workload of the HQ might also be quantified to give insight into the impact of interaction designs on the team performance.

The studies have highlighted the trade-off between establishing common ground and information overload (Section 8.5). It is believed that further studies can focus on exploring, how this the trade-off affects team coordination and developing guidelines for system designers to negotiate the trade-off.

Moreover, the interaction designs are not limited to on-the-loop and in-the-loop. The contribution of this PhD work can be extended by exploring some middle-ground designs. For example, the agent can be given the responsibility to decide when to perform a re-plan (as it is in the on-the-loop design), but also send the instructions to

HQ for approval (as it is in the in-the-loop design).

It should also be noted that centralised planning support is only one kind of agent planning support technology among many. For example, the planning support can also be decentralised. Planning agents could be built into personal assistant devices for every field responder. We believe that the change of agent technologies would have implications for interaction design. New interaction design patterns can be devised and adopted for changing of agent technologies. Therefore, this PhD work can potentially be extended by exploring different automated planning support technologies.

Finally, this work contributes to the research paradigm of Human Agent Collective Systems ([HACS](#)) system (Section [2.3.3](#)) by studying the interaction between one centralized planning agent and a Disaster Response team. However, the [HACS](#) researchers envision a scenario in which large number of computational entities (including both software agents and embodied agents such as rescue robots) and human teams collaborate at large scale. For future work, there is a potential to extend AtomicOrchid to incorporate multiple agent and human teams to study multi-agents and human interaction.

Part V

APPENDIX

A

APPENDIX TO ATOMICORCHID IMPLEMENTATION

This appendix illustrates details of AtomicOrchid implementation, which cover system architecture, data model and protocol between agent and AtomicOrchid server.

A.1 SYSTEM ARCHITECTURE

The following chart illustrates the system architecture. Numbers in the chart are explained in more details in table [12](#)

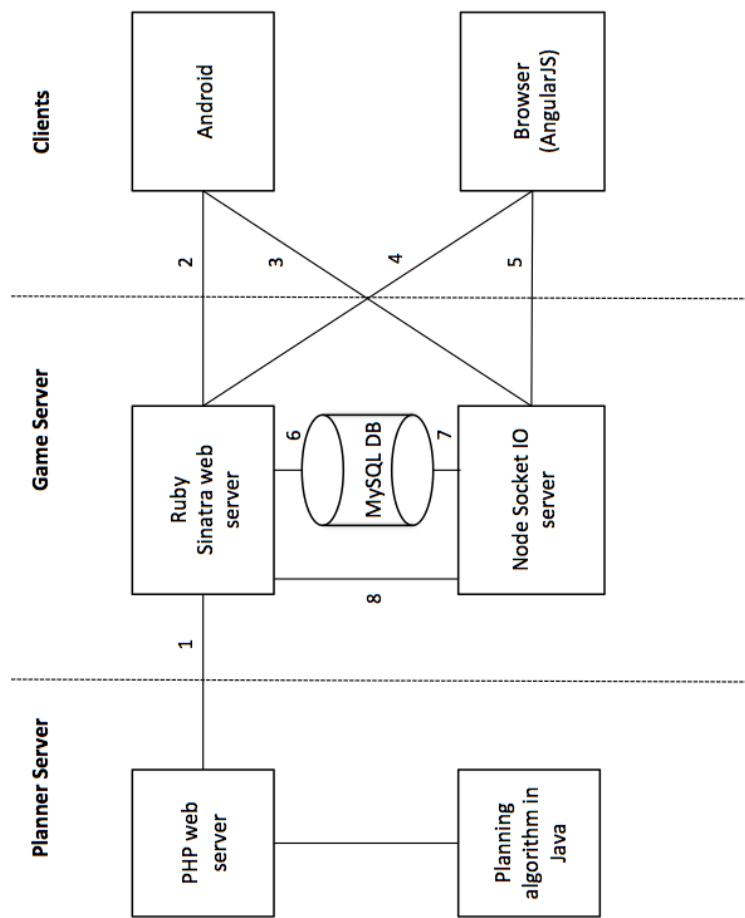


Figure 64: Technological illustration of AtomicOrchid

| Interactions | Components | Descriptions |
|--------------|---|--|
| 1 | Agent PHP server <-> Ruby game server | 1. game server requests to initialise planner for a particular game session. 2. game server requests to a plan to be computed with a game status attached. |
| 2 | Android Client <-> Ruby server | 1. Android client requests to join a game session 2. Android requests game setup info |
| 3 | Android Client <-> Node server | 1. Android reports GPS updates, player feedbacks chat messages. 2. Node server transmit updates of player health , target/player locations, messages, instructions. |
| 4 | Ruby game server <-> Browser | 1. Browser requests game setup info 3. Node server transmit updates of player health , target/player locations, messages, instructions. |
| 5 | Node server <-> Browser | 1. Browser reports plan edits and chat messages 2. Node server transmit updates of player health , target/player locations, messages, instructions. |
| 6 | DB <-> Ruby server | Ruby server retrieves and updates game status |
| 7 | DB <-> Node server | Node server retrieves and updates game status |
| 8 | Ruby server <-> Agent PHP server | Ruby server requests Node server to push latest status updates to clients including player health, cloud status target/player locations, messages, instructions |

Table 12: Communications between components

A.2 DATA MODEL

| Models | Properties | Associations | Descriptions |
|--------------|---|--|---|
| Game | id: [number] update_interval: [float] grid_size: [number] simulation_file: [string] terrain: [string] | has_many: targets has_many: players has_many: dropOffPoints has_many: plans | Represent config information of a particular game session |
| Player | id: [number] latitude: [float] longitude: [float] role: [number] initial: [string] name: [string] group: [string] | belongs_to: game | Represent players in the game |
| Task | id: [number] latitude: [float] longitude: [float] type: [number] status: [number] | belongs_to: game | Represent virtual targets in the game |
| DropOffPoint | id: [number] latitude: [float] longitude: [float] radius: [number] | belongs_to: game | Represent drop off zone |
| Plan | id: [number] | belongs_to: game | Represent a plan from agent |
| Frame | id: [number] | belongs_to: plan | Represent one particular stage in one frame |
| Instruction | id: [number] group: [string] player: [number] target: [number] status: [number] | belongs_to: frame | Represent one instruction sent to one player |

Table 13: AtomicOrchid DB model

A.3 AGENT-SERVER PROTOCOL

| Agent server interface | Path | Parameters | Return |
|--|---|---|--|
| Agent initialisation | /init_session.php? data=[JSON string] | { terrain: [string], session_id: [number], targets: [array of targets], dropoffzones: [array of dropoff zones] } | { status:[ok error] } |
| Check status | /check_status? session_id=[number] | a number representing game id | { status: [inprogress error done] } |
| Update target location (study 2 only) | /update_target? data=[JSON string] | { session_id: number, targets:[array of targets] } | { status: [ok error] } |
| Fetch plan | /fetch_plan? data=[JSON string] | { session_id: [number], frame: [time frame], rejections: [instructions rejected], (study 2 only) keep: [instructions to keep], (study 3 only) state: { players: [array of players], targets: [array of targets] } } | { plan: [[array of instructions], [array of instructions] ...] } |

Table 14: Agent server protocol

B

APPENDIX TO STUDY 1

This appendix includes additional materials to study 1, which cover target distribution, participant's demographics and chart of game event visualisation.

B.1 GAME AREA AND SET-UP

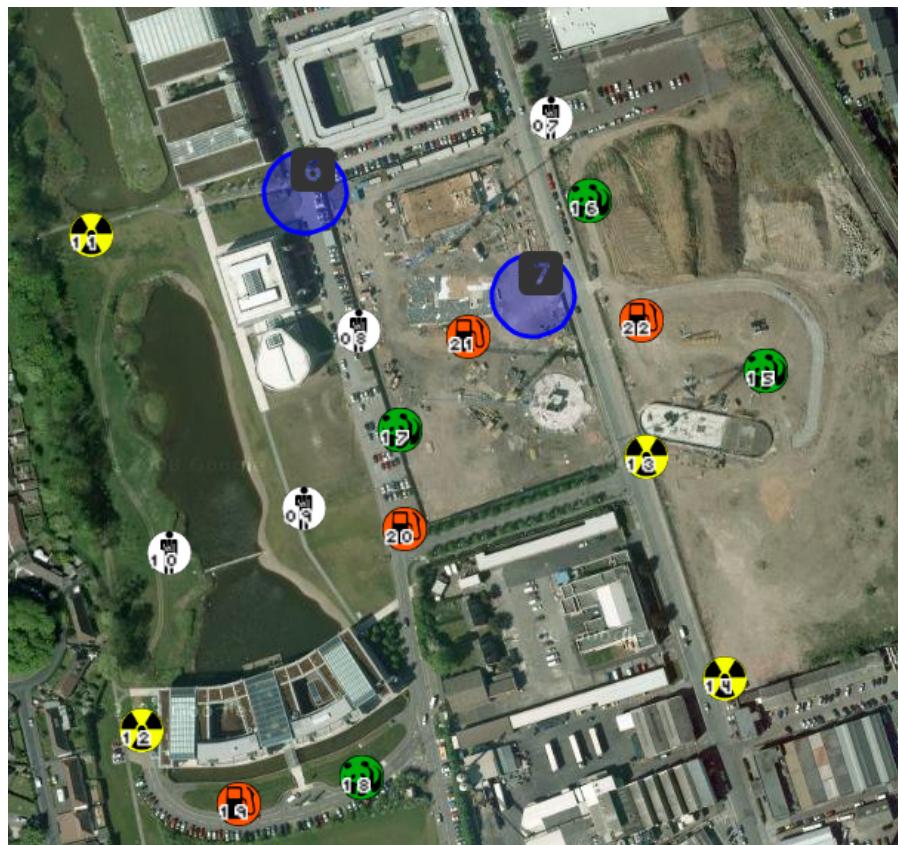


Figure 65: Target distribution in study 1

B.2 PARTICIPANT DEMOGRAPHIC

| Initials | Gender | Occupation | Mobile OS experience | Location game experence | Navigation app experience | Map reading skill | Know other participants |
|----------|--------|------------|----------------------|-------------------------|---------------------------|-------------------|-------------------------|
| NH | M | Student | Android | YES | YES | MEDIUM | |
| RL | M | Student | iOS | NO | YES | HIGH | |
| JP | F | Student | Android, Wndows | NO | YES | HIGH | TD |
| TD | M | Student | iOS | NO | YES | MEDIUM | JP |
| TV | M | Student | Android | NO | YES | HIGH | |
| HQ1 | M | Student | iOS, Blackberry | NO | YES | MEDIUM | |
| HQ2 | M | Researcher | iOS | NO | YES | HIGH | |

Table 15: Participants in session A, Study 1

| Initials | Gender | Occupation | Mobile OS experience | Location game experence | Navigation app experience | Map reading skill | Know other participants |
|----------|--------|------------|----------------------|-------------------------|---------------------------|-------------------|-------------------------|
| D2 | M | Student | Android,Windows | NO | YES | MEDIUM | |
| D1 | M | Student | Android | NO | YES | HIGH | |
| BR | F | Student | iOS | NO | YES | | KL,HQ1 |
| KL | F | Student | iOS | NO | NO | HIGH | BR,HQ1 |
| KY | F | Student | Android,Windows | NO | NO | MEDIUM | |
| JH | M | Student | Android,iOS | NO | YES | HIGH | |
| HD | F | Student | No experience | NO | NO | MEDIUM | |
| MF | M | Researcher | iOS | NO | NO | HIGH | |
| HQ1 | F | Student | iOS | NO | YES | | |
| HQ2 | F | Student | iOS | NO | YES | HD | |
| HQ3 | F | Student | iOS | NO | YES | | KL,BR |

Table 16: Participants in Session B, Study 1

B.3 GAME EVENT VISUALISATION

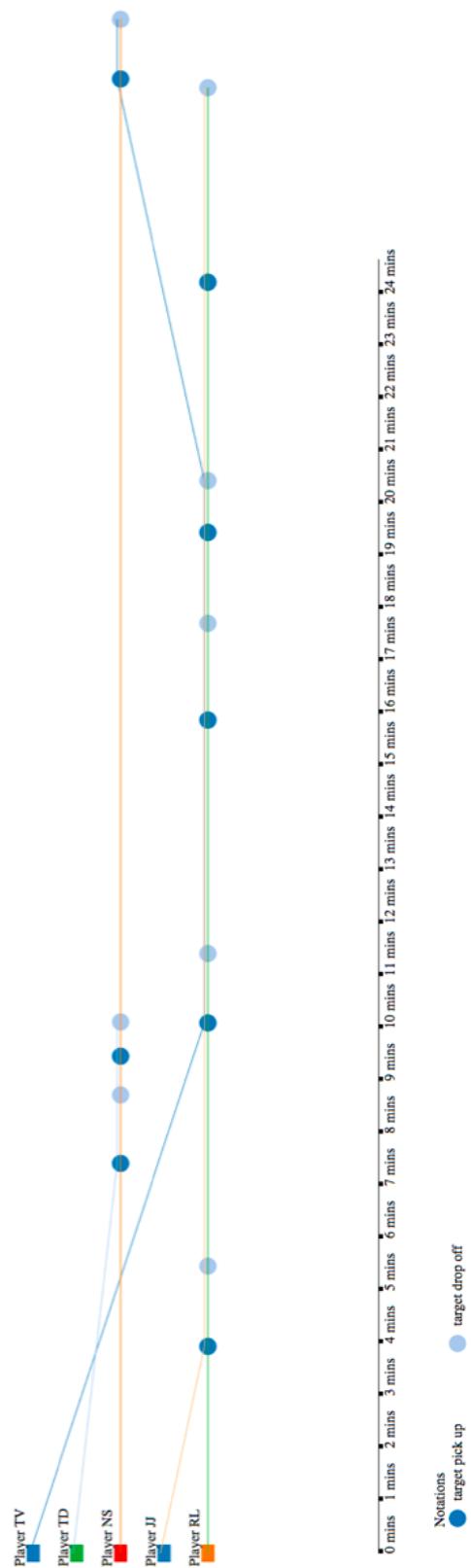


Figure 66: Event visualisation of Study 1, Session A

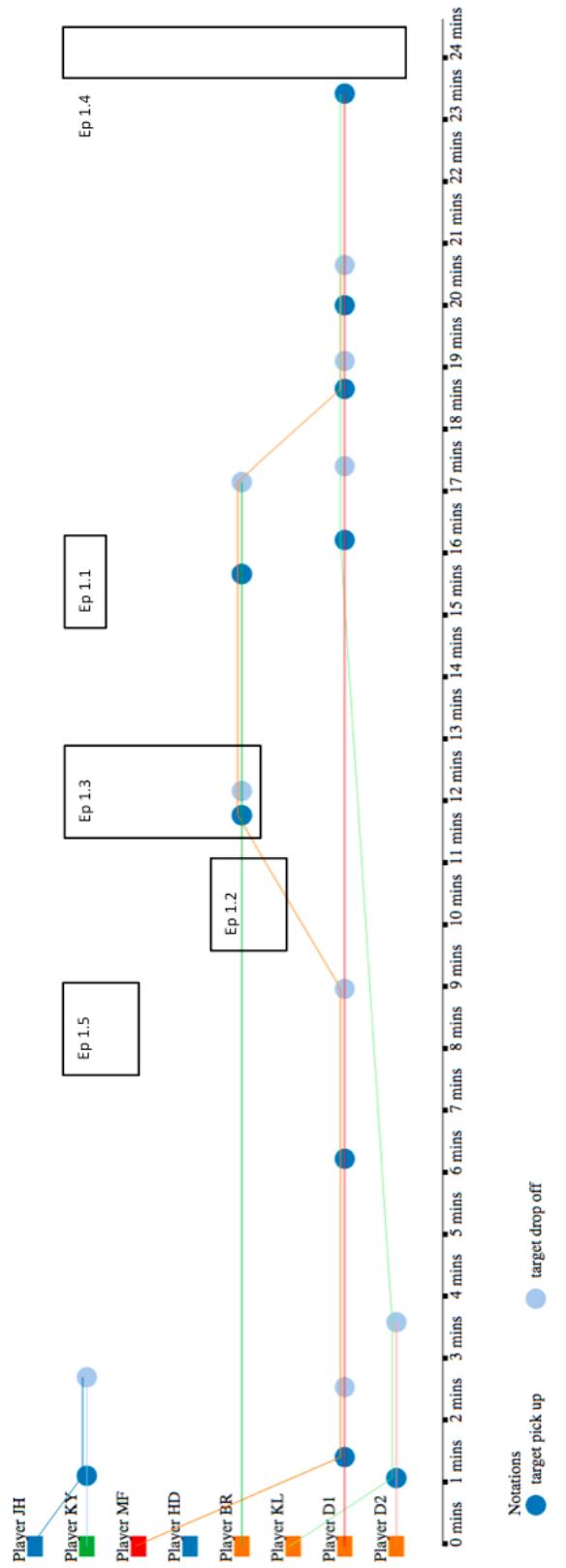


Figure 67: Event visualisation of Study 1, Session B

APPENDIX TO STUDY 2

This appendix includes additional materials to study 2, which cover target distribution, participant's demographics and chart of game event visualisation.

C.1 GAME AREA AND SET-UP

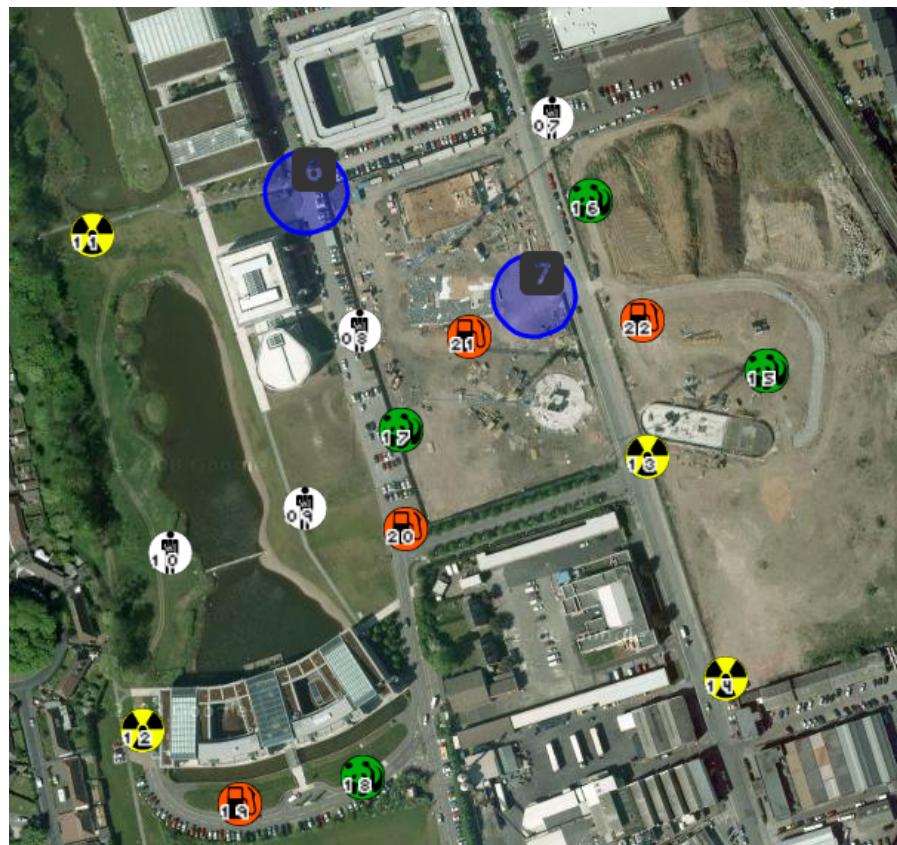


Figure 68: Target distribution in study 2

C.2 PARTICIPANT DEMOGRAPHIC

| Initials | Age | Gender | Occupation | Mobile OS experience | Location game experence | Navigation app experience | Map reading skill | Know other participants |
|----------|-----|--------|-----------------|----------------------|-------------------------|---------------------------|-------------------|-------------------------|
| AC | 26 | M | Student | iOS | YES | YES | MEDIUM | |
| HB | 19 | F | Student | iOS | NO | YES | MEDIUM | KD |
| AR | 18 | M | Student | iOS,Android | YES | YES | HIGH | |
| AW | 18 | M | Student | Android | NO | NO | MEDIUM | |
| LC | 27 | M | Event organiser | Nokia | NO | NO | HIGH | |
| YF | 24 | F | Student | Android | YES | YES | HIGH | |
| DD | 32 | F | Student | iOS | NO | YES | HIGH | |
| KD | 20 | M | Student | Android | YES | YES | HIGH | |
| HQ | 24 | M | Student | Android | NO | NO | MEDIUM | |

Table 17: Participants in Session A, Study 2

| Initials | Age | Gender | Occupation | Mobile OS experience | Location game experence | Navigation app experience | Map reading skill | Know other participants |
|----------|-----|--------|------------|----------------------|-------------------------|---------------------------|-------------------|-------------------------|
| DP | 26 | M | Student | iOS | NO | YES | MEDIUM | |
| LT | 19 | F | Student | iOS | NO | YES | HIGH | NK,NW |
| CR | 18 | M | Student | iOS,Android | NO | YES | HIGH | |
| SS | 18 | M | Student | Android | NO | YES | HIGH | PC |
| NK | 27 | F | Student | iOS | NO | YES | MEDIUM | LT,NW |
| NW | 24 | F | Student | iOS | NO | YES | MEDIUM | NK,LT |
| JL | 32 | F | Student | no experience | NO | YES | HIGH | |
| PC | 20 | F | Student | iOS | NO | YES | MEDIUM | SS |
| HQ | 24 | M | Student | iOS | NO | YES | MEDIUM | |

Table 18: Participants in Session B, Study 2

C.3 GAME EVENT VISUALISATION

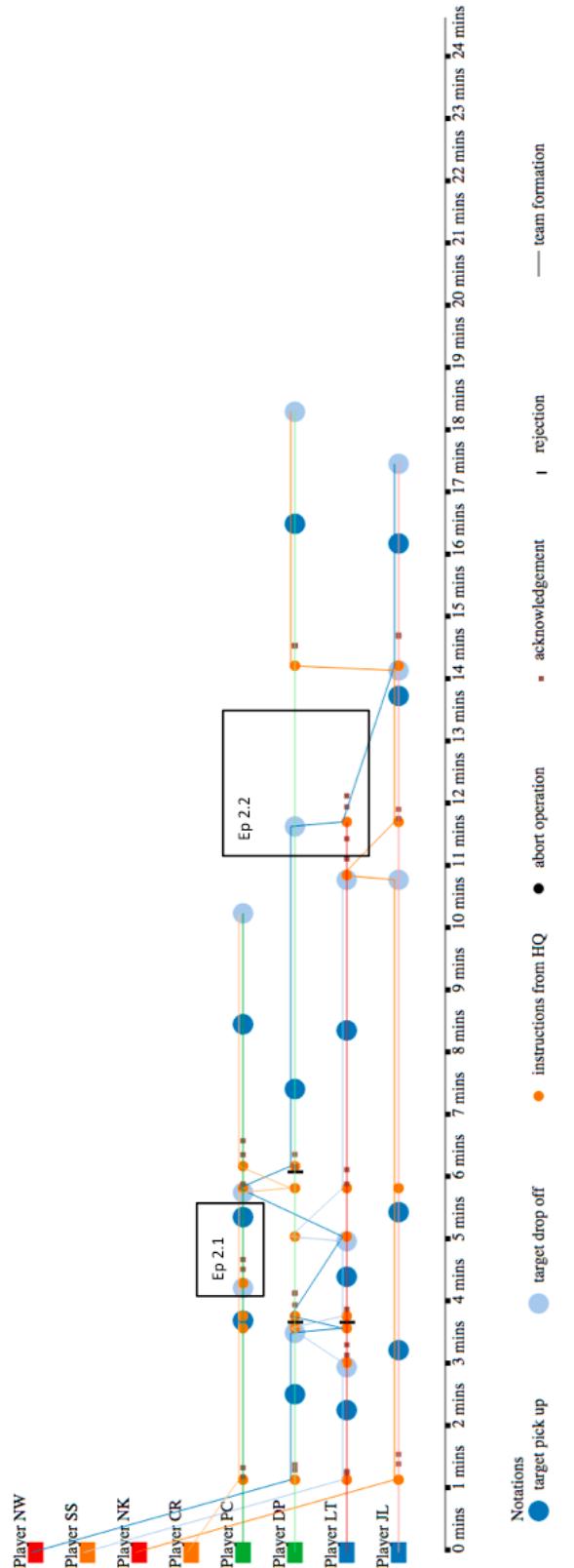


Figure 69: Event visualisation of Study 2, Session A

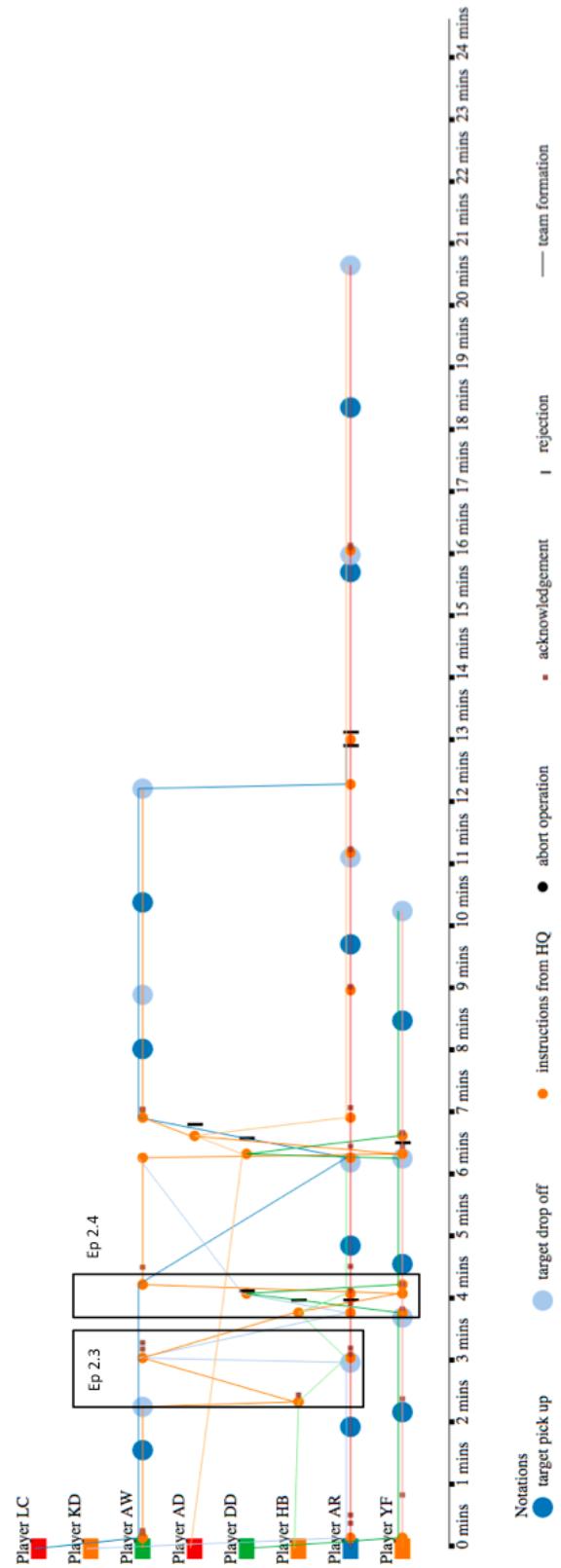


Figure 70: Event visualisation of Study 2, Session B

APPENDIX TO STUDY 3

This appendix includes additional materials to study 3, which cover target distribution, participant's demographics and chart of game event visualisation.

D.1 GAME AREA AND SET-UP

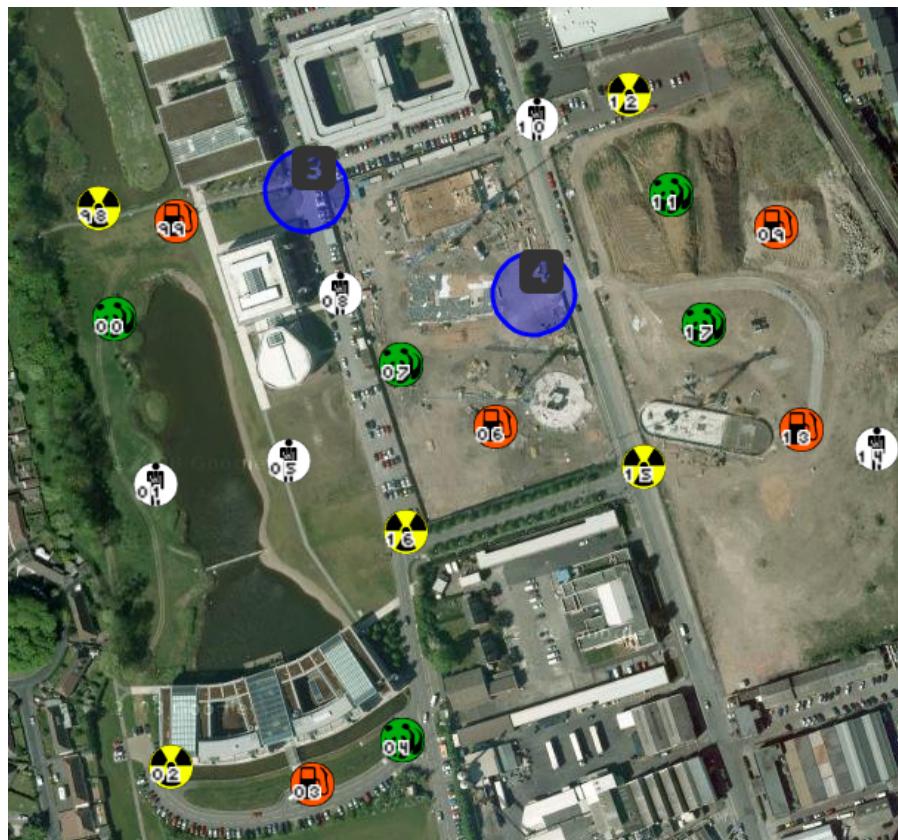


Figure 71: Target distribution in study 3

D.2 PARTICIPANT DEMOGRAPHIC

| Initials | Age | Gender | Occupation | Mobile OS experience | Location game experience | Navigation app experience | Map reading skill | Know other participants |
|----------|-----|------------|------------|----------------------|--------------------------|---------------------------|-------------------|-------------------------|
| WX | 25 | student | F | Android | NO | YES | MEDIUM | BW |
| LR | 23 | student | M | Android,iOS | YES | YES | MEDIUM | |
| NG | 26 | student | M | Android,iOS | YES | NO | HIGH | |
| MV | 22 | student | M | iOS | NO | YES | HIGH | |
| DI | 26 | researcher | M | Android | YES | YES | HIGH | |
| BW | 23 | student | M | Windows | NO | YES | MEDIUM | XW |
| LT | 23 | student | F | iOS | NO | YES | MEDIUM | DI |
| YI | 23 | student | M | iOS | NO | YES | MEDIUM | |
| HQ1 | 25 | student | M | iOS | YES | YES | MEDIUM | HQ2 |
| HQ2 | 24 | student | M | iOS | YES | YES | MEDIUM | HQ1 |

Table 19: Participants in Session A, Study 3

| Initials | Age | Gender | Occupation | Mobile OS experience | Location game experience | Navigation app experience | Map reading skill | Know other participants |
|----------|-----|------------|------------|-----------------------|--------------------------|---------------------------|-------------------|-------------------------|
| GS | 29 | student | M | Android | NO | YES | MEDIUM | BW |
| WB | 25 | student | F | Android | NO | YES | HIGH | |
| CE | 24 | student | F | Android, Windows, iOS | NO | YES | MEDIUM | |
| DS | 24 | student | M | Android | YES | NO | MEDIUM | |
| MP | 24 | researcher | M | Android, iOS | YES | YES | MEDIUM | |
| MB | 25 | student | M | Android | NO | NO | HIGH | CY, WB, HQ1, HQ2 |
| DS | 25 | student | M | Android | NO | YES | MEDIUM | DS |
| CY | 25 | student | M | iOS | YES | YES | MEDIUM | MB |
| HQ1 | 34 | student | M | Android | YES | YES | MEDIUM | HQ2, MP |
| HQ2 | 32 | student | M | Android | YES | YES | MEDIUM | HQ1, MP |

Table 20: Participants in Session B, Study 3

D.3 GAME EVENT VISUALISATION

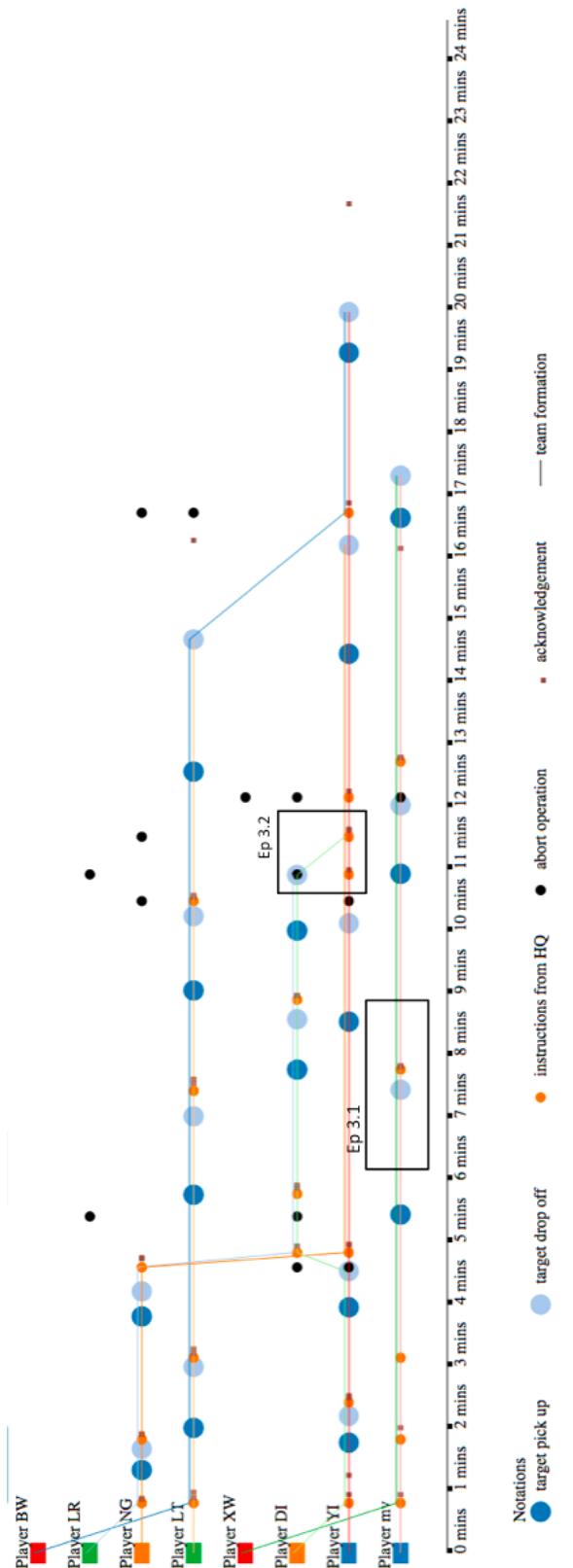


Figure 72: Event visualisation of Study 3, Session A

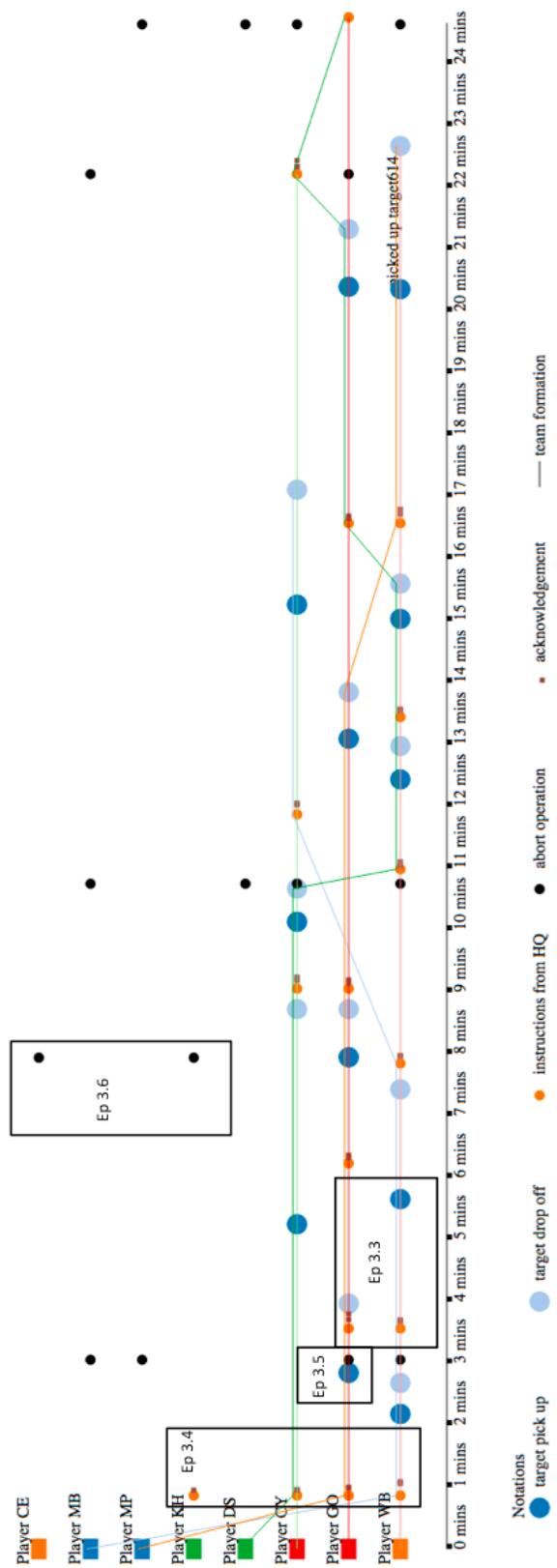


Figure 73: Event visualisation of Study 3, Session B

BIBLIOGRAPHY

- Abbott, K. R. and Sarin, S. K. (1994). Experiences with workflow management: issues for the next generation. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 113–120.
- Ackerman, M. (2000). The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human-Computer Interaction*, 15(2):179–203.
- Allen, J. E., Guinn, C. I., and Horvitz, E. (1999). Mixed-initiative interaction. *IEEE Intelligent Systems*, 14(5):14–23.
- Allen, J. F. (1984). Towards a general theory of action and time. *Artificial Intelligence*, 23(2):123–154.
- Armenakis, C. (2012). Systems approach to management of disasters: methods and applications. *Natural Hazards*, 66(1):139–140.
- Avrahami, D. and Hudson, S. E. (2004). QnA: Augmenting an instant messaging client to balance user responsiveness and performance. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, volume 6, pages 515–518.
- Bellotti, V. and Edwards, K. (2001). Intelligibility and Accountability: Human Considerations in Context-Aware Systems. *Human-Computer Interaction*, 16(2):193–212.
- Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M., and Row-Farr, J. (2006). Can you see me now? *ACM Transactions on Computer-Human Interaction*, 13(1):100–133.
- Benford, S., Magerkurth, C., and Ljungstrand, P. (2005). Bridging the PHYSICAL AND DIGITAL in Pervasive Gaming. *Communications of the ACM*, 48(3).

- Boersma, K., Groenewegen, P., and Wagenaar, P. (2009). Emergency Response Rooms in Action: an ethnographic case-study in Amsterdam. In *Proceedings of the 6th International ISCRAM Conference*, number May.
- Bowers, J., Button, G., and Sharrock, W. (1994). Workflow From Within and Without: Technology and Cooperative Work on the Print Industry Shopfloor Introduction. In *4th European Conference on Computer Supported Cooperative Work*, pages 51–66. Fourth European Conference on Computer-Supported Cooperative Work.
- Bradshaw, J., Sierhuis, M., and Acquisti, A. (2003). Adjustable autonomy and human-agent teamwork in practice: An interim report on space applications. *Agent autonomy*, (October 2002):7–9.
- Bradshaw, J. M., Feltovich, P., and Johnson, M. (2011). Human-Agent Interaction. In Boy, G., editor, *Handbook of HumanMachine Interaction*, chapter 13, pages 293–302. Ashgate.
- Brundell, P., Tennent, P., Greenhalgh, C., Knight, D., Crabtree, A., O’Malley, C., Ainsworth, S., Clarke, D., Carter, R., and Adolphs, S. (2008). Digital Replay System (DRS): A Tool for Interaction Analysis. In *International Conference on Learning Sciences Workshop on Interaction Analysis*.
- Burstein, M., Ferguson, G., and Allen, J. (2003). Integrating agent-based mixed-initiative control with an existing multi-agent planning system. *Proceedings Fourth International Conference on Multi-Agent Systems*, (January):389–390.
- Carstensen, P. and Schmidt, K. (1999). Computer supported cooperative work: New challenges to systems design. In K. Itoh (Ed.), *Handbook of Human Factors*.
- Carver, L. and Turoff, M. (2007). Human-computer interaction: the human and computer as a team in emergency management information systems. *Communications of the ACM*, 50(3):33–38.
- Chang, M.-S., Tseng, Y.-L., and Chen, J.-W. (2007). A scenario planning approach for the flood emergency logistics preparation prob-

- lem under uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 43(6):737–754.
- Chen, R., Sharman, R., Rao, H. R., and Upadhyaya, S. J. (2005). Design Principles of Coordinated Multi-incident Emergency Response Systems. *Simulation*, 3495:177–202.
- Chen, R., Sharman, R., Rao, H. R., and Upadhyaya, S. J. (2008). Coordination in emergency response management. *Communications of the ACM*, 51(5):66–73.
- Comfort, L. K. (2004). Coordination in Rapidly Evolving Disaster Response Systems: The Role of Information. *American Behavioral Scientist*, 48(3):295–313.
- Conti, G., Abdullah, K., Grizzard, J., Stasko, J., Copeland, J., Ahamad, M., Owen, H., and Lee, C. (2006). Countering security information overload through alert and packet visualization. *IEEE Computer Graphics and Applications*, 26(2):60–70.
- Costanza, E., Fischer, J. E., Colley, J. A., Rodden, T., Ramchurn, S. D., and Jennings, N. R. (2014). Doing the laundry with agents. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 813–822.
- Cova, T. (1999). GIS in emergency management. *Geographical information systems*, (Rejeski 1993).
- Crabtree, A., Benford, S., Greenhalgh, C., Tennent, P., Chalmers, M., and Brown, B. (2006). Supporting ethnographic studies of ubiquitous computing in the wild. *Proceedings of the 6th ACM conference on Designing Interactive systems - DIS '06*, page 60.
- Crabtree, A., Rouncefield, M., and Tolmie, P. (2012). *Doing Design Ethnography*. Springer.
- Crabtree, A., Tandavanitj, N., Steed, A., Benford, S., Rodden, T., Greenhalgh, C., Flintham, M., Anastasi, R., Drozd, A., Adams, M., and Row-Farr, J. (2004). Orchestrating a mixed reality game 'on the ground'. In *Proceedings of the 2004 conference on Human factors in*

- computing systems - CHI '04*, pages 391–398, New York, New York, USA. ACM Press.
- Dashti, S., Palen, L., Heris, M. P., Kenneth, M., Anderson, S., and Anderson, T. J. (2014). Supporting Disaster Reconnaissance with Social Media Data: A Design-Oriented Case Study of the 2013 Colorado Floods. *Information Systems for Crisis Response and Management (2014)*, (May).
- Dawes, S. S., Schneider, C. A., and Road, W. (2004). Information, Technology and Coordination: World Trade Center Response. *World Trade*, (June).
- Deconinck, S. (2013). Computational Disaster Management. In *23rd international joint conference on Artificial Intelligence. AAAI Press, 2013.*, pages 12–18.
- Dorais, G., Bonasso, R., and Kortenkamp, D. (1999). Adjustable autonomy for human-centered autonomous systems. *Working notes of the 16th International Joint Conference on Artificial Intelligence Workshop on Adjustable Autonomy Systems*, pages 16–35.
- Drury, J., Cocking, C., and Reicher, S. (2009). Everyone for themselves? A comparative study of crowd solidarity among emergency survivors. *The British journal of social psychology / the British Psychological Society*, 48(Pt 3):487–506.
- Eglese, R. W. and Pidd, M. (1994). A spatial decision support system for emergency planning. In *International Emergency Management and Engineering Conference research and applications*. Society for Computer Simulation.
- Endsley, M. (2001). Design of situation awareness in complex systems. In *Proceedings of the Second International Workshop on symbiosis of humans, artifacts and environment. 2001*.
- Faraj, S. and Xiao, Y. (2006). Coordination in Fast-Response Organizations.
- Ferguson, G., Allen, J., and Miller, B. (1996). Towards a Mixed Initiative Planning Assistant. In Drabble, B., editor, *Proceedings of the*

3rd International Conference on Artificial Intelligence Planning Systems AIPS96, pages 70–77. AAAI Press.

Fischer, J., Reeves, S., Rodden, T., Reece, S., and Ramchurn, S. D. (2015). Building a Birds Eye View: Collaborative Work in Disaster Response. In *The 33rd Annual ACM Conference on Human Factors in Computing Systems*.

Fischer, J. E., Rodden, T., Flintham, M., Price, D., Goulding, J., and Pantidi, N. (2012). Serious Mixed Reality Games. *CSCW2012: Mixed Reality Games Workshop*.

Fitts, P. M. (1951). Human Engineering for an Effective Air Navigation and Traffic-Control System. Technical report, National Research Council.

Flintham, M., Benford, S., Anastasi, R., Hemmings, T., Crabtree, A., Greenhalgh, C., Tandavanitj, N., Adams, M., and Row-Farr, J. (2003). Where On-Line Meets On-The-Streets: Experiences With Mobile Mixed Reality Games. In *Proceedings of the conference on Human factors in computing systems - CHI '03*, pages 569–576. ACM Press.

Geertman, S. and Stillwell, J. (2004). Planning support systems: an inventory of current practice. *Computers, Environment and Urban Systems*, 28(4):291–310.

Georgeff, M., Pell, B., Pollack, M., Tambe, M., and Wooldridge, M. (1998). The belief-desire-intention model of agency. In *Intelligent Agents V: Agents Theories, Architectures, and Languages. 5th International Workshop, ATAL'98.*, pages 1–10.

Goodrich, M. A., Jr, D. R. O., Crandall, J. W., and Palmer, T. J. (2001). Experiments in Adjustable Autonomy Robot Effectiveness Neglect. In *IJCAI workshop on Autonomy, Delegation and Control: Interacting with Intelligent Agents*, pages 1624–1629.

Grootjen, M., Neerincx, M. A., and Weert, J. C. M. V. (2007). Measuring Cognitive Task Load on a Naval Ship : Implications of a Real World Environment. *Foundations*, pages 147–156.

- Hardin, B. and Goodrich, M. A. (2009). On using mixed-initiative control: A perspective for managing large-scale robotic teams. In *Proceedings of the 4th ACM/IEEE international conference on human robot interaction (HRI '09)*, pages 165–172.
- Hayes-Roth, B. and Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science*, 3(157):275–310.
- Heath, C., Hindmarsh, J., and Luff, P. (2010). *Video in Qualitative Research*. Sage Publications Ltd.
- Heath, C. and Luff, P. (1992). Collaboration and controlCrisis management and multimedia technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work (CSCW)*, 1(1):24–48.
- Hiltz, S. R. and Turoff, M. (1985). Structuring Computer-Mediated Communication Systems to Avoid Information Overload. 28.
- HM Government (2008). Fire and Rescue Manual - Volume 2 Fire Service Operations - Incident Command.
- Hoffman, R. R., Feltovich, P. J., Ford, K. M., Woods, D. D., Klein, G., Feltovich, A., and Hayes, P. J. (2002). A Rose by Any Other Name ... Would Probably Be Given an Acronym. *IEEE Intelligent Systems*, 17(4):72–80.
- Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit CHI 99*, pages(May):159–166.
- Intergraph Corporation (2000). Intergraph MGE GIS. Technical report.
- Jennings, N., Moreau, L., Nicholson, D., Ramchurn, S. D., Roberts, S., Rodden, T., and Rogers, A. (2014). Human-agent collectives. *Communications of the ACM*, 52(12):80–88.
- Jordan, B. and Henderson, A. (1995). Interaction Analysis: Foundations and Practice.

- Kaber, D. and Endsley, M. (1995). Out of the loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, 16(3).
- Kadous, M. W., Sheh, R. K.-M., and Sammut, C. (2006). Effective user interface design for rescue robotics. *Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI '06*, page 250.
- Kitano, H. (2000). RoboCup Rescue: a grand challenge for multi-agent systems. *Proceedings 4th International Conference on MultiAgent Systems*.
- Koes, M., Nourbakhsh, I., Sycara, K., Ramchurn, S. D., and Others (2005). Heterogeneous multirobot coordination with spatial and temporal constraints. In *Proceedings of the National Conference on Artificial Intelligence*, volume 20, page 1292.
- Kogan, M., Palen, L., and Anderson, K. (2015). Think Local, Retweet Global: Retweeting by the Geographically-Vulnerable during Hurricane Sandy. *Computer-Supported Cooperative Work and Social Computing* (2015).
- Koleva, B., Adams, M., Taylor, I., Benford, S., Fraser, M., Greenhalgh, C., Schnädelbach, H., vom Lehn, D., Heath, C., and Row-Farr, J. (2001). Orchestrating a mixed reality performance. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '01*, (3):38–45.
- Kopena, J. B., Sultanik, E. a., Lass, R. N., Nguyen, D. N., Dugan, C. J., Modi, P. J., and Regli, W. C. (2008). Distributed Coordination of First Responders. *IEEE Internet Computing*, 12(1):45–47.
- Kozierok, R. and Maes, P. (1993). A learning interface agent for scheduling meetings. *Proceedings of the 1st international conference on Intelligent user interfaces IUI 93*, 11:81–88.
- Lagoudakis, M. and Markakis, E. (2005). Auction-Based Multi-Robot Routing. *Robotics: Science and Systems*, 5.

- Lieberman, H. (1997). Autonomous interface agents. *Proceedings of the ACM SIGCHI Conference on . . .*, pages 67–74.
- Lieberman, H. and Selker, T. (2003). Agents for the user interface. *Handbook of Agent Technology*, pages 1–21.
- Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM*, 37(7):30–40.
- Manzey, D. and Onnasch, L. (2012). Human Performance Consequences of Automated Decision Aids : The Impact of Degree of Automation and System Experience. *Human Factors*, 6(1):57–87.
- Mead, G. and Mind, H. (1934). Self and society. *Chicago: University of Chicago*, pages 1–8.
- Meesters, K. and Van De Walle, B. (2013). Sisaster in my backyard: A serious game introduction to disaster information management. In *ISCRAM 2013 Conference Proceedings - 10th International Conference on Information Systems for Crisis Response and Management*, number May, pages 145–150.
- Mercer, J. and Homola, J. (2014). Human-Automation Cooperation for Separation Assurance in Future NextGen Environments. *Proceedings of the HCI-Aero Conference*.
- Metral, Y. and Maes, P. (1998). Collaborative interface agents. *Readings in agents*.
- Miller, H. (1999). GIS-based dynamic traffic congestion modeling to support time-critical logistics. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, page 9. IEEE Comput. Soc.
- Morrow, N. and Mock, N. (2011). Independent evaluation of the Ushahidi Haiti project. *Development Information Systems International 8 (2011)*.
- Niazhkhani, Z., Pirnejad, H., Berg, M., and Aarts, J. (2009). The impact of computerized provider order entry systems on inpatient clinical workflow: a literature review. *Journal of the American Medical Informatics Association : JAMIA*, 16(4):539–49.

- Norman, D. A. (2013). *The design of everyday things: Revised and expanded*. Basic books.
- Nwana, H. (1996). Software agents: An overview. *The knowledge engineering review*, pages 1–49.
- Okaya, M., Niwa, T., and Takahashi, T. (2014). TENDENKO: agent-based evacuation drill and emergency planning system. *...agents and multi-agent systems*, pages 1669–1670.
- Palen, L. and Liu, S. B. (2007). Citizen communications in crisis: anticipating a future of ICT-supported public participation. *Natural Hazards*, pages 727–736.
- Palen, L., Soden, R., Anderson, T. J., and Barrenechea, M. (2015). Success & Scale in a Data - Producing Organization: The Socio - Technical Evolution of OpenStreetMap in Response to Humanitarian Events. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing*, pages 4113–4122.
- Parasuraman, R. and Manzey, D. H. (2010). Complacency and bias in human use of automation: an attentional integration. *Human factors*, 52(3):381–410.
- Parasuraman, R. and Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2):230–253.
- Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans*, 30(3):286–297.
- Petrescu-prahova, M. and Butts, C. T. (2005). Emergent Coordination in the World Trade Center. *Sociology The Journal Of The British Sociological Association*, pages 1–23.
- Pettit, C., Shyy, T., and Stimson, R. (2003). An on-line planning support system to evaluate urban and regional planning scenarios. In *Planning support systems in practice*, pages 331–347.

- Pillac, V., Cebrian, M., and Van Hentenryck, P. (2015). A column-generation approach for joint mobilization and evacuation planning. *Constraints*, 20(3):285–303.
- Provitolo, D. and Dubos-paillard, E. (2011). Emergent Human Behavior during a Disaster: Thematic versus Complex Systems Approaches. In *Emergent Properties in Natural and Artificial Complex Systems*, pages 47–57.
- Ramchurn, S., Simpson, E., Fischer, J., Huynh, T., Ikuno, Y., Reece, S., Jiang, W., Wu, F., Flann, J., Roberts, S., Moreau, L., Rodden, T., and Jennings, N. (2015a). HAC-ER: A disaster response system based on human-agent collectives. In *14th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 533–541.
- Ramchurn, S. D., Polukarov, M., Farinelli, A., Jennings, N., and Trong, C. (2010). Coalition Formation with Spatial and Temporal Constraints. *Constraints*, pages 1181–1188.
- Ramchurn, S. D., Wu, F., Jiang, W., Fischer, J. E., Reece, S., Roberts, S., Rodden, T., Greenhalgh, C., and Jennings, N. R. (2015b). Human agent collaboration for disaster response. *Journal of Autonomous Agents and Multi-Agent Systems*, pages 1–29.
- Rescue Global (2012). An Introduction to Rescue Global.
- Richard, H., Harper, J. A., and Hughes, D. Z. S. (1989). Working in harmony: an examination of computer technology in air traffic control. In *First European Conference on Computer Supported Cooperative Work*, pages 73–86.
- Ropohl, G. (1999). Philosophy of socio-technical systems. *Techné: Research in Philosophy and Technology*, pages 59–71.
- Sarcevic, A., Palen, L., White, J., Starbird, K., Bagdouri, M., and Anderson, K. (2012). Beacons of Hope in Decentralized Coordination: Learning from On-the-Ground Medical Twitterers During the 2010 Haiti Earthquake. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 47–56.

- Scerri, P., Farinelli, A., Okamoto, S., and Tambe, M. (2005a). Allocating tasks in extreme teams. *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems - AAMAS '05*, page 727.
- Scerri, P., Farinelli, A., Okamoto, S., and Tambe, M. (2005b). Allocating tasks in extreme teams. *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems - AAMAS '05*, page 727.
- Schooley, B., Hilton, B., Lee, Y., Mcclintock, R., Olusola, S.-o., Horan, T., Schooley, B., Mcclintock, R., and Horan, T. (2010). CrashHelp : A GIS Tool for Managing Emergency Medical Responses to Motor Vehicle Crashes. *Information Systems*, (May):1–11.
- Schwarz, M., Stückler, J., and Behnke, S. (2014). Mobile teleoperation interfaces with adjustable autonomy for personal service robots. *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14*, pages 288–289.
- Searle, J. R. (1976). A classification of illocutionary acts.
- Sharples, S., Millen, L., Golightly, D., and Balfe, N. (2011). The impact of automation in rail signalling operations. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 225(2):179–191.
- Simpson, E., Roberts, S., Psorakis, I., and Smith, A. (2013). Dynamic bayesian combination of multiple imperfect classifiers. *Studies in Computational Intelligence*, 474:1–35.
- Skeels, M., Lee, B., Smith, G., and Robertson, G. (2008). Revealing uncertainty for information visualization. *Proceedings of the working conference on Advanced visual interfaces - AVI '08*, page 376.
- Sobhaninejad, G., Hori, M., and Kabeyasawa, T. (2011). Enhancing integrated earthquake simulation with high performance computing. *Advances in Engineering Software*, 42(5):286–292.
- Squire, K. and Jenkins, H. (2003). Harnessing the power of games in education. *Insight*, 3:5–33.

- Starbird, K. and Palen, L. (2010). Pass it on?: Retweeting in mass emergency. In *Information Systems for Crisis Response and Management (2010)*, number December 2004, pages 1–10.
- Suchman, L. (1987). *Plans and situated actions: the problem of human-machine communication*. Cambridge university press.
- Sukthankar, G., Giampapa, J. A., and Burnett, C. (2009). Communications for Agent-Based Human Team Support. In *Organizations in Multi-Agent Systems*, number 1, pages 284–312. IGI-Global (<http://www.igi-pub.com/>) Handbook of Research, Information Science Reference.
- Sukthankar, G., Shumaker, R., and Lewis, M. (2011). Intelligent agents as teammates. *Theories of Team Cognition: Cross-Disciplinary Perspectives*, pages 313–343.
- Susi, T., Johannesson, M., and Backlund, P. (2007). Serious Games , An Overview. *Elearning*, 73(10):28.
- Sutton, J., Palen, L., and Shklovski, I. (2008). Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires. In *Proceedings of the 5th International ISCRAM Conference*, number May, pages 1–9.
- Tang, F. and Ren, A. (2012). GIS-based 3D evacuation simulation for indoor fire. *Building and Environment*, 49:193–202.
- Tolcher, R. (2005). *Ethnomethodological Studies of Conventional Software Engineering Methods in Military C2 Work Settings*. PhD thesis, University of Oxford.
- Toups, Z. O., Kerne, A., and Hamilton, W. (2009). Game design principles for engaging cooperative play. In *Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games - Sandbox '09*, page 71, New York, New York, USA. ACM Press.
- Toups, Z. O., Kerne, A., and Hamilton, W. A. (2011). The team coordination game. *ACM Transactions on Computer-Human Interaction*, 18(4):1–37.

- Turoff, M. and Chumer, M. (2004). The design of a dynamic emergency response management information system (DERMIS). *Journal of Information . . .*, pages 1–35.
- U.S. Department of Homeland Security (2008). National Incident Management System.
- Vickery, P. J., Masters, F. J., Powell, M. D., and Wadhera, D. (2009). Hurricane hazard modeling: The past, present, and future. *Journal of Wind Engineering and Industrial Aerodynamics*, 97(7-8):392–405.
- Vlassis, N. (2007). *A Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence*, volume 1.
- Wattegama, C. (2012). ICT for disaster management. *Information Economy, Society and Polity*.
- Wickens, C., Li, H., and Santamaria, A. (2010). Stages and levels of automation: An integrated meta-analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 4:389–393.
- Wu, F., Ramchurn, S. D., Jiang, W., Fischer, J. E., Rodden, T., and Jennings, N. R. (2015). Agile Planning for Real-World Disaster Response. In *International Joint Conference on Artificial Intelligence*.
- Yang, S.-Y. and Lee, D.-L. (2012). Developing a Cloud Intelligent and Energy-Saving Information Interface Agent with Web Services. *2012 26th International Conference on Advanced Information Networking and Applications Workshops*, pages 1310–1315.
- Zerger, A. and Smith, D. I. (2003). Impediments to using GIS for real-time disaster decision support. *Computers, Environment and Urban Systems*, 27(2):123–141.
- Zimmerman, J., Tomasic, A., Simmons, I., Hargraves, I., Mohnkern, K., Cornwell, J., and McGuire, R. M. (2007). Vio: a mixed-initiative approach to learning and automating procedural update tasks. In *CHI 07 Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1445–1454. ACM.

- Zuk, T. and Carpendale, S. (2007). Visualization of uncertainty and reasoning. *Smart Graphics*, pages 164–177.