

WENCHAO JIANG

INVESTIGATING INTERACTIONAL ISSUES OF
AGENT PLANNING SUPPORT FOR DISASTER
RESPONSE

INVESTIGATING INTERACTIONAL ISSUES OF AGENT PLANNING SUPPORT FOR DISASTER RESPONSE

WENCHAO JIANG

With a Mixed Reality Game Probe

August 2012 – version 0.1

Wenchao Jiang: *Investigating Interactional Issues of Agent Planning Support for Disaster Response* , With a Mixed Reality Game Probe , © August 2012

— something.

Family means nobody gets left behind, or forgotten.

— Lilo & Stitch

Dedicated to the loving memory of Rudolf Miede.

1939 – 2005

ABSTRACT

This thesis contributes to the understanding of the potential socio-technical issues that might emerge from the interaction between responder teams and automated planning support and propose design solutions to them.

(Problem) Recently, Frequent natural and man-made disasters in Haiti, Chile and Japan drew attention of Researchers. A lot 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 contains several incident commanders and field agents, is faced with the problem of carrying out geographically distributed tasks under spatial and time constraints in a quickly changing task environment.

Effective planning and coordination can be a key factor for the success of disaster operation but it is difficult to achieve. Recent advance in the multi-agent technologies leads to the possibility of building agent software which supports Team coordination by automating the task planning process. However , it is unknown how the agent software can fit into the team organisation in a way that improve rather than hinders the team performance. The interaction between human operators and planning support systems need to be carefully designed before technology deployment.

(method) This work presents three field studies which investigates the impact of different interactional arrangements between human teams and automated planning support. The studies adopt serious game approach which is arguably an established vehicle to 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 real aspects of 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 software. Interaction analysis is carried out on the data, leading to descriptive results which unpacks interactional issues. By iteratively designing and examining different interactional arrangements through three iteration of studies, we progressively explore requirements and social implications of planning support system for responder teams.

In the 1st study, field responders and incident commander coordinate without support of the intelligent planner. The study establish baseline performance of the game play and derived several requirements for planning support system. In the 2nd study, an intelligent planner was introduced to support field responders directly. In the third study, Incident commander mediate task assignment between field responders and the planning agent.

(results) Overall these studies show that ...

ACKNOWLEDGMENTS

To my parents, wife, friends

CONTENTS

1	INTRODUCTION	2
1.1	Problem Definition and Objectives	3
1.2	Approach	5
1.3	Scoping	6
1.4	Research questions	8
1.5	Contributions	9
1.6	Publications of this thesis	9
1.7	Structure of the Thesis	10
i	BACKGROUND	12
2	LITERATURE REVIEW	13
2.1	Automation and Human Agent interaction	13
2.1.1	A History of Automation	13
2.1.2	The Level of Automation	14
2.1.3	Definition of Software Agent	15
2.1.4	Human Agent Interactions	16
2.1.5	Issues in Human Agent Interactions	16
2.1.6	A summary	17
2.2	CSCW Workflow Management	17
2.3	Task Planning in Disaster Response	17
2.3.1	Task Planning	17
2.3.2	DR Command structure	19
2.4	Technology support for task planning	20
2.4.1	Disaster simulation, optimisation	20
2.5	Game based simulation	20
2.5.1	Serious Game	20
2.5.2	Mixed Reality Games	22
2.6	Mixed Reality Games	24
2.6.1	Definition and examples	24

2.6.2 Game for Serious Purpose	24
ii METHODOLOGY AND APPROACH	25
3 APPROACH	26
3.1 Socio-technical Perspective of technology support for organizational work	27
3.2 Serious Mixed Reality Game as a testbed	29
3.3 The AtomicOrchid Platform	30
3.3.1 Game Design Rationale	31
3.3.2 Iterative design and development	34
3.3.3 System Architecture	35
3.3.4 The planning agent	36
3.4 Explore interaction designs with three AtomicOrchid studies	37
3.4.1 The On-the-loop interaction	40
3.4.2 The In-the-loop interaction	41
3.4.3 The three AO game studies to explore socio-technical issue of human agent interaction	42
3.5 Collaboration with Professional Disaster Response Organisation	42
3.5.1 Introduction of Rescue Global	43
4 METHODOLOGIES TO INVESTIGATING HUMAN AGENT INTERACTION	45
4.1 Ethnomethodology perspective	45
4.2 Interaction Analysis	46
4.3 Mixed Methods Approach	47
4.3.1 Log data analysis	48
4.3.2 Message classification	50
4.3.3 Group interview	51
4.4 Summary	51
iii STUDIES	53
5 ATOMICORCHID STUDY 1: NON AGENT VERSION	54

5.1	Introduction	54
5.2	System Description	56
5.3	Study Design	58
5.4	Data analysis and results	60
5.4.1	Speech act analysis on text messages	61
5.4.2	Responding to directives from HQ	65
5.4.3	Local coordination without HQ	70
5.4.4	Messages as a resource of situational awareness	73
5.5	Performance implications	76
5.5.1	Division of labour between HQ and Field Responders	76
5.5.2	Enhancing remote coordination	77
5.6	Design Requirements	78
5.7	Summary	81
6	ATOMICORCHID STUDY 2: AGENT SUPPORTED HUOL DESIGN	83
6.1	Introduction	83
6.2	System Evolution	85
6.2.1	The planning agent	86
6.2.2	A feedback loop	87
6.2.3	Interface improvement	88
6.3	Study Design	88
6.4	Data analysis and results	88
6.4.1	Task assignment	89
6.4.2	Team reformation	91
6.4.3	Task interruption	94
6.4.4	Disagreement on task interruption	97
6.4.5	The headquarters	99
6.5	Discussion	100
6.5.1	Division of labour between the agent and the human teams	100
6.5.2	Hidden costs of team reformation and task interruption	101

6.5.3	Feedback to the agent	103
6.6	Design implications	104
6.6.1	Achieve common ground	104
6.6.2	Facilitate accountability	104
6.6.3	Balance responsibilities between humans and agent	105
6.7	Summary	105
7	ATOMICORCHID STUDY 3: AGENT-SUPPORTED IN-THE-LOOP DESIGN	106
7.1	Introduction	106
7.2	System Evolution	108
7.2.1	Interface functionalities	109
7.2.2	The planning agent	112
7.3	Study Design	112
7.4	Data Analysis	115
7.4.1	Messaging system	116
7.4.2	Overview of task assignments	117
7.4.3	Approve the plans generated by planner	119
7.4.4	Editing assignments generated by planner	120
7.4.5	Cancelling approved assignments	122
7.4.6	HQ avoiding task interruption	123
7.4.7	Consequence of task interruption	124
7.4.8	The consequence of missing player feedbacks	127
7.4.9	The incident of player death	128
7.4.10	Division of labour in HQ	131
7.5	Discussion	132
7.5.1	How does In-the-loop design play out?	132
7.5.2	Dealing with the imperfection of agent planner	136
7.5.3	Lack of explanation from agent	138
7.5.4	Unsupported forward planning activities	139
7.5.5	Coordination with field Players	140
7.6	Design Implications	143
7.6.1	Establishing common ground	143

7.6.2 Balance responsibilities between human and agent	144
7.7 Professional feedbacks	145
7.7.1 The Rescue Global feedbacks	146
7.7.2 Conclusion	150
7.8 Summary	151
iv CONCLUSION	154
8 CONCLUSION	155
8.1 Summary of approach	155
8.2 Summary of findings	155
8.3 Summary of interface evolutions	158
8.4 Summary of interactional issues	162
8.4.1 Division of labour	163
8.4.2 Establishing Common Ground	166
8.4.3 Faciliating Accountability	168
8.5 Limitation	168
8.6 future work	169
v APPENDIX	173
A APPENDIX TEST	174
A.1 Appendix Section Test	174
A.2 Another Appendix Section Test	175
BIBLIOGRAPHY	176

LIST OF FIGURES

Figure 1	HQ and field players in AO	30
Figure 2	The AO targets	32
Figure 3	The AO targets	32
Figure 4	Role target mapping	32
Figure 5	The radioactive cloud	33
Figure 6	System Architecture	35
Figure 7	Result for MMDP, Myopic and Greedy algorithms	37
Figure 8	Input and output of the agents	38
Figure 9	On-the-loop interaction design	41
Figure 10	In-the-loop interaction design	41
Figure 11	Log visualisation	49
Figure 12	Replay system	49
Figure 13	summary of methodologies	52
Figure 14	The HQ interface	56
Figure 15	The mobile responder app	58
Figure 16	Speech act classification	61
Figure 17	Adjacency pairs	64
Figure 18	JH (Behind Left), D2 (Middle Front), KY (Right behind)	66
Figure 19	KY (Left) , MF (Right) holding mobile phones	66
Figure 20	Caption	67
Figure 21	Caption	67
Figure 22	How responders addressed task allocation messages from HQ.	69
Figure 23	Caption	70
Figure 24	Caption	70
Figure 25	Caption	72

Figure 26	Caption	74
Figure 27	Caption	75
Figure 28	Caption	91
Figure 29	Caption	94
Figure 30	Caption	94
Figure 31	Result Overview	115
Figure 32	Task assignment overview	116
Figure 33	Task assignment in session 1	117
Figure 34	Task assignment in session 2	117
Figure 35	Interaction issues and interface improvements	156
Figure 36	Interface of study 1	159
Figure 37	Interface of study 2	160
Figure 38	HQ Interface of study 3	161
Figure 39	Mobile Interface of study 3	162

Figure 40 On-the-loop vs In-the-loop [163](#)

LIST OF TABLES

Table 1 Autem usu id [175](#)

LISTINGS

Listing 1 A floating example [175](#)

ACRONYMS

d

INTRODUCTION

Disaster response operations such as Urban Search And Rescue (USAR) can be very challenging. In large-scale disaster, DR teams may have limited resource and personnels 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 challenge requires DR teams to carry out highly coordinated activities in an uncertain task environment. The challenges create the opportunity 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 role in the disaster management. Moreover, Multi-agent system 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 intelligent planning support in the DR domain. However, before we apply this algorithm to support, we need to understand how can we configure the interactions between responder team and planning support system in a way that improve team rather than hinder. The aim of this work is aimed to explore potential interactional issues surrounding a planning support system, which in turn, informs interaction design of such systems.

This PhD work is a ORCHID sponsored research, which contributes to the understanding needed to build Human Agent Collectives (HACs)

in disaster domain. As computational systems becoming increasingly embedded into our life, the researchers from ORDHID project envision a future in which people and computational agents operate at a global scale, forming human agent collectives. The ORCHID project is aimed 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 systems (www.orchid.ac.uk).

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 thesis structure.

1.1 PROBLEM DEFINITION AND OBJECTIVES

In large scale disasters, Disaster Response team(DR) may have limited resource and personnel to deal with large amount of incidents across large geographic area under time pressure. In this situation, task and team allocation become a grand challenge for DR team. The responders and resources need to be assigned to teams and tasks in a way that to minimise 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., 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 leads to some real-time simulation and optimization technologies, some of which has great have potential to be adapted to support task planning for DR teams. One of a good example can be coalition formation algorithms devised by xxx (More details, see section x) Although the opportunity space has been recognised, most coalition formation algorithms has only be tested in computational simulations. None of them has be deployed to guide real human in DR situations. Many CSCW literatures have pointed out ill-designed work-flow management/automation system can lead to undesirable results, not only fail to improve work efficiency but also hinders 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. found that extreme difficulties might be encountered when introducing new technology support for human teams. New technologies might not support, but may disrupt smooth workflow if they are designed in an organisationally unacceptable way.

We believe the same is true for intelligent planning support. Before we can build intelligent systems that support human task planning, field trials are needed to understand the potential impact of technology support for team coordination. Although most multi-agent coordination algorithms have been tested to perform well in the computational simulations, they have never been excised to guide real human responders in the real world environment. Currently, there are few studies aimed to unpack the interactional issues relating to a socio-technical aspects of the intelligent planning support system. Interactional issues of a human-agent system can be defined as the issues related to interaction design and more importantly, the social aspects of the a human-agent. system. "Social issues" of human-agent systems are thought to be as important as technical issues []. There-

fore, this PhD work is aimed to fill this gap by exploring and unpacking interactional issues surrounding the intelligent planning support system from a HCI perspective.

1.2 APPROACH

To meet our research objective, we adopt a serious-mixed reality games approach (Fischer et al., 2012) to create a game probe (i.e. AtomicOrchid) that enables studying team interaction with planning support system in a real-world disaster scenario whilst providing confidence in the efficacy of behavioural observations. Mixed-reality games bridge the physical and the digital (Benford et al., 2005). Arguably, They serve as a vehicle to study distributed interactions across multiple devices and ubiquitous computing environments in the wild (Crabtree et al., 2006).

The AtomicOrchid [reference the AO2012 paper] is a serious mixed-reality game designed to mirror aspects of real-world disaster. In this game, field responders use smartphones to coordinate, via text messaging, GPS, and maps, with headquarters players and each other. The players in the game faces a distributed task planning problem with both time and spatial constraints. To achieve 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 software. The planning support agent is based on a state-of-art coalition formation optimisation technology. In Chapter x, design and implementation of AtomicOrchid will be introduced in more details.

In order to explore human agent interactional issues, three studies are conducted with different research focuses. In the 1st study, field

responders and incident commander coordinate without support of the intelligent planner. The study establish baseline performance of the game play and derived several requirements for planning support system. In the second and third studies, an intelligent planner was introduced to support task planning with two different interactional designs. The second study adopts a arrangement of human-ON-the-loop in which the planning agent automatically generate plans and instruct field players to execute plans. In the third study, we adopts an human-IN-the-loop design in which every plan generated by planning agent will need to be approved and edited before it is sent to field players before execution. More details of these two interactional arrangements will be introduced in Chapter x.

The work also adopted an ethnographically-inspired approach for data analysis. Game plays were recorded and qualitative interaction analysis were carried out to unpack human agent interactional issues.

1.3 SCOPING

This thesis is relevant to several research areas.

- Human Computer Interaction (HCI) for Disaster Response(DR).
With the vision of HACs system, the current ICT for DR may eventually evolve into HACs in the future. The thesis is aimed to help realise the vision by providing design implications for ICT systems with intelligent task planning agents from a HCI perspective.
- Multi-agent systems. The muli-agent simulation technologies underpin the technical possibility of intelligent planning support, providing the opportunity space of human agent collective planning.

- Human agent interaction. Existing human agent interaction research is the overarching research area of this thesis.
- Ethnography. The thesis adopts a Ethnographically-inspired approach to study human system interactions in field trials.

There are various ICT and AI-based technologies supporting disaster management activities in the different stages of the crisis circle including preparedness, response and recovery. This thesis is going to limit the scope on operations of rescue and evacuation in the immediate aftermath of a disaster impact, which typically requires high level of team coordination and real-time task planning and execution.

The thesis also focus on socio-technical issues related to human team interacting with the intelligent planning support from a HCI perspective. The work involves planning support agents based on multi-agent coordination algorithms, but the effectiveness of particular coordination algorithms are not concern of this work.

As part of ORCHID project, the AtomicOrhid serious game platform was developed as A research "Probe" to trial human agent collective planning in the domain of disaster response. As mentioned in section x.x, the AtomicOrchid platform consists of two major components: a game engine, and a embeded 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 Savapali Ramchun. Both Feng and Ramchun have expertise and research interest in the performance of task planning algorithm, where the author's research interest is the interaction between human and the intelligent task planning agent.

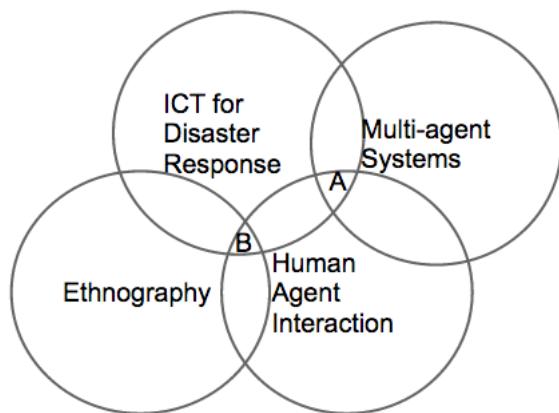
1.4 RESEARCH QUESTIONS

The recent advances in ICT and multi-agent optimisation technologies have created the opportunity space for intelligent planning support system in disaster response domain. Before we deploy such a planning support system, a deep understanding of interactional issues are required for appropriate interaction design between human teams and computational agents. This work adopted serious game approach to explore the interaction design space. Integrated with intelligent planning support, the AtomicOrchid game platform is used as a testbed for human agent interaction techniques. The AtomicOrchid is further configured with two different interaction techniques to produce game "probes" for field trials. Through field observation, this work is aimed answer the following two research questions:

- A What interactional issues will emerge if we try to automate planning process in a disaster response team? Interactional issues can be seen as an umbrella term for all issues relating to interaction design, which can range from social, organisational issues to interface design issues. This work is aimed to conduct an exploration of the interactional issues in the interaction design space of agent planning support.
- B How can we design interaction to support human agent collaboration in task planning? Following the first question, the emerged interactional issues will need to be handled with appropriate interaction design. This work seeks to produce interaction design implications through field observation and interaction analysis, both of which are grounded in literature review (Chapter x)

[The research is primarily interested in two interaction techniques, what are they and why?] Both two questions will be answered with respect to the two different interaction techniques which will be detailed in Chapter x.

1.5 CONTRIBUTIONS



This thesis contributes to the knowledge in the following areas:

- A A real-world interactive prototype and trials to investigate team coordination in a disaster response settings.
- B The field observation of serious game trials leads to enriched understanding of interactional issues surrounding human agent collective planning in disaster response domain.
- C For each study, field observations are further analysed to generate design implication which contribute to future deployment of agent-based planning support system.

1.6 PUBLICATIONS OF THIS THESIS

Parts of the contents of this thesis have been accepted by peer-review for publication in conference proceedings in HCI or are in submission:

Chapter x 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.

The exploration of Requirements for building coordination support system in Chapter x 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.

The exploration of interactional issues related to human-on-the-loop arrangement in Chapter x have 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.

1.7 STRUCTURE OF THE THESIS

This thesis is structured as four parts. Part I surveys the relevant background literatures. The chapter of literature review will firstly introduce a brief history of automation, agent software and Study of human agent interaction. In the following section, an overview of task planning activities will be introduced together with command and control structure of DR teams. In particular, the section will review the state-of-art agent simulation technologies that can help support task planning. The following section will review relevant CSCW Literatures which will lead to the potential challenges of real deployment of the agent technologies. The rest of the chapter give an overview of serious mixed reality games which underpins the foundation of research approach of this PhD work.

Part II develops the approach and methodology employed to study interactional issues of planning support system in two chapters. The first chapter develops framework of interactional arrangement under which interactional issues can be explored. The rest of this chapter will introduce serious mixed reality game as approach to study interactions ,followed by detailed description of a game used as testbed for this study - AtomicOrchid. Chapter 5 describes the methodology used to study the interactional issues. In particular, this chapter will describe ethnographic observation and interaction analysis, which is supplemented by interviews and questionnaires.

Part III covers observational studies in this thesis. Chaper x reports the first observational study with AtomicOrchid. This version of AtomicOrchid do not have planning support agent included. The study establishes baseline human performance of task planning and derives general requirements of communication support. The Chapter x give an account of second observational study of AtomicOrchid. In this study, an planning agent was built into the game with human-on-the-loop interactional arrangement. Chaper x reports the third field study of AtomicOrchid with human-in-the-loop arrangement.

Finally, Part IV concludes this thesis with a summary of contributions and future work.

Part I

BACKGROUND

You can put some informational part preamble text here.

LITERATURE REVIEW

2.1 AUTOMATION AND HUMAN AGENT INTERACTION

The study Automation design and Human agent interaction are two heavily intersected research areas which overarches this PhD work.

2.1.1 *A History of Automation*

The original goal of Automation is to replace the tasks originally performed by human with a machine. [1] . It can be defined as the execution by machine, usually computer, of a function previously performed by human [Parasuraman and Riley 1997] The tasks that can be automated is used to be limited by technical capabilities, but this is no longer the case. With quick growth of machine's speed and intelligence, the tasks that can be automated is rapidly increasing, including complex cognitive activities such as information analysis, planning and decision making.[RAJA] The boundary between human machine capabilities has blurred. With little cannot be automated, the automation designers have to make hard choice about what to automate and to what extent.

One traditional approach for automation design is to simply automate all system functions that can be automated easily in a cost-effective way, leaving the all 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 whole system with no

detrimental impact results from the automation. However this is not always the case, Large body of empirical work in automation design have shown that xxxx .

The other approach is to achieve division of labour between human and automation according to their strength and weakness. As in Fitts list [2] , a set of strengths and weaknesses of humans and machines is identified. Some study pointed the division of labour would not be as simple as a labour division according to strength and weakness[]. Firstly the time factor could be important because human and machine's availability may change overtime. Secondly By delegating the same task to machine, the nature of human tasks can be changed as well. Large body of work has shown clearly that automation does not simply supplant human activity but rather changes it, often in a way that is unanticipated by the system designer. [Coordination and Supervision interaction required, expand] [1] [Later , proposed by Licklider in the un-Fitts list [3], automation is aimed to leverage and extent human capability by using agents (not replacing).[1].]

2.1.2 *The Level of Automation*

More recently, another alternative model of automation to guide automation design is proposed by proposed by C. D. Wickens[],known as 10 level model of automation. The model 10 levels of automation (LOA) has been later extended by R Parasuraman [] with more detailed classification of automation types. The extended model of LOA informs study approach of this PhD work, which will be documented in chapter x. This section will give an introduction of this model.

Most tasks can be fully or partially automated, which implies that automation is not all or none, but can vary across a continuum of level[]. At the lowest level, all system functions are performed manu-

ally by human operators. At highest level, system are fully automated, taking over all system functions. In between this two extremes, 10 different levels of automation are proposed by xxx. As shown in table x.

2.1.3 *Definition of Software Agent*

Since creation of the term "software agent", there are a lot of debates about its definition . One definition commonly shared by the Multi-Agent system(MAS) literatures is that the software agents are designed to operate independently without constant human supervision. In the AI community, the software agent evolves from multi-agent system research which in turn, derived from the field of Distributed AI [6]. This strand of work investigates infrastructure, language and communication to realize coordinated agent software system. The goal was to specify, analyse, design and integrate systems comprising of multiple collaborative agents.[5]

More recently, the use of the word software agent become much more diversified. [5] has made attempts to investigate broader classes and types of agent. Nwana's [5] topology of agents identified three characteristics of agents, learning, autonomous and collabroitive, Based on this characteristics, software agents can be categorised into 8 classes, ranging from collaboration agents, information agents to interface agents.

With a broader definition, researchers begin to investigate broader issues surrounding development of agent systems, among which human agent interactions have attracted a lot of attention of researchers and practitioners.[get one reference here!!!]

2.1.4 Human Agent Interactions

Automation design and Human agent interactions have big overlaps as both concerns the impact of automation(by computational systems) on human performance. While automation design is trying to answer the question of what and to what extent the automation should be, the study of human agent interaction concerns the issue of interaction design related to development of human agent system, with the aim to develop agents which is good at teamworking with human operators.

The agent researchers adopted a view that human and agents are equally important team players in a human agent system.[[1](#)] It highlights the importance of mutual interaction between human and agents that can enhance the competencies of both human and systems. From this perspective, the aim of automation is no longer "replace" human but achieve an human-machine symbiosis through effective human agent coordination, which is similar to the view adopted by [xxx](#) in developing un-fitts list. As researchers realize the effective human-machine symbiosis requires sophisticated interactions design between human and agent, the "interactional" issues of automation is thought to be as important as technical issues.[\[Expand discussion about this issues.\]](#) [[1](#)]

Several interaction techniques have been devised by agent researchers to achieve effective coordination between human and agent. [\[expand on mixed initiative and flexible autonomy\]](#)

[\[define the interactional issues!!!!\]](#)

2.1.5 Issues in Human Agent Interactions

Combine both

Why the social issues of designing agent system is important? [Norman1994] state lesson from CSCW

2.1.6 *A summary*

Because this PhD work sits in the overlapping research area of human agent interaction and automation design, the later chapters will consider automation and agent software as two interchangeable concepts.

2.2 CSCW WORKFLOW MANAGEMENT

2.3 TASK PLANNING IN DISASTER RESPONSE

What is coordination(Coordination theory) Team Coordination. Malone (1990 361) defines coordination as the act of managing interdependencies between activities performed to achieve a goal. One of the very important component of coordination in DR, following sections will firstly... and then discuss how the coordination is carried out through command structure of DR team.

2.3.1 *Task Planning*

In disaster response, team coordination is essential in order that groups of people can carry out interdependent activities together in a timely and satisfactory manner (cf. Bradshaw et al., 2011). Disaster response experts report that failures in team coordination are the most significant factor in critical emergency re- sponse (Toups et al., 2011: 2) that can cost human lives. Shared understanding, situation awareness, and alignment of cooperative action through on-going communication are key requirements to enable successful coordination. Convertino et al. (2011) design and study a set of tools to support common

ground and awareness in emergency management.

One important characteristic of large-scale disaster is the presence of multiple spatially distributed incidents (Chen et al., 2005). 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). Depending on the proliferation of incidents, response personnel may need to dispatch, deploy and redeploy limited resources. Coordination is required to efficiently allocate limited resources to multiple incidents with temporal and spatial constraints imposed by the nature of disasters.

(The requirements of multiple incident response from Chen.)

((Coordination game) Below confirms that in a single target need multiple responder teams) While the present work applies generally to disaster response, our iterative design and theory-building processes have been specifically informed by work practice in the sub-domain of fire emergency response. We work from fire emergency response in small-scale structural fires, observing practice in the United States of America. Fire emergency response is undertaken by small teams distributed throughout the incident, coordinated by an incident commander (IC) [Toups and Kerne 2007; Landgren 2006; Jiang et al. 2004; Landgren and Nulden 2007; Wieder et al. 1993; Carlson 1983; U.S. Department of Homeland Security 2008]. Multiple response teams, or companies, are dispatched to any incident and cooperate around the fireground (Figure 1). A company officer leads each team, which consists of firefighters and/or engineers.² Normally, each company is associated with a firefighting vehicle; an apparatus, such as an ambulance, engine, or ladder truck.

The fireground and surrounding space constitute a dangerous and dynamic interface ecosystem [Kerne 2005] of distributed cognition, connecting responders, victims, fire-fighting equipment, communi-

cation media, and information artifacts. Upon arriving at an incident, multiple companies distribute in and around the fireground. Companies and their apparatuses are placed at strategic locations, and are moved as needed. Human operators work on and from these platforms. Firefighters and rescue workers deploy from them, taking equipment into the fireground; equipment, such as firehoses and radios, may be technologically supported by the apparatus itself (pumps and water sources, or high-power repeaters, respectively). Each apparatus, and in many cases, each human worker, is equipped with a half-duplex radio to facilitate long-range, broadcast communication.

(Geographic distribution as also a key factor)

Muti-objective nature of the DR response lead to its connection with multi-agent simulation and optimisation.

Therefore task planning envlove xxx challenges and opens the opportunity for technology intervention. Current state of art practice will be reviewed in section.....

2.3.2 DR Command structure

A relatively generic model of C and C is proposed by The development of Information Technology System (ITS) to support the emergency response team work has many work published in the literature. For long time, crisis management systems were based in the military model of CC [1]. For many authors [2, 3 and 4] the need to change the theories of emergency management creating new paradigms is imperative to improve the flexibility of the CC structures. Their aim is to make them more efficient, multi-disciplinary and multi-institutional, increasing the collaboration between CC and the field responders and allowing sharing planning and resources to stabilize the crisis [5]. Stanton et. Al. [7] proposes a new generic model of CC based

on field observation. One important conclusion he made concerning common characteristics in the different domains of CC: 1. The presence of a remote control room; 2. The great dependency on verbal communication and; 3. The existence of collaborative discussion between field teams and command.

Gold-Silver-Bronze control model is documented in xxxx(wiki)

2.4 TECHNOLOGY SUPPORT FOR TASK PLANNING

This section reviews state of art technological practice to support task planning. Task planning toolbox containing various components including data sets, (meta-)information, storage and query tools, analysis methods, theories, indicators.

Greetman review of planning system. not real time, The systems can be categorized according to several criteria (Aims, Capabilities, Content, Structure, Technology). None of them is adapted to real-time coordination.

Multi-agent paraghim

Graph to position the my system.

2.4.1 *Disaster simulation, optimisation*

The section will briefly review current status of the simulation and optimisation technologies for disaster response.

2.5 GAME BASED SIMULATION

2.5.1 *Serious Game*

There are various definitions for the terms ‘serious game’ in different literatures. One issue that most of the literatures agreed on about ‘Se-

'serious game', is that the term is concerned with the use of games and gaming technology for the purposes other than mere entertainment, which include education, training, healthcare, and advertisement. Although serious games have look and feels of game, They are actually 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. (Corti, 2006; Squire and Jenkins, 2003).

Recently, the serious games have been increasingly 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(Michael and Chen, 2006; Squire and Jenkins, 2003). One good example of serious game for disaster response is the Biohazard developed by MIT Comparative Media Studies.

The game is 3D . It is designed to help emergency first responders deal with toxic spills in public locations (including those caused by the deployment of weapons of mass destruction). Emergency responders work in teams to organize the response to a gas attack in a crowded suburban shopping mall. Players race against the clock to save as many civilians as possible. The game play involves quickly scanning and assessing the situation, dividing into teams, and coordinating efforts so that the likely point source is identified and isolated before more civilians (as well as the player) become sick. Students can thus practice recognizing the signs of different chemicals and viruses, examining victims' symptoms, and observing how different chemicals spread in differing conditions.

Another example could be the team coordination game developed by Toups, Kerne and Hamilton (2011), which teaches participants ef-

fective cooperation and in particular 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[].

The advantage of serious games in disaster response domain may be obvious. Actually releasing deadly agents into the environment is too dangerous. By changing only a few simple variables, players can experiment with a multitude of different conditions, examining how strategies should differ, depending on whether the mall is crowded or vacant, whether the contaminant is released near an air vent, whether the air conditioning is on or off, or how much of the contaminant is released.

limitation. People can be unpredictable, particularly in highly stressful, emergency situations. Dealing with a panicked public demands a type of interpersonal skill expressed through voice, gestures, and demeanor or that contemporary game interfaces cannot capture. Despite the many advances in artificial intelligence in games, characters still can't express a full range of emotional responses in dynamic reaction to events. Clearly, Biohazard is no replacement for experience or even field trials and manuals. Rather, it is a tool that responders can use to explore ideas and talk about their practice. [paraphrase] — people in DR can be , the game based simulation may fail to simulate panicked public demands and emotional human response. Therefore, the game may not be enough for training of the human and social aspects of DR domain. Therefore

2.5.2 *Mixed Reality Games*

Ubiquitous gaming is one of the most exciting ideas to emerge from the games industry in recent years. Ubiquitous games can be played anytime, anywhere and often play out across multiple media. For ex-

ample, Electronic Arts' Majestic uses cell phones, desktop computers, and fax machines (among other technologies) to immerse players in a multiplayer conspiracy theory game. Such games motivate players to investigate, weigh evidence, compare notes, test hypothesis, and synthesize information as they draw conclusions about what has occurred and why. We use the terms enhanced reality or augmented reality to refer to virtual experiences being played out in real-world spaces. While such game experiences may seem prohibitively expensive, recent handheld technologies, which can now combine cellular phone, GPS, Bluetooth, wireless Internet, multimedia capabilities, and infrared technologies into one machine, make such experiences possible even on the level of an individual school.

Due to some characteristic of, it is arguably a good way for simulating disaster response situations. Therefore, it become a vehicle for study.

Compare to the computational simulations, it could be an good alternative.

2.6 MIXED REALITY GAMES

This chapter takes an in-depth look at the methodology that underlies the empirical approach in the presented studies.

2.6.1 *Definition and examples*

2.6.2 *Game for Serious Purpose*

Part II

METHODOLOGY AND APPROACH

APPROACH

The PhD study adopts a socio-technical view towards the planning support systems in disaster response setting. Introducing a planning support to Disaster Response(DR) operations may create a socio-technical gap that need to be considered by system designers. We argue that the gap can be reduced by an appropriate interaction design supported by deep understanding 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 human and planning support agent in disaster setting [missing?]. A serious mixed reality game (MRG) approach is adopted to create MRG AtomicOrchid(AO), which are used to simulate DR operations. Using the AO as a testbed, we outlined two potential interaction designs that are later deployed for field observations. In addition, efforts have been made to establish contact with professional response agency Rescue Global(RG), which leads two workshops entered on AO. Professional feedbacks on AO are collected from the workshops.

In this chapter, we will go through the Socio-technical Perspective of planning support system adopted by this PhD work (section), followed by introduction serious mixed reality game (MRG) approach (section 3.2) and description of AtomicOrchid platform (section 3.3). The section x will outline the two interaction designs that will be deployed to AO for field studies, and the section x gives an introduction of workshops with RG for professional feedbacks.

3.1 SOCIO-TECHNICAL PERSPECTIVE OF TECHNOLOGY SUPPORT FOR ORGANIZATIONAL WORK

This research is aimed to inform the design of planning support for organizational work conducted by responder teams. This PhD work adopts the 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 challenge will be encountered for introducing the planning support to disaster response operations.

The term socio-technical systems was originally coined by Emery and Trist (1960) to describe systems that involve a complex interaction between humans, machines and the environmental aspects of the work system. nowadays, this interaction is true of most enterprise systems. The corollary of this definition is that all of these factors people, machines and context need to be considered when developing such systems. Within the context of CSCW the term stands for the recognition that aspects of both, technical as well as social subsystems, need to be considered when an organization introduces new technology and that there is a 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 and In contrast, technical systems are characterized by artifacts, control, anticipation, state-transitions, pre- programmed adaptability, learning in respect to purposes which are determined from outside the system. Introducing a tech support system into an organization requires the technical system and a social system to be integrated, which causes lots of well-known problems in CSCW studies.

In the context of this study, the technology support is based on coalition formation algorithm, which requires a set of rigid inputs and produces task assignments. On the other hand, planning activities of responder teams are characterized by natural social processes such as communication, negotiation and cooperation. To support the responder teams with coalition formation technologies, the confrontation between social and technical is inevitable.

Some researchers have pointed out the existence of the inherent social technical gap, the great divide between what we know we must support socially and what we can support technically. Some argued that human activity is highly flexible, nuanced, and contextualized and that computational entities such as information sharing, roles, and social norms need to be similarly flexible, nuanced, and contextualized. However, current technology support systems for organisations are often rigid and inflexible, failing to fully support the social world. Computer science is learning how to use techniques such as machine learning, user modelling to fill social-technical gap. It is hard to disprove that a technical solution is imminent. However, some argued that such a technical solution is unlikely, given that computer science, artificial intelligence (AI), information technology, and information science researchers have attempted to bridge the gap without success for at least 20 years.

We argue that the social aspect does not need to be fully supported through technological advance. A deep understanding of social issues and appropriate interaction design may lead to possible “workaround” of the socio-technical gap. Although technology support may be not fully integrated with the social aspects, We believe appropriate interaction design can reduce its negative impact on social process so that the benefits of technology can overweight its adverse social impact.

3.2 SERIOUS MIXED REALITY GAME 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, Serious game approach has been adopted by researchers to study technology interaction in disaster scenarios through game-like simulations, for example to prepare first responders for scenarios in which hazardous materials are involved (Losh, 2007). Abbasi et al. (2012) present a study in which locally distributed participants played the role of victims asking for help via social media in a simulated crisis, and participants that played the role of first responders used a coordination system to filter messages and mobilize the appropriate responder teams according to their assigned capabilities. (see section x).

The PhD work also adopts serious game approach to simulate a disaster response setting in which distributed responder teams are coordinated with a time and spatial constraints (see section x). 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. The Mixed-reality games are recreational experiences that make use of pervasive technologies such as smart phones, wireless technologies and sensors with the aim of blending game events into a real world environment. The MRGs serve as a vehicle to study distributed interactions across multiple devices and ubiquitous computing environments 'in the wild' (see section ??sec:LRMRgame)).

The MRG testbed called AtomicOrchid (AO) simulates a radioactive incident. Participants of the game plays the role of responders 'on

the ground', coordinating 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 on the game interface. Through agent integration and interface modifications, we created three 'probes' of agent planning support with different interaction designs. The three probes are then be used to conduct behavioural studies, which allows us to unpack human-agent interaction with different interaction designs.

3.3 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 (play as field responder) and online players in a control room (play as Headquarter).

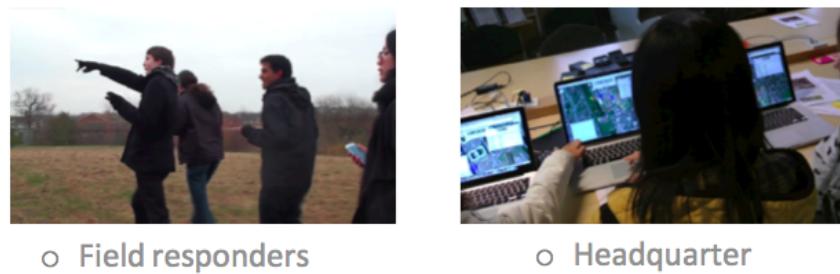


Figure 1: HQ and field players in AO

In the following sections, we describe the game design including grounding of the design rationale, iterative design process, and the system architecture.

3.3.1 Game Design Rationale

The AtomicOrchid is based on the fictitious scenario of 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 other than 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 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. Whilst formal response teams tend to use radio to communicate (e.g., Toups et al., 2011) we chose text-based messages for its flexibility to support scenarios with many distributed (volunteer) field responders.

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

The game's two-tiered organisational structure is derived from real world disaster response organisation and from NIMS (Homeland Security, 2008). The game's HQ is loosely modelled on sector coordinators, whose role is to manage resources and communications between their assigned teams, and command and coordinate action within their sector (INSARAG, 2012). Field responders are modelled on team leaders and members. We ignore this distinction to simplify roles, as

signments, and game mechanics.

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

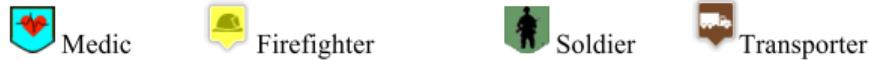


Figure 2: The AO targets

There are four types of (virtual) targets:



Figure 3: The AO targets

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 are required in immediate proximity to 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 (see figure 4).



Figure 4: 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. This is an example of what Toups et al call, information distribution (2011).

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). In require communication between HQ and field responders, the spatial position and movement of the cloud is only known to HQ. The cloud is shown in a heatmap style in the figure 5.

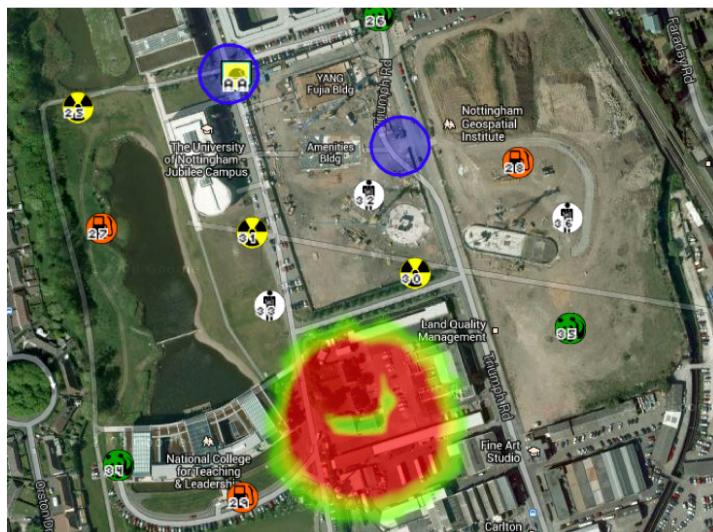


Figure 5: The radioactive cloud

Command-and-control structure. The division of responsibility into HQ and field responders simulates a situation where volunteer responders are connected to a simple two level Command-and-control structure, similar to the real-time layer of the existing professional disaster response organizations (e.g., Chen et al., 2005).

System interface. System interface design is closely related to specific interaction designs, and it keeps evolving throughout three iterations of field trials. Therefore, the details of interface evolution is left

to be introduced in the subsequent chapters (Chapter 5,6,7) of field trials.

3.3.2 *Iterative design and development*

Before the game is deployed for observational studies in this PhD work, the game went through a iterative design and development process to test, refine game concepts and system robustness. We briefly describe three cycles of iterative game design and evaluation before the system is 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 technology prototype was first tested with users in the second iteration. Users were equipped with the responder smartphone app to communicate, navi- gate, 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 les- sons learned in the pilot study revealed problems with user interaction, network- ing, and game parameter tuning, which we subsequently addressed.

In the third iteration, we improved system stability and interface designs. We conducted a pilot study at the campus of another university, to test the system in place. The full-fledged study we report on here was conducted shortly thereafter.

3.3.3 System Architecture

The AtomicOrchid is based on the open-sourced geo-fencing game MapAttack 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 6. Client-side requests for less dynamic content use HTTP. Frequent events, such as location updates and radiation exposure, are streamed to clients to avoid the overhead of HTTP. In this way, field responders are kept informed in near real-time.

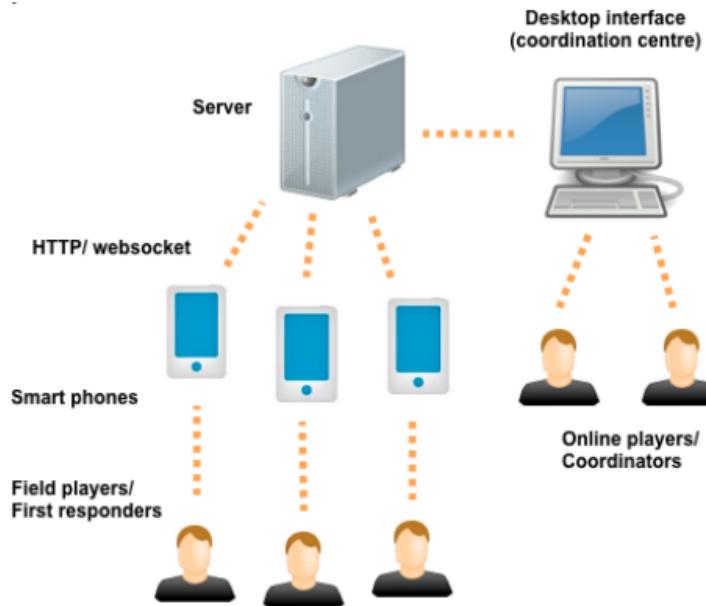


Figure 6: System Architecture

The platform is built using the Sinatra for Ruby, and state-of-the-art web technologies such as socket.io, node.js, redis and Synchrony for Sinatra, 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.

3.3.4 *The planning agent*

In study 2 and 3 (chapters 6,7), planning agents are integrated into the AtomicOrchid to support player's planning activities. Two types of agent has be used in study 2 and 3 respectively. In what follows, we briefly describe technical details of the agent and system integration between AO and the agents.

The coordination problem (described in section 3.3.1) is modelled using a Multi-Agent Markov Decision Process (MMDP) that captures the uncertainties of task execution, extending earlier work []. The modelling 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-sized problem. Hence, we devised an approximate solution 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 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 our MMDP algorithm against greedy and myopic methods (see figure 7). The results confirm that our algorithm produces efficient task allocations. It should the agents are developed by ORCHID Research

partners from Southampton, more technical details of the planning agent is available in [JAMASS paper].

<i>Metrics</i>	<i>MMDP</i>	<i>myopic</i>	<i>greedy</i>
#completed tasks	71%	65%	41%
#responders survived	100%	25%	0%

Figure 7: Result for MMDP, Myopic and Greedy algorithms

For integration, the agent is deployed on a separate server. It communicate with the AO game server through a pre-defined HTTP protocol (for details, see Appendix X). The agent takes game status from game server as input, which includes player's health, road connectivity, locations of players, targets and radioactive clouds. The output of the agent a set of task assignments like 'player A and player B, go to target C' (see figure 8) . The task assignments are sent to the AO game server and present to game players. Detailed interaction design between human and the agents will be presented in Chapter 6,7. In order to facilitate the different interaction designs, the input of the agents are slightly different between study 2 and study 3, which will be detailed in section 6.2.1 , 7.2.2.

3.4 EXPLORE INTERACTION DESIGNS WITH THREE ATOMICORCHID STUDIES

[Justify the relation between the three iterations] - avoiding pitfalls that undermines observation of interaction arrangement - HQ agent interaction can be parallel/ HQ FR interaction follows progressive design interaction - The first iteration: base case/ need to understand the organization before we do anything.

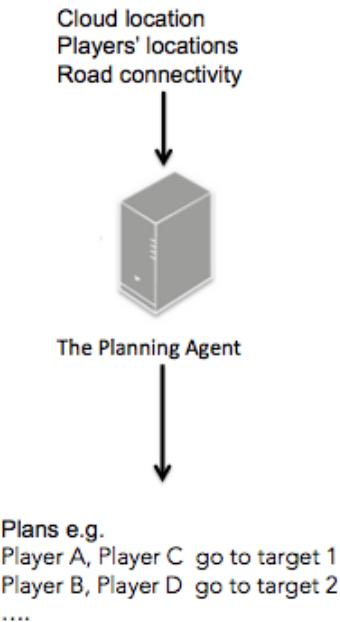


Figure 8: Input and output of the agents

Based on serious game approach, three studies are 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 agent. Inspired by the model of Level Of Automation (LOA) from the research of automation design, we outlined 4 paradigms of human agent interaction loosely based the automation level of planning activities: Full manual, Human-in-the-loop, Human-on-the-loop, and Human-out-of-loop. Arguably, the paradigm of Human Out-of-loop is believed to be unrealistic compared to In-the-loop, On-the-loop and full manual. Therefore, this PhD work will only consider the latter 3 notions of automation.

In research of automation design, the LOA model has been developed to categorise systems into a linear spectrum according to degree of automation (reference to literature review). Arguably, the model may not fit into context of socio-technical system due to some of its

limitations identified in section x.x.x. 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. Out-of-the-loop system is supposed completely run independently. Human is replaced with machine, therefore no human system interaction is required. It is unlikely to be realized in a socio-technical system in which organisational work is mainly carried out by human 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 minimum level of human intervention. Compared to Out-of-the-loop, the On-the-loop system is designed run without human intervention at 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 goal of the socio-technical system.
4. Full manual In full manual system describe a system without automation. In the context AtomicOrchid, the platform without integration of planning agent can be seen as a full manual system.

In the context of AtomicOrchid platform, the notions of In-the-loop and On-the-loop can be used to describe the degree to which the planning agent automate the real-time task planning and to what extent

human Headquarters need to be involved in the plan-execution loop. Guided by the 2 notions, we devised two detailed interaction designs for integrating the planning agent into AO game. In next the two sections, we will detail the two interaction designs, followed up by an overview of three field studies, which details how a series of system prototyping and field trials are organised based on the two interaction designs, and how they are designed to serve the research objectives.

3.4.1 *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 10). In this design, the agent can directly contact field responders to allocation tasks. The responsibility of the planning agent is to generate and send plans directly to field responders. The agent is also responsible for initiating re-planning according to the changes of game status. The agent can also directly handles feedbacks from the agent. i.e. the field players can feedback to the agent by accepting or rejecting the plan, while the agent can generate new plans according to the feedbacks.

The role of the HQ is to monitor the planning process and provide support when contingency rises. For example, the HQ may decide to stop some tasks issued by the agent if 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.

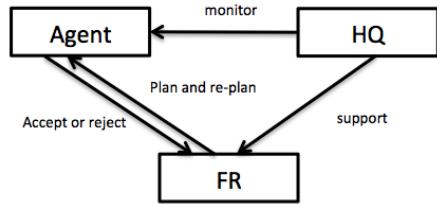


Figure 9: On-the-loop interaction design

3.4.2 *The In-the-loop interaction*

The In-the-loop interaction is designed to facilitate a different pattern of labour division between humans and agent: a planning agent propose the task assignments, while the human HQ need to approve the tasks before it is 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 don't agree with a task allocations from agents, they can intervene by directly editing part of the plan or require the agent to re-plan.

The feedbacks from the field responders (i.e. accept/reject) are delivered to HQ before any actions are taken. The HQ are responsible to review the feedbacks and decide the actions to be taken (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 can not operate without HQ's input, i.e. the HQ have to make decisions on every agent proposed tasks, and take actions on the feedbacks from field responders.

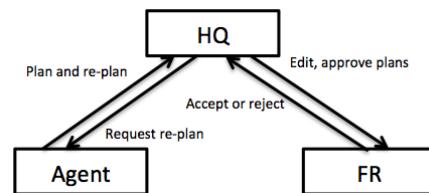


Figure 10: In-the-loop interaction design

3.4.3 *The three AO game studies to explore socio-technical issue of human agent interaction*

Three studies are conducted to explore the socio-technical issues related human agent interaction. The first study focus on a ‘manual’ version of AtomicOrchid without agent support, while the latter two study focus on On-the-loop and In-the-loop respectively. For each study, we develop a game probe through prototyping, to facilitate the interaction design to be studied. Through field trial of the AO game probes, The studies seek to unpack human agent interactions with the three interaction designs.

The first study is aimed to observe and explore human coordination without planning agent support. The non-agent trial supports the two later (chapter 6, 7) agent-supported system trials by 1) Revealing baseline performance of human coordination without agent support 2) Generate design requirements which feeds into subsequent prototyping of AtomicOrchid. The purpose of the second and third studies are aimed 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.

3.5 COLLABORATION WITH PROFESSIONAL DISASTER RESPONSE ORGANISATION

In addition to the three observational studies of AtomicOrchid games, two workshops with a Rescue Global (a professional disaster response agency) was organised to get professional feedbacks about the AtomicOrchid system and planning support component. Because the contact with Rescue Global was established in very late stage of this PhD work, the feedbacks from Rescue Global (RG) workshop are not used

to drive the development of AO simulation and interaction design, but to get an insight into similarity and difference between AO simulation and the real world disaster response(DR) operations, which help us understand limitations and strengths of our observational study. The first RG workshop happened between study two and study three (section 7.7). The In-the-loop AO probe was demonstrated to RG team and a discussion was organised to get feedbacks from RG. The second RG workshop happened after the study three (section x). It contains a hands-on session for RG to experience AO game, and the feedbacks are collected from discussions during and after the game session.

3.5.1 *Introduction of Rescue Global*

[Need paraphrase] 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.” [30]. RG has adopted a framework of procedures that implements ISO 9001 Quality Management (QM) principles, which commits RG for example to conduct risk assessments and to record decisions for accountability purposes. This has implications on the ways in which missions are planned and carried out (the focus of our field work).

[Example of operations carried out? , give context for next feedback section below]

RG’s organisational structure represents a typical hierarchy found in emergency services (cf. [39]), termed Gold, Silver and Bronze. Gold

denotes the strategic lead, which is associated with RG's senior officers (often referred to as the 'head shed') 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 onsite. Bronze refers to the operational level, in which 'Pathfinders' (field responders) carry out operations 'on the ground' supported by Silver command [29]. RG's core staff consists of around 20 highly specialised experts and admin support, many of whom have had prior careers in the military, and emergency and first response services.

METHODOLOGIES TO INVESTIGATING 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 study is aimed to conduct ethnographic-oriented field studies based on AtomicOrchid(AO) platform to generate descriptive results, which contains rich interactions among participants and planning support system. Ethnographic observations and interaction analysis are central to all three field studies, while group interviews, message classification, and system log analysis are introduced to supplement the two former in-situ methods.

4.1 ETHNOMETHODOLOGY PERSPECTIVE

The method of observing participants in the field study is informed by Ethnomethodology (EM). Following the tradition of ethnography, EM seeks to explicate real-world organisation of works by adopting the naturalistic stance. The 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. The EM-informed ethnography arguably helps answering what might be regarded as an essential question in design: what to automate and (Crabtree et al. 3) what to leave to human skill, competence, judgement, experience and expertise. By producing description of the actions and practices in and through which the work ‘gets done’ time and time by the members, The EM could inform the system design by uncovering what actions and activities

we should therefore support.

For the purpose of this thesis, the social situation the interaction with and around the planning support was argued to be a critical factor to understand how social organisation of work is achieved participants with the existence of a planning support system. Observation of the situated actions and practice employed by the participants was 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 ethnographies around technology use (Crabtree et al., 2006). The next section will go through the method of video-based interaction analysis for unpacking the interactions observed in the field.

4.2 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). In the context of HCI study, it is a method of analysing video-recordings of naturally occurring talk and activity, with the aim of uncovering, 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 relies upon report from participants, rather than actual, reasoning and behaviour. The over-reliance on participants' report make those methods vulner-

able 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 number of participants. The 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.

For the purpose of this PhD work, interaction analysis is applied to the evaluation of game probes undergoing field trials in selected work settings of Disaster Response. 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.

4.3 MIXED METHODS APPROACH

Ethnographic observations and interaction analysis have been introduced as the main methods for investigating socio-technical issues in our studies. Some additional methods companies the two main methods by providing extra insights and context information. In particular, the group interview supplements field observation and interaction analysis by providing subjective description of game play experience. The message classification method gives an quantitative

insights into remote communication. It also provides context for interactions in the field and help to identify interesting game events. The log analysis produces game events visualisation and replay. When triangulated with the video data, the log data analysis also supports interaction analysis by providing context and help to identify interesting episodes in video data.

4.3.1 Log data analysis

Now days, the HCI researchers often collect rich dataset for investigating interactions. The data set becomes larger and larger as digital record systems increase in availability. Automated tools are increasingly necessary for managing the organisation, replaying, structured and free coding (and annotation) and analysis of these growing data sets (). The logging system of AtomicOrhid produces time-stamped system logs. In order to triangulate the log data with video and audio data from field observations, A data visualisation tool and a log replay system are developed.

To recap (see section xxx), the logging systems records players's location, health, targets' location and status (pick up/drop off), task assignments and players' feedback (reject/accept). The data visualisation tool (see figure xxx) 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 (see figure 11). Detailed information of the event can be displayed when mouse hover on. As you can see for figure 11, the deep blue dot with 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 game-event sequence. It assists interaction analysis by providing context and guidance on episode selection process.

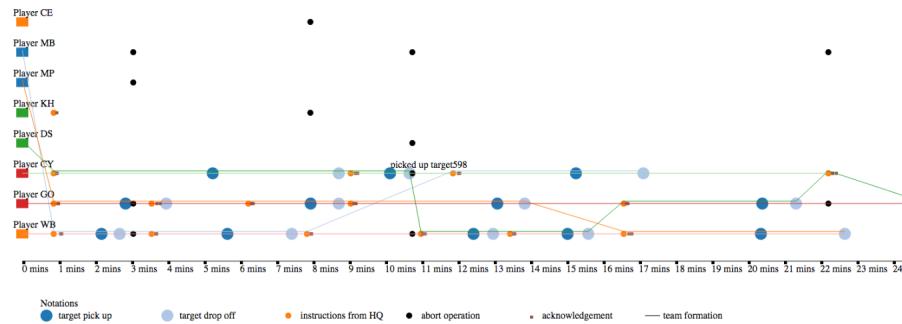


Figure 11: Log visualisation

A replay system is also built to triangulate multiple video with log data. The main map view on the replay interface (see figure 12) displays game status reconstructed from log data, in a way that similar to the HQ interface does (see section x for details). By giving a time offset to video file, the videos are synchronised with the main map view of game status. The replay system is important tool for interaction analysis, as it presents distributed game play with a single interface in a synchronised way, providing insights into the distributed interaction among participants and the system.

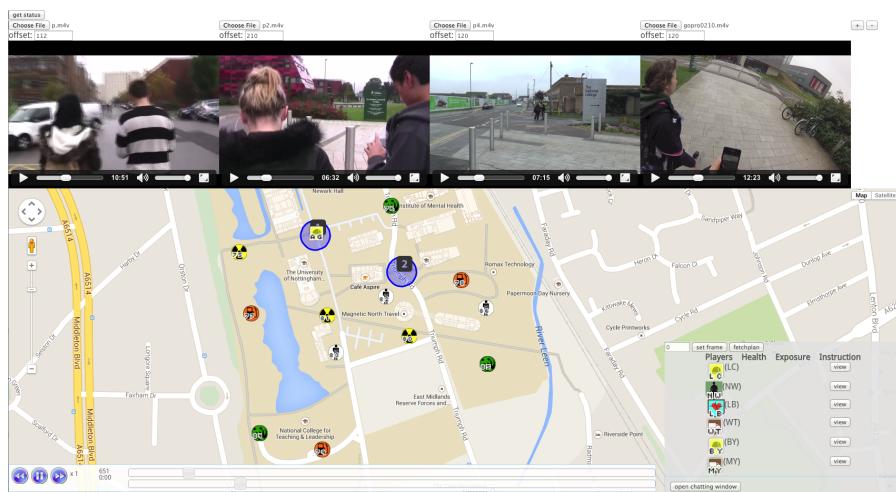


Figure 12: Replay system

4.3.2 *Message classification*

For AtomicOrchid, remote coordination between field and human HQ is achieved through a text messaging channel. To understand how the team members interact through the remote messages, we devised a message classification method based on speech act theory (Searle, 1975). We used speech-act theory and the notion of adjacency pairs in linguistics 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. We were primarily concerned with the illocutionary dimension of speech acts.

Searle's classification of illocutionary acts (Searle, 1975) 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. In the Radiation Re-

sponse Game, we expected many actions in remote conversation to be accomplished through pairs of utterances such as request-response, question-answer, or inform-acknowledge.

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

4.3.3 *Group interview*

[read andy's interview chapter] For all three field studies, group interviews were conducted with all participants after each game sessions. The interviews consists of open-ended questions with the aim to supplementing the field observation and interaction analysis with rich descriptions of participants' experiences of the game. Emergence of unanticipated issues and events was also fostered by asking open-ended questions, which in turn, are used to guide the interaction analysis of field observations.

4.4 SUMMARY

To sum up, ethnographic observation (?) and interaction analysis are the main in situ methods applied in the studies . The two methods are supported by message classification, group interviews and log data analysis. Message classification produces a message catalogue which gives quantitative insight into the remote communication. It also contribute to episode selection and provide context for further episode

analysis.

The log data analysis produces replay and visualisation of game events. The replay system help to triangulate game events in log data with multiple videos from field observation, revealing the relationship between system log and interactions in the field. It also help us identify important episodes of interactions and provide context for further episode analysis.

Comments from participants are collected from group interviews. The comments give insights into participants' subjective game experience, which are also important context for episode analysis.

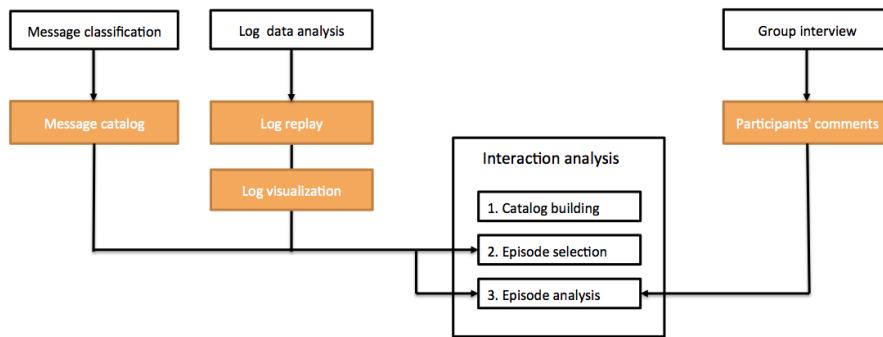


Figure 13: summary of methodologies

Part III
STUDIES

ATOMICORCHID STUDY 1: NON AGENT VERSION

In this study, we focus on investigating requirements for time critical distributed team support relevant for domains such as disaster response. In this Field responders of the AtomicOrchid game use smartphones to coordinate, via text messaging, GPS, and maps, with headquarters and each other. 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 HACs system to support team coordination and uncover requirements that highlight the role of local coordination, decision-making resources, geospatial referencing and message handling.

5.1 INTRODUCTION

Disaster response (DR) has been characterised as highly coordinated, time-critical collaborative activities [4]. Coordination is essential in such settings so that time critical interdependent activities such as search and rescue can be completed in a timely and satisfactory manner [1]. Opportunity space for building ‘intelligent’ coordination support for such activities has been recognised by the researchers of HACs systems (see chapter 1). However, little study has explored the design space for HAC systems to support time-critical coordination settings. Therefore, little is known about the challenges and requirements in building systems to support team coordination in such settings.

Due to the critical nature of the disaster operations, it would be hard to design and deploy ‘intelligent’ coordination in the field before we thoroughly explored the requirements of interaction design. On the other hand, computational simulation of an ‘intelligent’ system is fundamentally insufficient for studying interactional issues (see section ??). Therefore, in this study, we are aimed to use AtomicOrchid game as a game probe to uncover the requirements and design implication for building ‘intelligent’ coordination support system.

The AtomicOrchid game creates socio-technical setting in which player teams plan and executes spatially distributed tasks (see section 3.1). Although HACs researchers has envisioned that an intelligent agent can team coordination by providing computational optimised task allocations in real-time, we focus on a base version of AtomicOrchid which does not involve any agent support in study. The aim is to unpack how human teams coordinate in the time and space constrained task setting, establish baseline performance of human coordination in the game, generate design requirements and implications for HACs systems, which in turn, supports our later system prototyping and studies.

This study is the first of three iterative system trials in this PhD work. The non-agent trial supports the two later (chapter 6, 7) agent-supported system trials by 1) Revealing baseline performance of human coordination without agent support 2) Generate design requirements which feeds into subsequent prototyping 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 prototyping.

In the following section, we will go through AtomicOrchid system description with a focus on interface and interaction design.

5.2 SYSTEM DESCRIPTION

Basic game mechanic and system architecture has been introduced in the chapter 3. This section will give a detailed description of the system interface that support coordination between the field responders and headquarters players.

The HQ is manned by two coordinators. All of the coordinators are provided with a web-based coordination interface (see figure 14). The interface gives them an overview of the game status and enable them to communicate with the field responders.

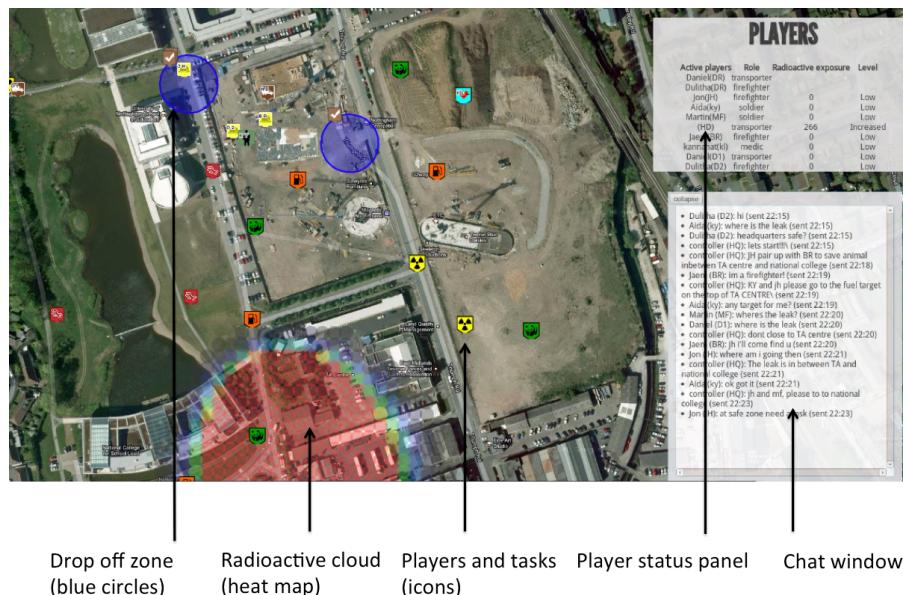


Figure 14: The HQ interface

As you can see in the figure 14, the majority space of the interface is occupied by a map-based presentation of the 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 on 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 play through the mobile and HQ interface.

Field responders are equipped with a mobile responder app providing them with sensing and awareness capabilities (figure 15). There are two tabs in the responder apps. The "map" tab displays a map showing locations of field responders and targets, which is similar to the map on HQ interface, except that the radioactivity is not shown. The radio level of 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 field player to receive and send messages.

The app shows a reading of radioactivity, their health level based on radioactive exposure, and a GPS-enabled map of the game area with the targets to be collected and the drop off zones for the targets. Icons according to responder roles that additionally have their initials on them can be used to identify individuals. Another tab reveals the messaging widget to broadcast messages to the other field responders, and to headquarters.

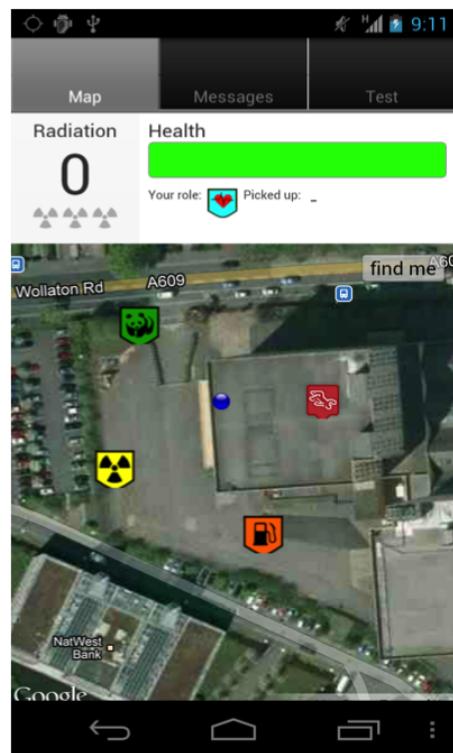


Figure 15: The mobile responder app

5.3 STUDY DESIGN

To explore socio-technical issues around team coordination, we ran two Radiation Response Game sessions, with volunteers recruited from the local university. 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 (8 female); 7 participated in session A and 11 in session B. All participants were reimbursed with 15 pounds for 1.5 hours of study. The majority of participants were students of the local university. Procedure. 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. The post-game session consisted of a questionnaire aimed at collecting participants’ feedback on (1) first impressions of the game; (2) usability of the system, and; (3) coordination issues in the game. A group interview was then conducted, before participants were debriefed and dismissed.

The size of the game area on the local university campus was 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. The pilot study showed that this was a challenging, yet not too 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.

We took a mixed methods approach to data collection and analysis. In addition to quantitative questionnaires, a semi-structured group interview was conducted aimed at eliciting important decision points, strategies and the overall decision-making process. Furthermore, 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.

We developed a log file replay tool to help with data analysis of time stamped system logs that contain 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.

Interaction analysis of local coordination 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 (cf. Heath et al., 2010). Key decision points in team- ing 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 this paper.

Message classification How are remote messages 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. According to speech act theory, utterances in dialogues can be considered as speech acts from three dimensions. We were primarily concerned with the illocutionary dimension of speech acts. Searle's classification of illocutionary acts (Searle, 1975) is used to categorize messages in the communication system.

5.4 DATA ANALYSIS AND RESULTS

Here, we present findings from interaction analysis supported by message classification that reveal how team coordination was achieved.

Results of the message classification will be presented first, followed by detailed analysis of episodes.

5.4.1 *Speech act analysis on text messages*

Overall, responders rescued 7 and 9 targets in session A and B respectively, out of 16 targets in total per session. Two players were incapacitated in session A, and 1 player was incapacitated in session B. 117 and 70 messages were sent in session A and B, respectively. We used Searle's classification of speech acts to categorize messages (see table reffig:speachact). We also add requests to the table to categorize all of the messages (Searle does not classify those as speech acts).

Speech acts	Session A		Session B		Example	Total
	HQ	FR	HQ	FR		
Directives	57	0	32	0	<i>JH pair up with BR to save animal in between TA centre and national college.</i>	89 (47%)
Assertives	25	2	8	4	<i>The leak around geospatial is bigger.</i>	39 (20%)
Expressives	5	0	0	0	<i>Good Job, JJ, TV and RL.</i>	5 (2%)
Declarations	3	0	0	0	<i>NOTICE - TEAM B: NS + TD.</i>	3 (1.6%)
Commissives	0	4	0	4	<i>ok got it.</i>	8 (4%)
Requests	8	6	1	19	<i>wheres the leak?</i>	34 (18%)
Unclassified		7		2		9 (5%)

Figure 16: Speech act classification

5.4.1.1 *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. Therefore, we further categories the instructions into two categories: instructions for task allocation, and instructions for 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. Let's take a look at the example below:

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

The instruction sent from Headquarter consists of two parts: (1) Description of Team- ing (who are involved) (2) Description of Location (Targets) to go. It is worth mentioning that HQ players use different strategies when they try to de- scribe a location to field players. HQ players in session D frequently referred to land- mark on the map in their description, while HQ in session C used simple directions (north, west, south east). For example:

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

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 radio- active cloud. To help field players avoid 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"

5.4.1.2 *Assertives*

in this game, assertives provide plain information to recipients. Most assertives are sent by headquarter because they have access to critical information the cloud location. Followings are some 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 maintain situational awareness. We will talk more about this shortly in the section " awareness".

5.4.1.3 Commissives, expressives and declarations

We also identified a small number of commissives, expressives and declarations. Commissives are field player's response 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 and declarations are only found in session C. Expressives are typically HQ's congratulations to field players. (e.g. "HQ:Good Job, JJ, TV and RL") In session C, HQ players sometimes declare field players to be in a team (e.g. "NOTICE - TEAM B: NS + TD"). The declaration helps HQ to refer to a team easier.

5.4.1.4 Requests and Adjacency pairs

In linguistics, adjacency pair is a term to describe conversational turn taking. The pairs can be question->answer, inform -> acknowledge, offer ->acceptance et al. For simplicity, we ignore the typology of adjacency pairs and treat all pairs as re- quest-> response. Any utterance that expects a response is considered as a request. We found a number of requests sent from Headquarters and field players (14 in session A and 20 in session B). Those requests can be related to a number of themes.

Field players may send request for:

1. Task assignment (e.g. “anything for us to do”)
2. Teaming (e.g. “firefighter with me for fuel?”)
3. Information about the cloud (e.g. “wheres the leak?”)

For headquarters:

1. They may ask idle players to respond (e.g. “fighter who is free now?”)
2. They may request for acknowledgement (e.g. “HQ:firefighter, respond”)

In comparison, only a small number of adjacency pairs are found in both sessions (8 in C and 8 in D), which means not all requests are responded. (see figure 17)

Total requests	Requests from HQ /with no response	Requests from field players /with no response
14	8/7	6/1

Requests in Session A

Total requests	Requests from HQ /with no response	Requests from field players /with no response
20	1/0	19/14

Requests in Session B

Figure 17: 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 expect an answer, the headquarter

players express their frustration for not having response to their instructions. A HQ player said in the group interview:

"I guess they did not look at it, they could not respond 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)"

5.4.2 *Responding to directives from HQ*

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 (see fig 16). Directives index (attempted) instances of remote coordination of field responders by HQ. The observed response to messages is critical to understanding relations between local and remote coordination. 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) has just joined soldier KY (on the right), and firefighter D2 who have just been allocated a task in a message by HQ. (see figure 18)



Figure 18: JH (Behind Left), D2 (Middle Front), KY (Right behind)

KY: ((reading out 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 19: 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)) (see figure 19)

KY: Yah this is what the headquarters told us to do
((looking at messages)) (see figure 19)

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] (see figure fig:studylep13)



Figure 20: Caption



Figure 21: Caption

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.] (see figure 21)

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.

Overall, out of the 43 task allocation directives HQ sent, the recipient field responders brought up only 15 messages in conversation in the team. 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 coincided 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. (see figure 25)

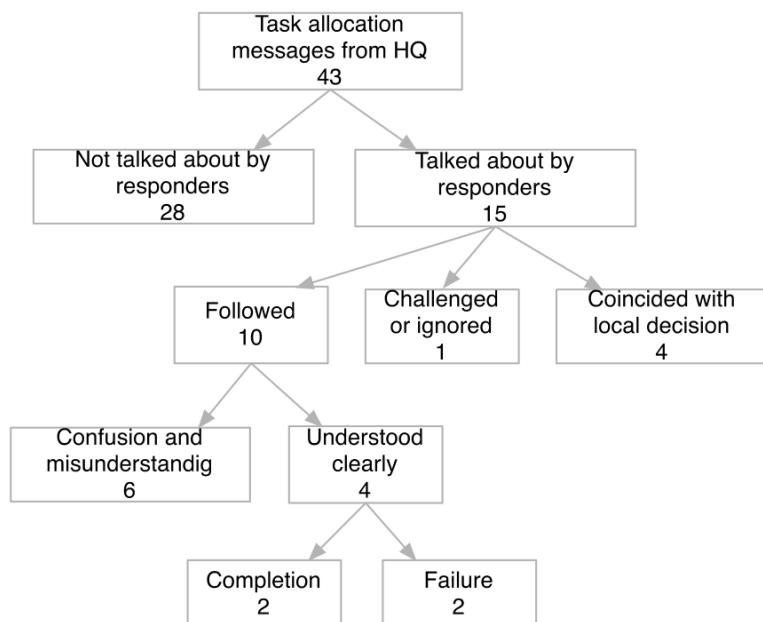


Figure 22: How responders addressed task allocation messages from HQ.

5.4.3 Local coordination without HQ

As presented, field responders predominantly coordinated teaming and task allocation of targets that were saved without HQ involvement. 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 23: Caption



Figure 24: Caption

(on the right) 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. They realise one of them is already being pursued by another team. They agree on another target 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 episode, which demonstrates how two teams resolve the conflict when they approach a same target.



Figure 25: Caption

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 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 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 (see ep x) 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 a same target, MF suggest 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 that how teaming and task allocation are achieved in a “ad-hoc” manner. By using the word “ad hoc”, we stress the the actions of field responders are typically not planned ahead adequately. The players often exchange information through conversations when co-located, and their plans are ready to be changed when new information is acquired. Take episode x 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 have not been aimed by others. In episode y, 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 closet one we can get.”

5.4.4 *Messages as a resource of situational awareness*

In the Radiation Response 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 messages as a resource to gain situational awareness.

The episode takes place towards the end of game session B. The radioac- tive 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.



Figure 26: Caption

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.

MF brings up HQ's message of 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 teammate.



Figure 27: Caption

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]

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)).

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 re-sponse (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 were regularly not reciprocated with a response: out of 14 requests in session A, 8 were not responded upon; and in session B,

14 out of 20 requests were not responded upon.

5.5 PERFORMANCE IMPLICATIONS

This section will present broader concerns emerged from the game for the design of HAC systems that support team coordination.

5.5.1 *Division of labour between HQ and Field Responders*

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 location of 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.(see episode x) 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 x and y, face-to-face conversation is 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 screen. The decisions like 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 remote coordination. 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 com-

munication channel (misunderstanding and confusions), field players always managed to find team-mates and avoid conflicts in their plans (e.g. pursuing the same target). [connect to situated planning by Lucy suchman]. However, the “ad-hoc” local coordination indicates the absence of the notion of resource optimisation, which is often the main concern of multi-agent coordination algorithm. In a sense, it opens opportunity for computational support, but it also highlights some potential challenges for such support system.

Firstly, it would be problematic if globally optimised choices conflict with the “ad-hoc” choices that are obvious for human field responders. As we can not assume either the agent or human choice will be always correct, we need to carefully design the interaction between human and agent to ensure that they can collectively reach an informed decision.

Secondly, there is a danger to impose an inappropriate “work model” (assumed by agent support) on the human team. For example, current division of labour between HQ and field responder suggest that field responders are comfortable with HQ being a supportive role (providing situational awareness). However, a centralized coordination algorithm may need to coordinate the whole team, requiring every players to follow top-down instructions to reach global optimum of resource allocation. In that case, the role of control room may need to change and it is unknown whether the change will disrupt or support natural human workflow.]

5.5.2 *Enhancing remote coordination*

The heavy reliance on the local coordination can partly be a result of the lack of remote coordination support. In other words, local co-

ordination is easier to achieve and can be seen as fallback of remote coordination. In a co-located setting, players can naturally make their actions observable and accountable to each other through conversations, body languages, gestures et al.[need to connect to data] and organise coordination activities reflexively. However, in the remote setting, the natural accountability of their activities become opaque.

A set of functionalities including GPS/map sharing, broadcasting has been provided. We have observed players utilise the functionalities to make sense of other team members' actions (see section 5.4.3) and act accordingly. However, coordination with remote players is still overall problematic which can be evidenced by frustrations related to remote communication (will be detailed in section 5.6).

We suggest the future planning support should properly support remote coordination in a way that facilitate accountability among distributed team members. The next section discuss some detailed requirements of remote coordination support drawn from the field observations.

5.6 DESIGN REQUIREMENTS

We now discuss the requirements for team coordination that emerged from the game. 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.

Local decision-making The study showed that teaming and task allocation were predominantly organised locally among field responders, in an “ad-hoc” manner, on-the-fly fashion. Despite the fact that headquarters 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. Local decision-making needs to be supported by HAC systems that aim to support time-critical team coordination, and need 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 HAC 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 landmark 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 Toups et al. (2009), designers need to think carefully about how the presentation layer of HAC 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. Problems arose from erroneous instructions or otherwise out-dated messages sent to field responders. In one case HQ sent a message in which two players with non-compatible roles were instructed to team up. This was particularly costly, as the players attempted to team up, and lost valuable time until they realised the game mechanics barred them from forming a team.

Freshness of messages As demonstrated in one of the episodes, 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, HAC systems should address these issues at the UI level, both for responders and for HQ. Develop functionality to flag messages as out-dated or retract incorrect messages. Up-to-date messages are particularly valuable. Thus, our findings support the use of fresh social media as a source of information for disaster response, despite problems that can arise with validation, because crowdsourced information will in many cases provide better coverage than official sources.

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 HAC 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.

5.7 SUMMARY

The objective of the research presented here was to generate requirements for supporting time-critical team coordination. 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.

We presented the design and study of the Radiation Response 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 eschew high-fidelity simulations in favour of 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 generated requirements for supporting team coordination, emphasising the roles of local coordination, decision-making resources, geospatial referencing and message handling. These requirements inform future work on building human-centred HAC systems by emphasising the role of human interaction in team coordination in time-critical settings.

ATOMICORCHID STUDY 2: AGENT SUPPORTED HUOL 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 HuOL interaction design, in which the HQ only monitor and occasionally intervene the planning agent. The purpose of the trials is to investigate socio-technical issues relating to integrating the planning agent to rescue team with the HuOL interaction design. Interaction analysis is conducted to examine log data and field observations unfolding how human agent interaction plays out, which in turn, revealing the process by which players interpret and negotiate the agent guidance as well as how these are intertwined with social dynamics of the teams.

6.1 INTRODUCTION

Task planning in teams can be complicated by both spatial and temporal constraints, particularly in time-critical task domains such as disaster response (DR). In a DR setting, responder teams have to coordinate sparse resources and personnel to prioritize geographically distributed tasks, forming and disbanding teams dynamically to carry out DR operations [4]. Multi-agent researcher has devised a number of agent coordination algorithm to coordinate task allocations for multi-agent systems, which create opportunity for computational planning support for DR teams.

However, these algorithms typically model humans as computational agents with respective capabilities, for example to dynamically

allocate teams of agents to tasks in order to maximise an objective (e.g., number of lives saved), taking into account other aspects of the real world (environment, infrastructures, victims, etc.) [14]. Therefore, the quality of the planning results can be constrained by limited assumptions of human behaviour (e.g., human psychosocial characteristics, movement, and learning ability) and real world environment [18]. These limitations highlight importance of human input in the planning process. Thus, we argue that effective collaboration between human and agent is required to produce and execute high quality plans in the disaster setting. (refer to background section?).

In order to support effective Human agent collaboration, a number of interaction design strategies have been identified, including design patterns such as HuOL and HuIL (see section ??). In this study, the planning agent is integrated into the AtomicOrchid system with a straightforward HuOL design pattern. This interaction design pattern assumes minimal Headquarter intervention and simple feedbacks from field responders are required by the agent to perform task planning. The interaction design is aimed 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. 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 iteration of task assignment.

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 HuOL design for human team performance. The implications for the interaction design are discussed that may improve 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; the hidden costs of instructions that suggest team reformation and interrupt on-going tasks. We conclude the paper with a number of emerging interaction design recommendations to consider when building agent-based support systems for human teams, which emphasise 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.

6.2 SYSTEM EVOLUTION

Compared to study 1, the system has evolved to provide agent planning support through HuOL interaction pattern. This section will go through the changes of system, which includes integration of a planning agent, implementation of a quick feedback system, and improvement in both HQ and mobile interface.

6.2.1 *The planning agent*

The 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, Savapali Ramchun. The coordination problem (described in section ??) of the AtomicOrchid is modelled using a Multi-Agent Markov Decision Process (MMDP) that captures the uncertainties of task execution, extending earlier work [15]. The modelling 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-sized problem. Hence, we devised an approximate solution 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 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 done by Wu Feng to benchmark our MMDP algorithm against greedy and myopic methods (see Table 1). The results confirm that our algorithm produces efficient task allocations.

The agent is wrapped in a Java server framework and deployed on an independent server separate from AtomicOrchid. The agent 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 ad-

justed according to task execution status. 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, will be introduced in next section.

6.2.2 *A feedback loop*

The feedback system is part of the HuOL interaction design, which enables the agent to take into account simple human feedback. It is also partly inspired by a requirement generated in the study 1, which highlights the importance quick acknowledgement from field responders. The feedback system can be seen as system level support for the field team to provide quick response to the HQ. This section will go through 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. The app displays the task allocation in a pop-up and details it in the 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).

On receiving a instruction from agent, the field responder can choose to either reject or accept the instruction. In the case of rejection, 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 rejected it, the planning agent would return a new task allocation with the constraint that this soldier should not be allocated this task.

6.2.3 *Interface improvement*

Apart from the integration and feedback system, two small modification of the interface is inspired by the requirements of coordination system generated in the previous section (See section 5.2). 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 identify outdated messages.

6.3 STUDY DESIGN

Just introduce the procedure.

6.4 DATA ANALYSIS AND RESULTS

This section presents episodes 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. In the following fragments, players can be uniquely identified by their initials. Targets are denoted by their unique numeric target id. Task assignments from the agent 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. A standard orthographic notation [8] is complemented by timestamps [0:00], and system messages from remote players and HQ.

6.4.1 *Task assignment*

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

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 (PC: I think we dropped off now. OK). After about 7 seconds, PC says she received a new task allocation from the agent (PC: 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 (CR: 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.

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

[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?

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.

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

6.4.2 Team reformation

Unlike episode A, 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. 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 28: Caption

[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.

[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

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 (LT: 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 (Fig. 3). 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 (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 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 struggles to leave. She does not leave the group to follow the instructions until prompted by someone. When CR points out NW message, 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 B 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.

6.4.3 *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 C and D describe how players handled task interruptions caused by the agent.



Figure 29: Caption



Figure 30: Caption

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

AW, 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.

[Fig. 5] 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 look-

ing 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 (AW: I got a new instruction), followed by both teammates simultaneously turning towards each other (Fig. 5). 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 (HB: I need to team up with AR, AW: 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 too far from the drop off zone; so everyone was still within line of sight, further facilitating successful reformation.

6.4.4 Disagreement on task interruption

[Following on from Episode C]

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 C, player AW disbards his team with HB and teams up with KD. However, 20 seconds after the ref-

ormation, 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 KDs 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 (KD: Alright, Lets go to 46), indicating he does not agree with AWs 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 assignments. 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 C 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 coalition formation 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.

6.4.5 *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. The majority of assertives were focused on providing situational awareness and safe 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 then 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 instruction, 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 take control.

6.5 DISCUSSION

In the previous sections, we described how the agent guidance is interleaved with the 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 (A) how division of labour is achieved; (B) the social implications and hidden cost incurred by team reformation and task interruption; and (C) the limited feedback mechanism.

6.5.1 *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, which indicates that, to a large extent, the agent successfully supported routine task planning activities. Episode A demonstrates a typical case of division of labour: the agent handles planning of teaming and task assignment, freeing the team to focus on navigational issues (identifying the target on the interactive map and finding directions). 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 labour division.

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 safety routing the responders to avoid radiation exposure. However, the fact that only 5 (out of 16) HQ instructions are implemented sug-

gests 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. This has implications for interaction design that are discussed in the next section.

6.5.2 *Hidden costs of team reformation and task interruption*

While team 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) (see Table 2). 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.

Firstly, 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 (section V.B), 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 (V.C), while instructions for only one member have coincided with interactional trouble, disagreement and delays (V.D).

Secondly, 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 D, 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 of accountability between remote members.

A further observation is that players were unwilling to give up on-going tasks after a certain time. In episode D, 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 simplistic model of interaction held by the agent. The agents coalition formation re-plans and reshuffles teams, in order to optimise group performance by minimising the travel distance to the targets. However, our study has revealed the ways in which social norms and the accountability of social conduct get in the way. This raises questions of 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. Our field 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 [1]. 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).

6.5.3 Feedback to the agent

To recap, a feedback mechanism is included in the interaction design to give responders some control over the task assignment. On receiving an instruction, players can either accept or reject instructions. 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 planner agent, which means the rejected target will not be assigned to the rejecting player for a while (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. 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). 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 are more likely to be rejected. Players statements indicate they perceive the rejection as a way to reverse to previous states (see ep. D). Other statements indicate rejections were expected to pair them with a new teammate instead of a new target. The mismatch between expected and actual effect highlights the 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.

6.6 DESIGN IMPLICATIONS

Our observations reveal the tension between agent planning support and the social organisation of teamwork. The tension does not simply mean the model held by the agent is incorrect; it highlights potential trade-offs we need to consider in system design [1]. Providing a detailed design solution is beyond the scope of this paper. Instead, we propose three design implications to scaffold the division of labour when building agent-based planning support for human teams.

6.6.1 *Achieve common ground*

Two main issues arose that challenged this basis for collaboration [2,19]. 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 received a new instruction. Secondly, a mismatch between the expected and actual function of rejections further shows intelligibility needs to be improved. Therefore, we suggest the design of agent support that a) takes 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).

6.6.2 *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 shall reveal the hidden cost of certain actions (e.g., rejections) to facilitate making local decisions accountable to remote team members, ensuring consequences of local decisions for the welfare of all teams are understood.

6.6.3 *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 [2]. The ways in which the HQ used messsages to provide situational information that complemented the agent instructions show that humans are readily able to deal with arising situational contingencies. The division of labour between humans and the agent appeared most 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 our interactional arrangement, the role of the human HQ was relatively weak. For example, the HQ struggled to overwrite the agents instructions through the messaging channel. In the future, we seek 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).

6.7 SUMMARY

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

This chapter presents the third iteration of AtomicOrchid field trials. Based on the second prototype of AtomicOrchid (see section 6.2), the system have gone through another development iteration to facilitate a In-the-loop interaction pattern. The In-the-loop pattern assume that constant human agent interaction is required for planning. The purpose of the trials is to investigate socio-technical issues relating to agent planning support with the In-the-loop interaction design. Interaction analysis is conducted to examine log data and field observations, unfolding how human agent interaction plays out. The results of the analysis reveals the process by which HQ, field players and the agents collectively generate and executes plans, as well as socio-technical issues emerged from this process.

7.1 INTRODUCTION

Most disaster operations require responder teams to plan and carry out task under spatial and time constraints. which means the teams often have limited resource and personnel to deal with large amount of geopolitically distributed tasks in limited amount of time. How do they optimise the use their rescue resources become computationally complicated problem. Multi-agent system researchers have developed multi-agent task allocation algorithms. As software components, they have all done very well in the computational simulation, therefore there is potential to apply those algorithms to support planning

activity of human responder teams.

However, there is concern that the algorithms hold over-simplified model of the environment and human behaviours. Therefore, effective interactions with the responder teams may be required to ensure the planning agent actually support (rather than hinder) the planning process of responder teams, which highlights the importance of appropriate interaction design. The In-the-loop design pattern assumes frequent HQ supervision and intervention is required to ensure the agent works properly (section 3.4). In this study, the AtomicOrchid system has evolved from the zed game probe in the study 2 (chapter 6), 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 and approve every instructions generated by agent.
2. allowing HQ to reject and override instructions generated by agent.
3. allowing HQ to cancel the plan in execution.

Compared to the On-the-loop design in the previous study (section 3.4), the major change is the rebalance 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 strong focus on control room interaction compared to the study 2. More specifically, this chapter addresses the following research questions on how a In-the-loop interaction design affects the team workflow: [preliminary]

1. How does the division of labour between human and agent play out in the In-the-loop paradigm, particularly in the con-

trol room.

2. What are the design implications that help the HQ involve in the planning loop.

Findings from the study highlight the processes in which the agent and HQ players collectively generate task assignments for field players. We discuss the division of labour between agent and the human team; the impact of HQ's involvement in the planning; the lack of information sharing between HQ and agent; and the issues about Field - Control room communications. The chapter is then concluded with the reflection on some of design implications, which emphasise the need for common ground between humans and the agent, and balance responsibilities between humans and the planning agent appropriately.

7.2 SYSTEM EVOLUTION

The In-the-loop version of AtomicOrchid is not designed from scratch, but evolved from the On-the-loop version introduced in previous study. In the study 2, we have observed HQ struggling to get involved in the planning loop. Combining the observations from the study 2 and the principals of In-the-loop interaction pattern (3.4), we further generate several system requirements for realizing the In-the-loop interaction.

1. HQ should be able to review and approve every instructions generated by agent.[HQ take to responsibility for final decision, can be kept aware of all the allocations sent]

2. HQ should be able to override instructions generated by agent and create new instructions.[HQ can send instructions in a structured way]
3. HQ should be able to cancel the plan in execution. [In study 2, HQ is unable to override the existing plan]
4. HQ should be able to send task specific information. [Dividing communication channel to deliver info accurately]
5. HQ should be able to be aware of the feedback (accept/reject) from the field players. [Maintain situational awareness of HQ]
6. HQ should be able to decide when the agent should re-plan.
7. HQ should be able plan for part of the team, leaving the agent to plan for the team left.

While requirements 1-6 are inspired by implications from study 2, the requirement 7 enable HQ to intervene in the agent planning without having to take full control from the agent. A set of interface functionalities have been implemented to support above requirements, which will be introduced in next section.

7.2.1 *Interface functionalities*

A set of fully functional interactive web-based and mobile interfaces have been developed to facilitate the In-the-loop design. There are two interfaces designed for the 2 HQ players in the control room. One mobile responder interface is designed for the field players.

The task assignment interface is operated by one of the two HQ players. As an overview, the interface has a map on the left [Fig x,]. Player/target locations, assignments are presented on the map. At

right side of the interface is a task assignment panel. The left (1) column of the panel shows pending assignments while right column (2) shows existing task status. The (x) shows an example of individual assignment. (Player a and b are assigned to target x)

1. Plan request button:

This button triggers agent re-planning. The agent will calculate an optimized plan based on task status, and present it to the HQ on the pending panel. The button allow HQ decide when to initiate a re-plan (meeting requirement 6)

2. Plan keeping checkbox:

These checkboxes are attached to every task assignments in the confirmed panel. If the checkbox is ticked, the planner will keep the corresponding assignment in next re-planning. In other words, the planner will keep the assignment fixed, performing partial planning for the rest of the team. The partial planning functionality is implemented to meet requirement 7.

3. Plan edit panel:

Manual plan edits will be activated by clicking this button. The assignments in pending area will change to edit mode. Assignment can be created, modified and deleted through drag and drop interaction. The functionality is implemented to meet requirement 2

4. Plan approval button:

This button approves all pending assignments. All pending assignment will switch to confirmed area. Alternatively, assignments can be approved individually by clicking individual confirm button on the pending assignment when the edit mode is activated. The functionality is implemented to meet require-

ment 1

5. Text messaging panel:

The messaging panel can be toggled by clicking msg button on the confirmed assignments. The panel is supposed to be used for assignment-specific information. Therefore, the messages in this plan are only visible to the two involved players and HQ. The functionality is implemented to meet requirement 4

6. The Feedback indicator:

The feedback indicators are attached to the right hand side of the confirmed assignments. The field players can easily provide feedbacks of their assignment through 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 with red color to grab attention of field players. The functionality is implemented to meet requirement 5

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

7. Stop button:

The stop button can be used to indicate an emergency termination of an assignment. If this button is clicked, the assignment will be dismissed both in the mobile and HQ interface. The functionality is implemented to meet requirement 3

Compared to the HuOL version of AtomicOrchid, the mobile interface is mostly unchanged except for the HQ task tab (see figure x). The task tab now displays a task with text description and map visualisation on 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.

7.2.2 *The planning agent*

One big change of the planner (compared to iteration 2) is the partial planning feature. The agent can take a list of fixed assignment as input. It then allows optimize the players and targets that are not involved in the fixed assignment list. The functionality has two potential usages:

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

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

7.3 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. 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 are recruited from researchers in the computer science department.

Because the HQ interface is a lot complicated compare to the 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 anticipate the workload of operating the HuIL interface would be a lot more then that of operating the HuOL interface. Therefore, there are two HQ payers recruited in each session to split work in the control room. One of the two HQ player operates the new Hull interface (described in section x.x.x) to manage task allocation, while the other player operates the old HuOL interface described in study 2 (see section x.x.x) to assist the other HQ 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 (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.

Before the formal session begins, there is 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 start-

ing point. The training session ends when field responders collect all four targets nearby. Once field responders were ready to start formal session, One research start the game engine, triggering a “game start” message to be sent to mobile interface. 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 post-game session consisted of a questionnaire aimed at collecting participants’ feedback on (1) first impressions of the game; (2) usability of the system, and; (3) coordination issues in the game. A group interview was then conducted, before participants were debriefed and dismissed.

The size of the game area on the local university campus was 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 20 targets. The pilot study showed that this was a challenging, yet not too 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.

We took a mixed methods approach to data collection and analysis. In addition to quantitative questionnaires, a semi-structured group interview was conducted aimed at eliciting important decision points, strategies and the overall decision-making process. Furthermore, 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.

We developed a log file replay tool to help with data analysis of time stamped system logs that contain a complete record of the game play, including responders’ GPS location, their health status and ra-

dioactive exposure, messages, cloud location, locations of target objects and task status.

We focus on the analysis of players' interaction with the system to unpack interactional issues. As this study have strong focus on the interaction in the control room, the interaction analysis started with observations in the control room. Video recordings of control room field action were catalogued to identify sequences (episodes) of interest (cf. Heath et al., 2010). 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 and field actions for deeper analysis that we present in this paper.

7.4 DATA ANALYSIS

This section will start with overview of game results, messaging system usage and task assignments. Detailed episodes will also be presented unpack interactions between participants and the system.

Overall the 28 of 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
16	99	75	90.75	9.337	0
12	97	0	48.125	41	2

Figure 31: Result Overview

7.4.1 Messaging system

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

Session 1	Sent by HQs	Sent by FR	(Total)
General message	35	3	38
Assignment-specific	22	15	37
Total	57	18	75
Session 2	Sent by HQ	Sent by FR	(Total)
General message	31	13	44
Assignment-specific	19	11	30
Total	50	24	74

Figure 32: Task assignment overview

We found HQs frequently send messages to update location of 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 wy 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 individual teams. The specific player initials are mentioned in those

messages. (E.g. “NG and YI approach quicker to 41 drop off to 8”)

7.4.2 Overview of task assignments

In the following tree diagrams, Plans are broken into individual task assignment. 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.

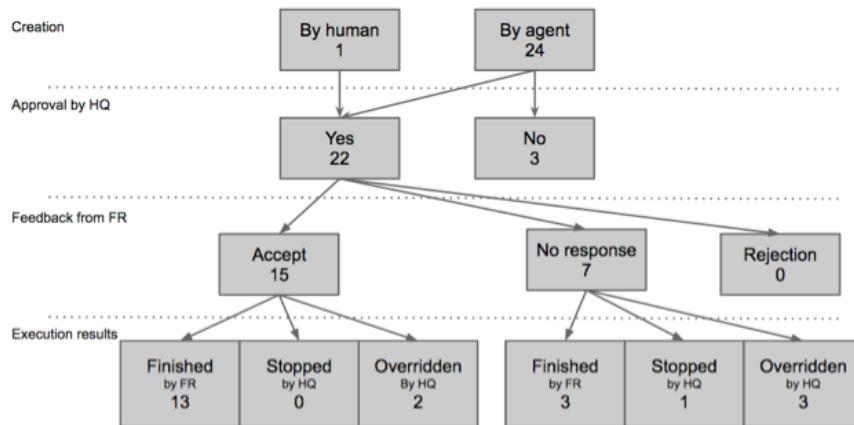


Figure 33: Task assignment in session 1

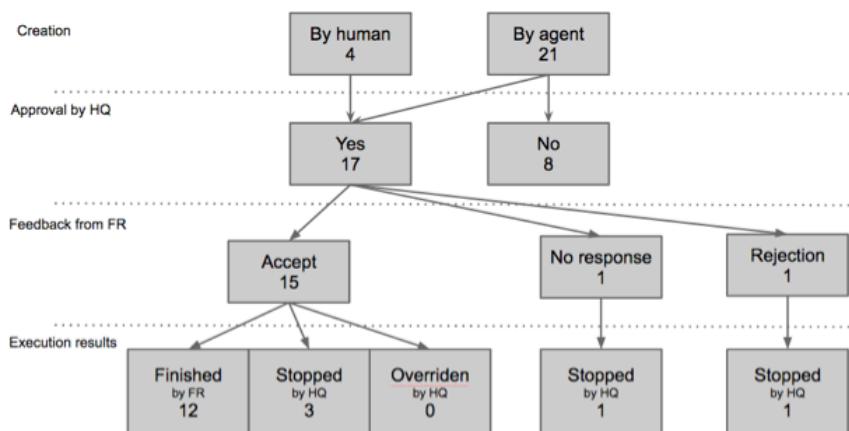


Figure 34: Task assignment in session 2

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

Although the diagrams (figure 33, 34) show 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.

In the following fragments and episodes, players will be uniquely identified by their initials. Targets are denoted by their unique numeric target id. Task assignments from the agent 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. A standard orthographic notation is complemented by timestamps [0:00], and System Message from remote players and HQ. The HQ players will be denoted by HQ1 and HQ2. HQ1 refers to the player operating the task-assignment interface, while HQ2 refers to the player operating the situational-awareness interface.

7.4.3 Approve the plans generated by planner

Firstly, HQ players in the control room are observed to carry out a set of routine activities, collaborating with the planning agent to organize task assignment. The episode 1 demonstrates a typical case of routine task planning process in control room.

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 [Fig. 1]

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? [Fig. 2]

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, (HQ click re-plan)

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

18.28, HQ1 36, yes (Click confirm)

At the beginning of this episode, the team MV, XW was carrying a target (43), approaching drop off zone. At [16:45], the HQ2 noticed the team was going to be freed up soon, so he began considering

next target for them [16:51]. HQ2 proposed to HQ that 36 should be prioritized [Fig. 1]. HQ1 agreed with the suggestion and decided to send the assignment after the team free up from existing task [Fig. 2][17:07]. At 18:10, the team dropped off target. After the drop-off, the HQ requested a re-plan. The agent assigned 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, we found HQ closely monitored the task execution, and request new task assignments when field team freed up. HQ's discussion suggested that the agent task assignment was approved after a careful review of player, target and radiation status. In addition, the HQ's planning happen before the team dropped off the target. It can be seen as a forward planning activity, which we observed several times in this observation.

7.4.4 *Editing assignments generated by planner*

Apart from routine task monitoring, HQ players occasionally chose to intervene the task planning activities by utilizing functionalities provided by the task assignment interface. The following simple episode demonstrates how HQ manually priorities target in order to implement their planning strategy.

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) [Fig 3]

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

01:33, (HQ1 click edit)

01:53, HQ1: I want this one (HQ1 drag target x to replace agent planning, see Fig.4) 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) [Fig 4]

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

This episode happened at the beginning of the session 2. It 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 closet to the radiation cloud [fig 3]. 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 finished 1 assignment is cancelled later, 1 assignment leads to player death.

7.4.5 Cancelling approved assignments

HQ's interventions are not limited to modification of agent proposed assignment. HQ also manually cancels (stop) approved assignments 5 times due to changing task status. The next episode exemplifies how HQ adapts task assignments to changing task status.

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

02:09, HQ2: they cannot walk there straight. [Fig 5]

Referring to team MB, G0, 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)

After confirming assignments in Episode 1, the HQ2 point out one assignment may be impractical because the route to target has already been blocked by a radiation cloud [02:09] [Fig. 5]. HQ immediately

open message panel to send warnings and urging the team to move fast [Fig. 6]. However the cloud expansion seems to be faster than HQ originally expected [02:51]. Apart from sending route guidance in the team, HQ started to consider cancellation of the assignment. After HQ2 agree with the cancellation, HQ send a message to the team to inform them the assignment is going to be cancelled and instruct them to go to new target o7. The assignment is formally cancelled in [04:11]. After HQ1 cancelled the task, he started to request new assignments from planner and allocated target o7 to the team later.

7.4.6 HQ avoiding task interruption

As it is shown in the previous episodes (3), the existing tasks may be interrupted by HQ interventions. Apart from HQ interventions, the task interruption can also be caused by the re-planing performed by the planning agent. In one case, we observed HQ picked up this kind of interruption proposed by the agent and veto it in the approval stage. This next episode illustrates the case in more details.

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. [Fig 7]

(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 requested a re-plan, agent proposed a set of assignments, one of which (CE KH -> o6) interrupts existing task of a team (CE KH -> o3). HQ immediately picked up this interruption ("What? Why I am getting?") and figured out the reason [Fig 7] ("Ahh, one of these guys does not accepted"). It should be noted that this kind of task interruption only happens when field players did not accept the tasks (See interaction design section x), making agent think they are idle at that moment. After finding this problem, HQ then requires the agent to "keep" existing assignment [04:29] and performs a plan request again to get desired assignments.

7.4.7 Consequence of task interruption

The next episode will show a problematic case of task interruption caused by agent re-planing . This episode involves 3 field players NG, YI and DI. The NG, YI formed a collocated team, and they just finished their previous task. DI is a player remote to NG and YI, alone at a drop off zone waiting for new task. This episode is divided into 2 parts. Part 1 is about observation in control room and part 2 is about team reaction in the field.

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 request 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 [fig], 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. HQ1 quickly approves this assignment (within x seconds). Although HQ1 did not verbalize his reasoning process, the quick approval may indicate HQ1 has not inspected new assignment carefully and did not notice the approval will cause task interruption.

Although the episode does not seem to be problematic from the HQ's perspective, we actually observed problematic sense making process in the field. The part 2 of this episode (following) reveals how field responders react to the task interruptions.

Context: NG YI received a task update after they finished their previous task.

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

23:50, NG: Task changed, to what?

24:06, NG: Oh sh*t, there is a radiation zone sh*t.

24:15, NG: It (target 50) is close to the lake. It is triumph road we have to go that way.

24:28, YI: Yea, we need to. No this is the lake.

24:33, NG: Oh that is the river, so it is that way, Can we go? Yes we can go, this road?

24:50, YI: Yes it should be.

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

24:52, YI: Task changed.

24:53, NG: Task changed? What?

24:57, YI: No, it is the same.

25:00, NG: I do not see anything.

25:02, YI: You are DI, right?

25:03, NG: No.

25:04, YI: En? You are no DI? DI, wait.

25:11, NG: I can not see my Task.

25:19, YI: You are NG, Oh no, I need DI.

25:21, NG: You should go back then. Or send a message, Oh I can not see my task.

25:34, NG: I think the page is not loading.

This part begins with a new task assignment for team NG, YI (NG, YI -> 50). This assignment is followed by a discussion about the location of the new target. At [24:50], the team figured out the route to pursue the target. However, a new task assignment arrived, requiring 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 thought the task did not change because the target is still 50 [24:57, YI: No, it is the same.],

and NG thought the interface is no longer working [25:34, NG: I think the page is not loading.]. After YI confirms NG's initials, she realized she needs to switch teammate (to DI).

Overall, we found the task interrupt resulted in a problematic sense making process in the field. The control room observation also reveals that several factors could have led to the task interruption including the absent 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. This episode can be an example showing that computationally 'optimized' assignment is not socially optimized for teams and HQ supervision failed to tackle the issue.

7.4.8 *The consequence of missing player feedbacks*

In session 7.4.7, we have already demonstrated the consequence of missing player feedbacks from field players. Lack of feedback leads to false assumptions of the agent (players are still available), which in turn compromises its optimization. In one case, we also found that HQ may also be distracted by the lack of response, in that they have to guess player's intention without their feedback.

(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]

7.4.9 *The incident of player death*

Two players are dead in the session 2 , the following episode will report this particular incident. This episode can be seen as a follow up of episode 2, in which the HQ modified 3 agent assignments. In that plan, the team CE, KM is instructed to go for a target very close to a radiation cloud. The episode is also divided into two parts. The part 1 illustrates what happened in control room while part 2 shows the field observation.

(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)

4:56, (HQ messaging) Careful, you are approaching the cloud. Move extremely fast to the cloud. Go around the lake to return.

5:32, HQ: God, too slow.

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

5.46, (HQ 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, (HQ messaging) Abort 03, return top to (target) 99.

9.04, (HQ1 stopped task of CE KM)

9.05-10:08, (HQ 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 has already got in the way between target and the field team CE and KM. The cloud was expanding quickly, which caught the HQ1 in surprise. As a result, 2 messages were sent by HQ1 to guide the team to avoid the cloud [4:56,5:46]. However the team was still excessively exposed to the radiation when they reached the side of the target, so HQ1 decide to abort the task. After HQ1 cancelled the assignment, the team

was still exposed to radiation. HQ1 then tried to assign the players to other targets by requesting agent for new assignment. 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 continuously requested plans for three times and said "No targets for them? There are lots of them (refer to targets)."[9:50] During this process, the team was still standing in the cloud. They finally lost all the health points and become incapacitated. The second part of the episode showed what happened in the

05:06, CE: (Reading out the message) You are approaching the cloud, move extremely fast to the cloud, Ok!

05:12, (CE grabbed KM and started running) [Fig 8]

06:03, (The team CW, KM ran all the way across the cloud)

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

06:53, CE: how dead are you?

06:55, KM: Pretty much dead.

06:56, CE: I am pretty dead as well.

(*The team is trying to locate the target, but it is unsuccessful*) [Fig, 9]

07:19, CE: Do you know what is funny? We have to get back.

08:15, CE: It should be somewhere around here! We are getting close.

(*Assignment cancelled*)

08:43, CE: there is a new task, No task at the moment?

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

09:35, CE: I am pretty much dead and radiation is 12.

Where is it! It says it is on this street but it is not.

09:57, CE: Oh I am dead.

The part 2 began with player CE reading the HQ message ("You are approaching the cloud, move extremely fast"). After this message, CE grabbed teammate KM and starts to run through the cloud [Fig. 8]. After they ran through the cloud, the team checked their remaining health value [07:19] and started searching for the target at the edge of the radiation cloud. About half minute later, the assignment was cancelled by HQ [08:43]. Although the task has been cancelled, the HQ failed to assign new task to the team [part 1, 9:50]. The team simply ignored the cancellation and kept searching for the target until they loss all their health [09:57].

7.4.10 *Division of labour in HQ*

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

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 public message channel.

In session 2, we do not find similar negotiations. However, in the post game discussion, the HQ2 expressed that he has too little to

things to do (other then sending general messages), while HQ1 has too much to do:

“The interface I use was completely useless, became 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, Khaled (HQ1) had to message to everybody privately. While he was doing that, he coundn’t do new plans. So there is a lot stuff for him be I coundn’t do anything.”

7.5 DISCUSSION

In previous section, episodes are presented to illustrate how the In-the-loop interaction design plays out in the field. The episodes reveal that, to large extent, human the HQ players were successfully involved in the control loop. However, we also found some confusions and misunderstandings occur between agent and human which may leads to issues and implications for interaction designs.

7.5.1 *How does In-the-loop design play out?*

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

1. **Plan request and approval** The plan request function can be seen as a tool for HQ to trigger the agent re-planning. The design of the plan approval stage gives HQ an opportunity moment to review and influence the final plans before it is sent to field responders. Both functions are essential for human and agent to collectively produce task assignment, constituting the routine task planning work in control room. As presented in previous section, the two functions are the two most frequently used interface functionalities. [HQ requested plans for 45 times; approved 39 task assignments] Deciding appropriate moment to request plan and approve desired assignments require 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 HQs are engaged in the planning-execution loop with a supervisory role.

2. **Plan edits** The “plan edits” enables HQs to directly intervene the planning. We observed HQ’s modified undesirable agent plans twice throughout the field study [section x and y]. 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 aspects planning activities.

3. **Plan cancellation** Further, the task cancellation functionality allows HQ to influence task execution after the plan approval. In episode 3, we observed assignment being cancelled and teams being reassigned due to unexpected cloud activities. This may suggests that 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 HQ's to indirectly influence on the planner. For example in episode 4, HQ identified task interruption in 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 assumption of the planner agent.

The usage of individual interface functionalities indicates the interface functionalities may actually serve its design purpose, which is to support the HQ requesting, editing, approving and cancelling the plans. In what follows, we will examine how the division of labour play out as a result of the interface support. the observation reveals that HQ players contribute to the planning with a set of supervisory activities, while the agent take over computational optimization of task allocation.

- 1) Firstly, HQ players are observed to closely monitor the plan execution. For example in Ep 1, field players' task progress is mentioned by the HQs. In episode 3, HQs expressed their concerns about cloud and player locations, while in episode 6, HQ's are more concerned with the uncertainty of field players' intention. The task
- 2) Occasionally, HQs also need to modify plans when agent proposed plan is not desirable. For instance in Ep 2, HQ overridden 3 our of 4 agent proposed assignments to implement his own strategy. Further, HQ's intervention is also required when contingency rises from task execution. For example, The HQ terminated an approved assignment due to unexpected cloud activities. A re-plan is quickly followed up to assign new assignments to the affected teams.

3) The forward planning activities are occasionally observed. For example, HQ's discussion of future plans for a team is observed in Episode 1. The forward planning activity is thought to be not well supported by the system and it is unclear whether this activity contributes to the planning. More details will be discussed in section x. On the other hand, the agent is found to take over 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 for 49 times. The agent generated 45 task assignments, 34 out of which are approved by Headquarters. In comparison with the agent planner, Human only created 5 task assignments. The fact that agent creates large proportion of assignments suggests the agent successfully takes over the routine planning to a large extent.

A pattern of labour division is also observed between the two HQ players in control room. To recap, the two HQ players in control room are split into two roles. The role of HQ1 is responsible for handle task assignment and send assignment-specific messages to field responders. [Have we fixed the role in briefing] The role of HQ2 is to support monitoring task status and send general chat messages. As a result, the messaging system of the messaging is split in to two channels (general information and assignment-specific, see section x). The messaging interface allows each HQ player to take control of one of the messaging channel. However, the system does not play out as designed in the field study. 11 messages in the general message channel are task specific, which indicate the HQ players may occasionally violate the designed division of responsibility. The field observation further confirms that the responsibilities to send messages are dynamically negotiated between the two HQ players regardless of the designed channel division [see fragment 1]. Further in the interview section, HQ players commented that one HQ is always doing too much while the other do not have anything to do [see section x]. The comments suggest there may exist imbalance of workload between

the two HQ players. The observation is consistent with findings of the CSCW researches. Some CSCW researchers argued that the norms for using a CSCW system are often actively negotiated among users. These norms of use are also subject to renegotiation (Strauss, 1991). CSCW systems should have some secondary mechanism or communication back channel to allow users to negotiate the norms of use, exceptions, and breakdowns among themselves, making the system more flexible.

7.5.2 Dealing with the imperfection of agent planner

In the game probe, the planning can be seen as partially automated by the planner agent. The task optimization performed by agent is thought 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 exemplified in previous chapter (study iteration 2), and also Ep 5. In more details, In that episode, HQs failed to pick up unnecessary task interruption in the approval stage. Consequently, extra coordination overhead and problematic sense making process for field responders are observed. This case of failure reveals several factors that contribute to the failure:

1. Lack of feedback,
2. Failures of HQ monitoring,
3. Computational optimization performed by agent, which ignores overhead of task interruption.

The point 3 can be seen as the main cause of imperfection of the agent. In this study, imperfection is thought to be “natural” for agent planning support, as computationally “optimized” assignments from the planner may not socially optimized for the responder teams [ref to study 2]. It highlights the importance of human involvement in the planning activities.

The failure case in this episode is similar to the well-observed phenomena related to imperfection of automation called complacency. The complacency phenomenon refers to the human failure of detecting occasional automation failure. Study of automation suggests that complacency can be classified as attention-based monitoring failure, which is likely to happen when human is engaged in multi-thread tasking. In the case of episode 5, the quick approval of problematic assignments suggests 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 only focus on the new assignments for the idle players rather than other 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 failed in episode 5, the interface highlight does grabbed attention of the HQ in episode 4, which suggests that the interaction design that properly drive the HQ’s attention can help to reduce complacency effects. [Connect to more literatures] Some study shows that complacency is found in both naive and expert participants and cannot be overcome with simple practice. Therefore, mechanism to counter the complacency effect may be an important factor that need to be considered in the interaction design. [can we claim?]

7.5.3 *Lack of explanation from agent*

Some observations in the study showed that some agent actions are not sufficiently explained to HQ, which results in undesirable HQ performance. This section will exam two particular cases in details.

Firstly, the HQ sometimes become confused when the planner stop giving back assignments. As presented in episode 7, the HQ repeated requesting plans four times and complains about no response from planner. However, the planner does not produce more assignments simply because the idle players have little health to take any more assignments. This misunderstanding indicates there is lack of explanation for the HQ to make sense of agent's reasoning process.

In another one case, we observed HQ drastically modified agent proposed plan (Ep 2). In episode 2, the HQ modified 3 out of 4 the agent proposed assignments. As the new targets are closer to cloud, the HQ's plan could be more risky. The risky plan turned out to be undesirable as it leads to 2 assignment cancellations and 2 player deaths (episode 7). As the planner has ruled out the risky plan at the first place, the reasoning behind the original plan may be important information that should have been exposed to HQ to inform their decisions. Depending on the implementation of the planner, evaluation of human proposed alternatives (or comparison with preferred plan) may also be available during the agent planning and can be exposed to imply risk. If HQ is aware of why the agent chose the original plan, HQ may be able to make informed decision when overriding the agent plans. However, the extra information sharing may further complicate the interface. For example, for the agent to share information of its reasoning behind a particular task assignment, it may need to present its prediction of the future game status and its assessment of current game status. It would be problematic for HQ to digest the

extra information in the time-constrained task environment. Various studies have identified the problem of information overload in the system interaction design [1] [2] [3]. The problem occurs when information presented at a rate too fast for a person to process [4], and it may lead to a list of human performance consequence [5]:

1. Fail to respond to certain inputs,
2. Respond less accurately than they would otherwise,
3. Respond incorrectly,
4. Store inputs and then respond to them as time permitted,
5. Systematically ignore (i.e., filter) some features of the input,
6. Recode the inputs in a more compact or effective form, or quit (in extreme cases) [23].

Various techniques have been explored to alleviate the problem of information overload in different application domains such as security monitoring [6], teleoperation [7], and communication systems [8]. However, the specific research of information overload for task allocation support in DR domain is still rare. We think 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 human performance consequence.

7.5.4 *Unsupported forward planning activities*

The agent performs n-step [need to confirm] looking ahead planning (see section of agent algorithm) for each re-plan. Similar to the agent, HQ players are observed to engage in forward planning activities. For example, in Ep 1, the HQ anticipated a team was going to finish assignment soon and plan to assign a new target for them. Later,

HQ triggered a re-plan and the result was consistent with the HQ's previous plan. In some other cases, the result of agent re-planning was not consistent with HQ's plan. Therefore, HQ has to make decisions about whether to follow the agent or not. As both HQ players and the agent conducted forward planning, sharing of relevant information may help HQ to understand the agent's reasoning and make informed decisions. Again, excessive and ineffective explanation of agent plans may introduce the danger of overloading the HQ players.

[Connecting to information overload]

7.5.5 *Coordination with field Players*

Coordinating with field players is also a key component of plan-execution loop. For the coordination, there are two issues emerged from the field study. The first is about confusions around the task cancellation. The second issue is about lack of response from field players. This section will exam these two issues in details.

For field players, the task cancellation is presented a notification ("task changed") followed by a blank task page only with the text ("No task assigned at the moment"). The field study suggests the presentation could be problematic for the field players. Firstly, being assigned with no task does not successfully convey the meaning task cancellation ("Don't carry on"). The episode 7 (part 2) shows a case in which the players completely ignores the task cancellation and carry on doing and risky task. In episode 5 (part 2), the player NG even think the blank task interface implies malfunctionality of the mobile interface. Even if the players understand task cancellation, its implication 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 life. 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 high radiation reading and loss of target, they don't know what to do next so they send messages to HQ for clarification. They finally left 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. However, apart from explicit task assignments, the communication system in the game probe does not support any other form of "explicit" instructions. Although it is hard to cover all possible explicit instructions, some built-in support for frequent ones can be useful. In the case of AtomicOrchid, the frequent instructions can be "stand by", "reverse back" and "escape".

Apart from the issues of task cancellation, the issues of field player feedback also emerged. The field observation showed that the feedback system is crucial for both agent and HQ players to keep track of the task status. Overall, most of (30 out of 39) the assignments have been responded by the field players. However, there are still a number of un-responded assignments that causes issues in coordination. For agent planner, the lack of field response may lead to false assumption of availability of the players, which in turn, compromises the subsequent re-planning (as presented in Ep 4 and 5). 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 (Ep 6). One strategy to view the issue of field response is to treat the lack of response as an uncertainty in the task execution. According to a classification of uncertainty by [], lack of field response can be classified as incompleteness of dataset. In Ep 6, we observed HQ's attempt's to eliminate the uncertainty by contacting the field responders. The unsuccessful attempt suggests more interface support shall be considered to help reduce uncertainty. However, it is often the case that uncertainties are inevitable. Interactive system researchers

[] have outlined a list of strategies to tackle the inevitable uncertainties [Can you see me now]. It can be argued that for safe critical domains like DR, it would be important to reveal the uncertainties to operators so that they can make informed decisions []. Not only the uncertainties, but also all the plans and actions that may be affected by the uncertainties could be revealed. For example, in the scenario of AtomicOrchid, no matter how fast the responders accept assignments, agent planning may happen in parallel. The agent planning can suffer from the lack of response if it is triggered when players have not responded (Ep 4 and 5). Therefore, it is important for operators to be aware of both the uncertainties and affected plans. One straightforward approach for visualizing the uncertainties is to highlight the affected data. Different methods for highlighting have been proposed in literature []. However, in the 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 resources. There are evidences showing that, visualization of uncertainties is likely to overload decision makers in time limited decision-making, which in turn, degrade their ability to respond in a timely manner []. Further the observed complacency effects may also compromise the effectiveness of uncertainties highlighting. Take ep 5 as example, although the system highlighted potential task interruptions as a result of lack of response, operator still ignored it.

Therefore, the interface features that counters the effects of uncertainties needs to be carefully designed. Excessive uncertainty visualization can cause information overload, and ineffective representations can be easily compromised by complacency effects.

7.6 DESIGN IMPLICATIONS

In previous studies have proposed some design implications in terms of Establishing Common Ground (CG), Balancing Division of Labour and Facilitating accountability. The interactional issues identified in this study also echo and provide some extra insights into to the design implications with regards to establishing CG and balancing the responsibilities between human and agent.

7.6.1 *Establishing common ground*

Establishing Common Ground between human and agent has been recognised an important aspect of interaction design for the AO platform in study 2. In this study, the task assignment interface is deployed to keep human HQ in the planning loop. As a result, the interaction between human HQ and the agent become more complicated than that in the study 2. Additional interactional issues have emerged from the trial of In-the-loop interaction , some of which highlight the need of more information sharing between human and agent for establishing common ground.

In more details, the study discovered that the forwarding planning activities of human HQ is not supported by the agent. The agent may be designed to provide some information with its prediction ability (e.g. prediction of future game status) to support forward planning. Further, the reasoning behind agent task allocation would also be useful for sharing. In section 7.4.9 , HQ is found overriding agent plans, which leads to undesirable results (i.e. player ‘death’). The HQ is also found being confused when agent stop assigning tasks to players with low health. Exposing internal reasoning of the agent can help the HQ to make informed decisions. However, the information shar-

ing may raise the chance information overload. More study about the trade-off might be required in the context of agent planning support to generate guidelines for interaction design.

7.6.2 Balance responsibilities between human and agent

The interaction design of Human In-the-loop studied in this field trial has some implications with regards to the balancing the responsibilities between human and agent. Firstly, there are evidence showing that the human manage to avoid some unnecessary task interruptions and team reformations caused by agent planning. It indicates that, given appropriate interface support for human monitoring and intervention, that human and agent could collectively contribute to planning, generating socially-acceptable and computationally optimized plans.

However, stronger human involvements also incurs some negative impact. It increases workload of HQ player because need to check each instruction proposed by agent. HQs sometimes failed to recognise task interruptions even with interface highlighting support, which suggest HQ can not help to eliminate all the unnecessary interruptions. Effectiveness of HQ may depends on a balanced workload [Literature!]. Achieving a balanced workload has long been recognised as a HCI challenge (Literatures). We suggest more study may be required to gain insight into the issues of HQ workload for agent planning support systems.

Furthermore, stronger human involvement may also increase the chance of human mistakes. The evidence can be found in the episode of responder "death" (ref). Although there are lots of factors contributing to the 'death' case (e.g. insufficient training for field play, commu-

nication breakdown), the task assignments sent by HQ players are thought to be too risky and not recommended by the agent at the first place. It is believed that the chance of human mistakes can be reduced by appropriate design interaction which help establishing the ‘common ground’ between human and agent. However, such a design would not be as easy task as it involves multiple design trade-offs (see section ??).

7.7 PROFESSIONAL FEEDBACKS

The section will present the results of a workshop we conducted with a professional disaster response charity called Rescue Global (RG). The purpose of these workshop is to get professional feedbacks about the AtomicOrchid(AO) platform to understand realism of the disaster simulation based on AO. Based on the feedbacks, we reflects on [strength], limitations and potential improvements of AO platform.

The workshop is centred on a demonstration of HACs system for disaster response (DR). The AO is a component of the HACs demo system. Others components include information crowdsourcing, multi-UAV control interface for DR, and DR operation provenance store [see HACer paper]. Each component are allocated one session (about 45 mins). The AO session is structured as follows:

1. A presentation introducing the AO platform and HQ interface.
2. A handles-on session for RG members to operate the HQ interface. For the purpose of demo, the field players in AO are simulated.

3. A discussion of AO platform for Rescue Global to reflect on some themes including: Command and Control structure, Division of Labour and Task and team planning.

We also encourages free discussions throughout the AO session, which leads to some emerged themes including: Data connection, time-scale of planning in real DR operations and other interface improvement suggestions. In what follows, we will briefly introduce Rescue Global and then summarise the feedbacks from the AO workshop session.

7.7.1 The Rescue Global feedbacks

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

"This is actually very similar to what we do when we do reconnaissance now. We identify needs, and we are aware of what the resources that we have and match it into what resources we have and match the needs into resources."

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

"What we did in Philippine is that , say we found school, and we talked to the [xxx] we have bumped into. They have medics, can run some clinics there. Then [MFS] 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 informal way"

The statements of RG indicates that their collaboration with the MFS medic teams are based on informal relationship with them and the formal organisational collaboration is often missing. The 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 the major challenge for realising AO paradigm is organisational. As the one RG member stated:

"The challenge with that is that UN clustering system, make separation of different tasks, they then delegate the tasks down to say Health cluster, [wash] report, 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 current paradigm would not allow it "

Based on RG's feedback, we also identifies some unrealistic aspects of AO setting. Firstly, RG member expressed the concerns related to data connection. They highlighted that the activity in AO requires good connection between field and HQ, which is not always the case in the disaster situation. Given that, the RG member can still imagine a picture of AO being used in the situation 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 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 will 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 manually input, which do not require broadband. ”

In terms of Command and Control(CC) structure, RG frequently refer to the two frameworks of CC, Bronze, Silver, Gold model (BSG, see chapter x) and National incident command framework (NICF see [x]). 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 sectorising units (geographical, functional). In the AO setting , both the command hierarchy and units sectorisation are missing. As stated by the At 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, enhanced data, at gold, stripping away all details. ”

For feedbacks, we also noticed that the time-scale of the RG's operation is much longer the AO. As one RG member describe 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 philipine, 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, dissaminate and give out jobs, it is just unbelievably slow."

Although the time-scale of AO is unrealistic for RG time, they think the AO can be seen as a ideal, speed-up version of mulit-agency operation, because it automates lots of manual process in terms of information capturing and presentation and promote 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 capturing and presentation, the RG team also see the value of AO as a tool to balance the work between field the HQ:

"I think it (AO) could be useful that you can use to help ppl in the field to gether 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.

7.7.2 Conclusion

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

Although AO are 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 future planning support system desirable to them. Firstly, the AO portrayed a picture of automated information capturing and visualisation which is appreci-

ated by RG. Secondly, RG also appreciate that the division of labour facilitated by the AO setting, 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”

7.8 SUMMARY

The objective of the research presented here was to explore the design issues surrounding an In-the-loop interaction pattern. The exploration is conducted through field trial of an AtomicOrchid(AO) probe with In-the-loop support, with particular focus in the planning activities in control room. The Game probe is evolved from the On-the-loop version of AO that is trialled in study 2 (see ref). To realise a Human In the loop interaction, the game probe allows 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. By examining the usages of interface functionalities and presenting episodes of task planning, we argue that a pattern of division of labour is achieved, in which HQ players contribute to the planning with supervisory activities and interventions, while the agent takes over computational optimization of task allocation [ref of interaction design]. Meanwhile, a set of interactional issues around In-the-loop design also emerged.

Firstly, the human supervision may not be the perfect solution for the imperfection agents. We found human HQ players sometimes quickly approve assignments without consideration of task interruptions caused by agent planning [ref]. The HQ players also override (safer) agent plans to implement risky plans, which lead to player ‘death’. Therefore, we argue that stronger human involvement may

come with undesirable human performance consequence (e.g. complacency, uninformed decisions).

Secondly, some of the observed undesirable human performance may arguably be a result of insufficient information sharing between human and agent. However, we also argue that the excessive information sharing may complicate interface and interactions between human and agent, which in turn, leads to information overload. Therefore, the trade-off may need to be carefully considered for interaction design. Further, forward planning activities are frequently observed and they are not well supported by the agent. One option to support the activity is to make the agent reveal some information regarding to future planning (e.g. prediction of future game status).

Finally, the lack of acknowledges and ineffectiveness of task cancellation have been identified as communication issues between field and the control room in section x.

The discovered interactional issues echo some design implications that we have highlighted in the study 1 and 2. Firstly, the insufficient information sharing may need to be addressed for establishing a common ground between human and the agent. Further, the Human-in-the-loop pattern is found to be useful to avoid unnecessary task interruptions and team reformation, generating socially acceptable plans. However, the greater human involve may also increase the chance of human errors, which can be an important factor to be considered in terms of balancing the labour division between human and agent.

In addition, a workshop with professional organisation has been reported in this chapter. The feedback reveals strength and weakness of AO simulation. In more details, the resource target mapping and teaming aspects of AO setting are found to be similar to activities in the DR operations. However, the time-scale, command control struc-

ture and the assumption of data connectivities does not match to current paradigm of DR operation. Although there are several unrealistic aspects of AO simulation, the responders still appreciate the core planning support aspects of AO system.

Part IV
CONCLUSION

CONCLUSION

This chapter briefly summarises the approach of this thesis, and then goes on to sum up the key findings from the field works. Finally, the thesis concludes with limitations and an outlook on future work that this thesis might inspire.

8.1 SUMMARY OF APPROACH

This PhD work use Mixed reality game - AtomicOrchid as a testbed to probe the design implications for agent planning support system. Field trials were conducted for the three versions of AtomicOrchid game with different interaction design patterns. The first one is a non agent version of the game. The second and third studies are planned to probe two interaction design patterns (On-the-loop and In-the-loop). Video and system logs of the field trails has been collected and interaction analysis has been conducted to generate requirements and interaction design implications.

8.2 SUMMARY OF FINDINGS

Three studies are carried out iteratively. The AtomicOrchid system evolves throughout the three studies. The interaction design for the study 2 and 3 follows the design patterns (On-the-loop and In-the-loop), but are also inspired by key observations from previous studies. The key observations leads to both interactional issues and detailed

system requirements which inspires system evolution across the three trials (see figure 35). The section 8.3 will give details about how the system evolves and the interface design rationales. The interactional issues emerged from field observations will be detailed in section 8.4. In what follows, the section will give a brief overview of the key observations from the three studies.

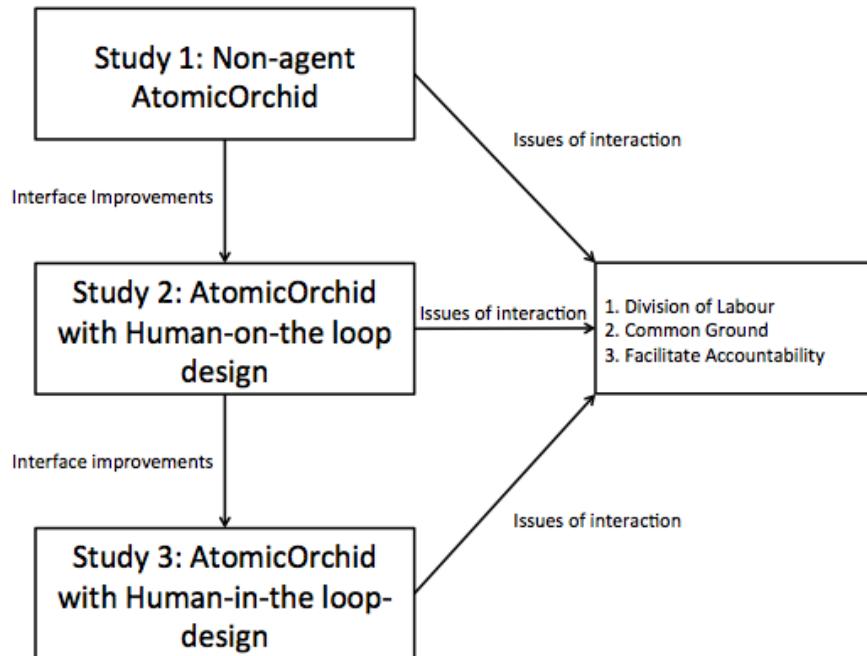


Figure 35: Interaction issues and interface improvements

The non-agent AtomicOrchid trial (study 1) is aimed to give insight into how human conduct planning (without agent support) in the time and spatially constrained disaster settings and to generate general requirements of coordination support system which is also applicable to the next two agent-integrated version of AtomicOrchid. The result of interaction analysis showed the team planning is dominated by local coordination between field players with in an “ad-hoc” manner. The field teams managed organise their team and task allocations without conflicts. The HQ is found to successfully provide awareness of the “danger zone” to the field teams. However, they

have little influence on the planning activities of the team. One potential reason could be the breakdown of communication between HQ and Field responders. The communication is thought to be affected by a set of factors including communication modalities (text vs voice), training level of players and HQ and system interface designs. A number of system requirements has also been introduced in the study 1, which inspires system evolutions in the subsequent studies.

In study 2, a planning agent is integrated to the system with the HuOL interaction design. The general requirements from study 1 leads to a number of system improvements (see figure 8.3) in study 2, which help us to avoid non-agent related factors in field trials. Through interaction analysis, we gain insight into the division of labour between human and agent (see section 6.1) in which the agent takes over routine planning activities while the human focus on other issues such as navigation. However, there are also evidence showing the agent planning occasionally interrupts workflow of human team because it fails to consider social cost of task changes. We also observed HQ player struggled to influence the plan because the lack of interface level support. Further, a set of misconceptions in feedback loops (see section 6.5.3) are also observed.

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, edit agent planning and monitor player feedbacks. Through observing the usage of this new functionalities in the control room, we found a new pattern division of labour between in which:

1. The HQ decide when to perform re-plan.
2. The HQ review every agent instruction for routine task planning.
3. The planning agent deals with player feedback.

4. The Agent propose task allocation

Further observation shows that task interruptions caused by the agent planning (observed in study 2) may have been reduced as a result of greater HQ involvement in the planning. Analysis of some failed cases of coordination also points out a number of issues, which highlights potential human performance issues as a result of the greater HQ involvement in the planning. [TO be finalised]

8.3 SUMMARY OF INTERFACE EVOLUTIONS

The section presents the interface evolutions of AtomicOrchid interface across three field studies and the design rationales behind it. The initial interface of AtomicOrchid (used in 1st study) is designed to display game status and provide broadcasting channel to support communication. In the first study, the control room is manned by 2-3 HQ players. Every HQ shares a same interface which shows game status on a map. A chat box is provided to send and receive broadcasting messages. The mobile interface is used by field responders. The responders share a same map with HQ, which displays game status with the exception of radioactive cloud. The messaging interface is placed in another tab.(see figure 36)

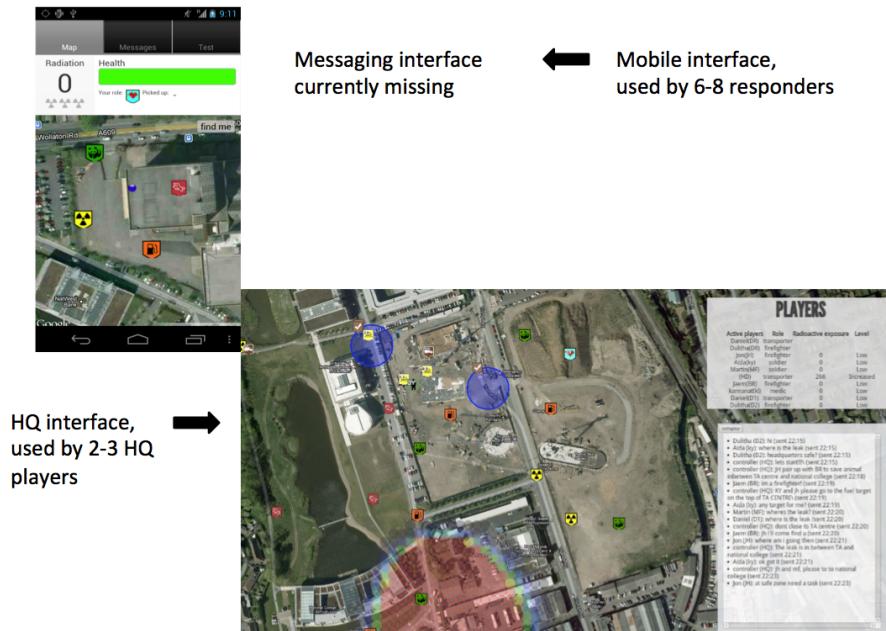


Figure 36: Interface of study 1

The 1st study discovered several flaws of the initial interface, which leads to a number of design requirements.

1. The interface should support quick acknowledgement and feedback in communication channel.
2. The interface should freshness of messages in communication channel.
3. The interface should provide support for geo-reference.

A set of changes in the 2ed version (used in study 2) of AtomioOrchid are partly inspired by the requirements (see figure 37). In 2nd version, all targets are identified by unique task ids to support geo-referencing. Text messages in broadcasting channel are labelled by time stamp to flag potentially outdated information. With introduction of planning agent, the interface also supports quick feedback to the agent with one button press. To integrate agent planning support, a new tab ("task") is implemented in the mobile app to show a text-based description of agent task-assignment. In HQ interface, the agent assignments can be revealed on request of the HQ players (HQ

clicking on "show task" button).

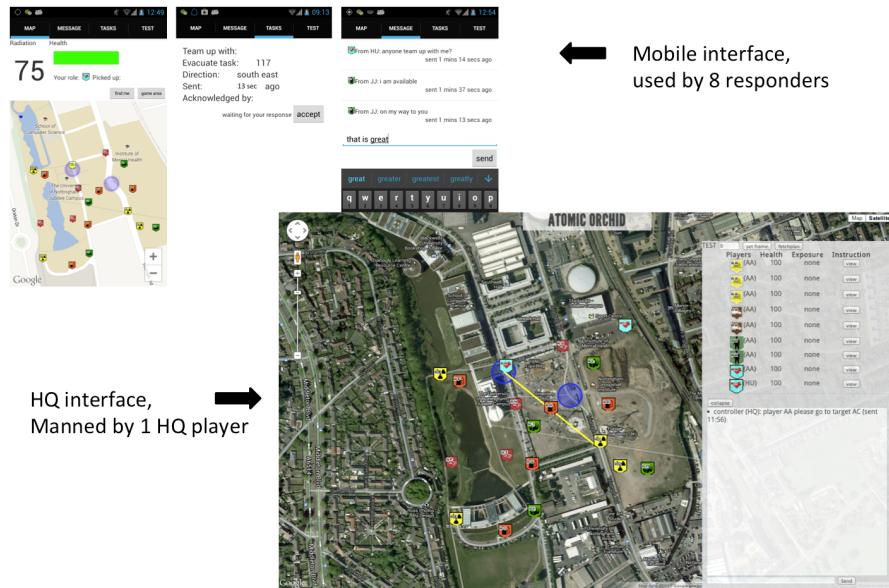


Figure 37: Interface of study 2

The result of the 2ed study revealed that interface level support for HQ intervention is missing in the 2ed version of AtomicOrchid. It is believed that the requirement for interface intervention support would also be important in the In-the-loop (the 3rd) study. Therefore, a task assignment (see figure 39) interface are introduced to enhance the HQ's ability to intervene the task planning. A task-specific communication channel is also introduced in the interface for players to avoid information overload in broadcasting channel. Compared to previous HQ interfaces, the operations on the task assignment interface is a lot more complicated. Therefore, one HQ player is dedicated to operate the interface and another player is provided with 2ed version of HQ interface by providing situational awareness and handling broadcasting messages.

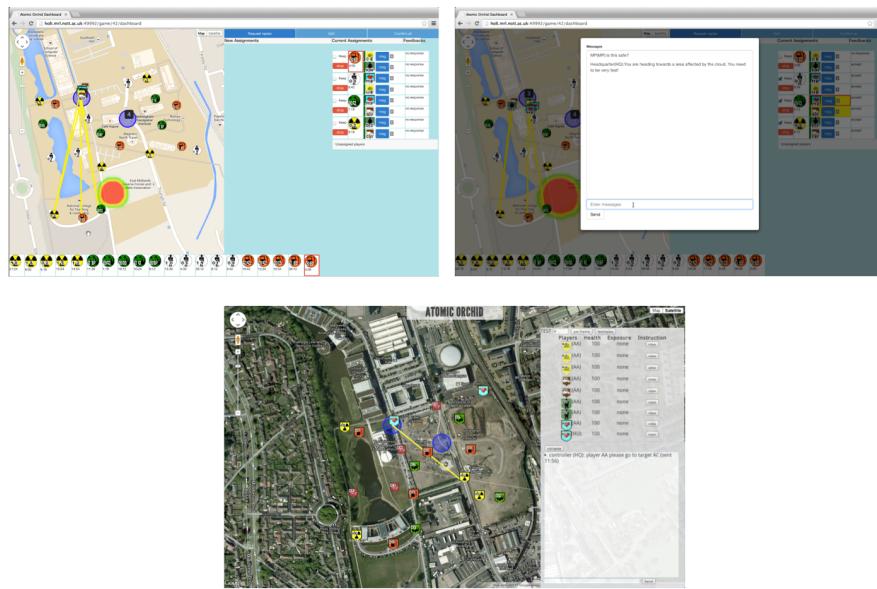


Figure 38: HQ Interface of study 3

For mobile interface, the tabs of “staus” and “chat” kept unchanged from the version 2 (named “map” and “messages”). The task interface is enhanced by a map-based presentation of task assignment and the task-specific chat box is displayed below the assignment (see figure 39).

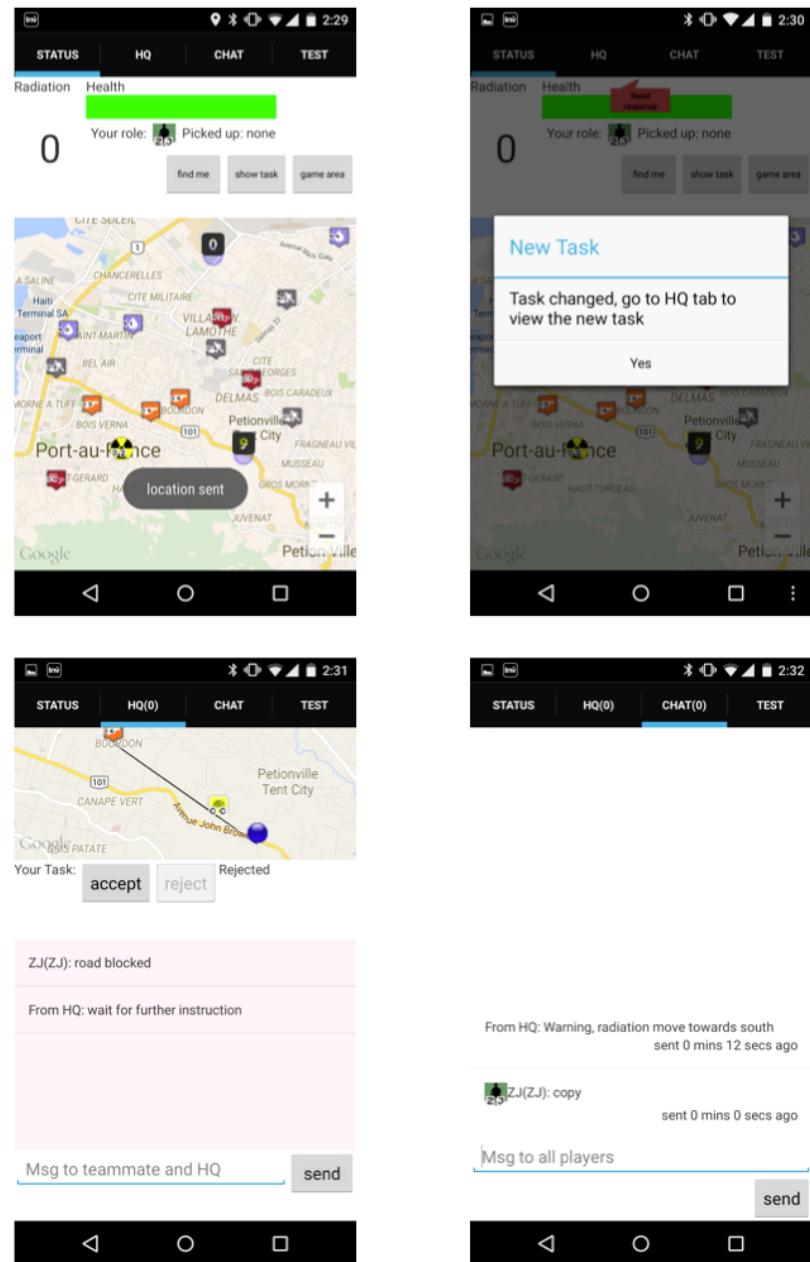


Figure 39: Mobile Interface of study 3

8.4 SUMMARY OF INTERACTIONAL ISSUES

There are a number of interactional issues emerged from the field observations of the three studies. This section will reflect on interac-

tional issues related to 3 themes including Division of labour, common ground, and accountability.

8.4.1 *Division of labour*

The On-the-loop and In-the-loop interaction designs have been trialled in two studies. The section will reflect on some of the findings related to the two interaction design patterns.

To recap, the main distinction between the two interaction design is the extent to which the human HQ is involved in routine task planning. The On-the-loop argues the minimal involvement of human HQ, leaving the agent to deal with the planning. HQ only need to deal with occasional contingencies. The In-the-loop requires constant HQ agent interaction to ensure the planning quality. Guided by these two patterns, detailed system design has been implemented.

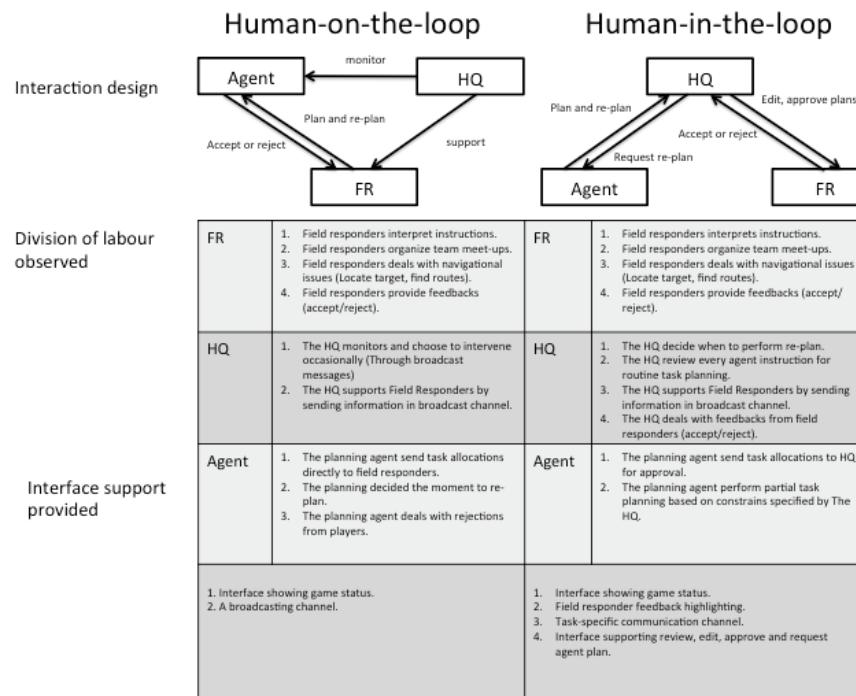


Figure 40: On-the-loop vs In-the-loop

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 applicable because the study 2 introduces a number of interface support (see figure 40) that influences team performance as well. However, the difference of performance can still leads to some implications of interaction design.

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

As we have summarised in section 7.5.2, there are evidence showing that the HuOL design is more likely to cause extra re-teaming and task interruptions 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 lack of reliability of the agent. Advanced agent planner which can better model the social process involved in task changes and take into account the possible overhead. However, given the social process might be hard to be modelled, the In-the-loop design may be useful to overcome the agent's limitation and utilise its capability at the same time.

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

crease of workload and the reduction of task-interruption

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

Compared to In-the-loop study, the HQ players in On-the-loop study are found to struggle to intervene the planning process. This performance difference can not be directly linked to the distinction between In-the-loop and On-the-loop. In On-the-loop study, the only way for HQ to intervene the planning is to send unstructured text messages in 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 On-the-loop study. It highlights the need of interface support for HQ intervention because both In-the-loop and On-the-loop requires HQ get involved when necessary.

3. Greater human involvement may introduce human mistakes.

Compared to On-the-loop, the In-the-loop design is more likely to introduce human mistakes as the design advocates greater human involvements. There is a case of human "death" 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 play, communication breakdown), the task assignment is thought to be too risky and not recommended by the agent at the first place. It is believed that the chance of human mistakes can be reduced by appropriate design interaction which help establishing the 'common ground' between human and agent. However, such a design would not be as easy task as it involves multiple design trade-offs (see section 8.4.2)[Need to expand in

study section]

4. Model of Accountability

To sum up, the reliability of agent could be one factor when considering interaction design[]. The improvement of reliability can reduce required human involvement, thus, allowing the HuOL design. However, in this PhD study, we assume human behaviour and disaster environment can be hard to be perfectly modelled. Therefore, In-the-loop design can be employed to overcome the (limited) reliability of the agent. Secondly, situational awareness of HQ can be important for both In-the-loop and On-the-loop design, though the study shows In-the-loop design help HQ to maintain situational awareness. Further, the interface support for HQ intervention has been proved to important in both In-the-loop and On-the-loop settings. Finally , a greater human involvement in planning (In-the-loop) are likely to introduce more human mistakes. The issue highlights the need for effective information sharing between human and agent, which help establishing the 'common ground' for coordination.

8.4.2 *Establishing Common Ground*

Establishing a common ground between human and agent has emerged as a common topic in discussions of both study 2 and study 3. The section will summaries the implications on establishing "Common Ground" from the two agent studies.

The "common ground" could be improved from 2 aspects: 1) Appropriate interaction design which allow human operators to understand the agent. 2) technical advancement in terms of algorithm and modelling techniques which enable agents to model human behaviours and process human feedbacks (e.g. agent being able to consider so-

cial cost of team reformation). The latter concerns about technical aspect of agent design which is beyond the scope of this PhD work. Therefore, this section will focus on building "Common ground" from the perspective of interaction design.

In both study 2 and 3, the agent behaves like a "black box" which gives the results (task allocation). The interface only exposes the results of agent for HQ to monitoring. Through our observations in chapter 6, 7, we find there could be some extra information shared between human and agent to improve planning.

The HQ players is found (section ??) to occasionally make some forward planning for field players. Because the forward planning is also performed by the agent to derive current plan, the information could be shared so that the HQ players can be provided with the agent suggestion when doing forward planning.

The reasoning behind current task allocation would also be useful for sharing. In chapter 7 , HQ is found overriding agent plans, which leads to undesirable results. The HQ is also found being confused when agent stop assigning tasks to players with low health. Exposing internal reasoning of the agent can help the HQ to make informed decisions.

Misunderstanding between agent and human is also observed in the feedback loop in study 6. Firstly, human respond do not know how agent is going to handle the rejection. They try to use rejection to reverse back to previous tasks, while the agent will give them more new instructions. Secondly, it is unknown to field responders that their rejection will cause replanning for the whole team which can lead to lots of costly task interruptions. Therefore, information indicating consequence of interface interactions should be also made available to human to facilitate accountability and ensure informed

decisions.

Although, we have identified a range of information which is missing for establishing "Common Ground", presenting the information could be also challenging. The information should be delivered in right form (e.g text, visualisation, dialogs) and in right time (e.g. pop up or on HQ request?)[] Especially in the multi-tasking, time-critical settings like AtomicOrchid, multiple sources of information can compete for attention of the human operator. Information overload could be a real danger of interaction design in this setting. Therefore, the way for exposing agent's information and its human performance consequence may need to be further studied.

8.4.3 *Faciliating Accountability*

8.5 LIMITATION

This section will outline 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 the distributed, time-critical task setting. The fictitious game scenario can not completely mirror the setting of disaster operations. Firstly, the time-scale of planning is much longer in a rescue and reconnaissance mission (from hours up to days), while the AtomicOrchid only have 30 mins time-scale. Secondly, the disaster response teams always prioritise critical tasks according to code of conduct (e.g. task with human injuries will always be prioritised. see section of RG interview). Therefore, the game objective of maximising targets saved may not match real goal of DR team.

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

The untrained participants are recruited for the field trials. We anticipate that the behaviours of players in the game may change with their training experience. Further, the professional responders can also behave differently because they have high level of training, real experience of DR operations, different code of conduct and organisation structure.

Overall, the PhD work employed a game approach, which has a number limitations in terms of game scenario, participants' altitude, and participants' selection. We can not claim the observations of human behaviours would be exactly the same as what happens in the real disasters. However, we argue that the game-based approach managed to introduce some critical factors of DR operations, including time pressure, distributed team setting, and the mental/physical stress. Compared to computational simulation, the field trials can reveal rich human-system interactions under time and spatially constrained task environment. Therefore, we argue the observations from game trials are still valid and can be used to generate design implications for future HACs systems.

8.6 FUTURE WORK

Following the limitations and contributions of the thesis, its impact can be extended into future work in several possible ways. Firstly, the planning agent itself might be improved with various AI tech-

nologies. For example, the agent in this study does not consider the social cost of re-teaming and task interruption. Some user modelling technologies such as [] have potential to be used predict human behaviour, which in turn, helps to model the social cost of re-teaming and task interruption [] . However, when 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 operation team. In order to do that, the close collaboration with disaster response teams may be required. The collaboration with Global Rescue happened in the late stage of this PhD work, which help 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 can make the game setting more realistic. Secondly, radio rather than text messages is the communication modality used by operation teams, which may have impact on our behaviour observations. Thirdly, the AtomicOrchid assume constant connectivity between HQ and field responders, but the 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 operation. As we believe that the trained professional responders may behave differently with the general public. It would

be desirable to recruit professionals in the future work to validate 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 difference in terms of games results (including players death, targets saved) can be used to imply effectiveness of interaction designs. Also the social cost and workload of HQ may also be quantified (?) to give insight into the impact of interaction designs on the team performance. (more?)

The studies have highlighted the trade-off between establishing common ground and information overload (see section x). It is believed that further studies may be needed to unpack how the trade-off affects team coordination and develop systematic approach for system designer to balance trade-off.

Moreover, the interaction designs are not limited to the On-the-loop and In-the-loop. 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 the centralised planning support is only one kind of agent planning support technology among many. For example, the planning support can also be decentralised. The planning agents can be built into personal assistant devices for every field responders. The change of agent technologies certainly have implications on interaction design. New interaction design pattern can be devised and adopted with the change of agent technologies. There-

fore, this PhD work can potentially be extended with exploration of different agent technologies.

Further, this work contributes to the research paradigm of HACs system (see section x) by studying the interaction between one centralized planning agent and a disaster response team. However, the HACs researchers envision a scenario in which large amount 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 incorporate multiple agents and human teams to study multi-agents and human interaction.

Part V

APPENDIX

A

APPENDIX TEST

Lorem ipsum at nusquam appellantur his, ut eos erant homero concludaturque. Albucius appellantur deterruisset id eam, vivendum partiendo dissentiet ei ius. Vis melius facilisis ea, sea id convenire referrentur, takimata adolescens ex duo. Ei harum argumentum per. Eam vedit exerci appetere ad, ut vel zzril intellegam interpretaris.

Errem omnium ea per, pro congue populo ornatus cu, ex qui dicant nemore melius. No pri diam iriure euismod. Graecis eleifend appellantur quo id. Id corpora inimicus nam, facer nonummy ne pro, kasd repudiandae ei mei. Mea menandri mediocrem dissentiet cu, ex nominati imperdiet nec, sea odio duis vocent ei. Tempor everti appareat cu ius, ridens audiam an qui, aliquid admodum conceptam ne qui. Vis ea melius nostrum, mel alienum euripidis eu.

A.1 APPENDIX SECTION TEST

Ei choro aeterno antiopam mea, labitur bonorum pri no. His no decore nemore graecis. In eos meis nominavi, liber soluta vim cu. Sea commune suavitate interpretaris eu, vix eu libris efficiantur.

More dummy text.

Nulla fastidii ea ius, exerci suscipit instructior te nam, in ullum postulant quo. Congue quaestio philosophia his at, sea odio autem vulputate ex. Cu usu mucius iisque voluptua. Sit maiorum propriae at, ea cum primis intellegat. Hinc cotidieque reprehendunt eu nec. Autem timeam deleniti usu id, in nec nibh altera.

LABITUR BONORUM PRI NO	QUE VISTA	HUMAN
fastidii ea ius	germano	demonstratea
suscipit instructior	titulo	personas
quaestio philosophia	facto	demonstrated

Table 1: Autem usu id.

Listing 1: A floating example

```
for i:=maxint to 0 do
begin
{ do nothing }
end;
```

A.2 ANOTHER APPENDIX SECTION TEST

Equidem detraxit cu nam, vix eu delenit periculis. Eos ut vero constituto, no vidit propriae complectitur sea. Diceret nonummy in has, no qui eligendi recteque consetetur. Mel eu dictas suscipiantur, et sed placerat oporteat. At ipsum electram mei, ad aeque atomorum mea.

Ei solet nemore consecetur nam. Ad eam porro impetus, te choro omnes evertitur mel. Molestie conclusionemque vel at, no qui omittam expetenda efficiendi. Eu quo nobis offendit, verterem scriptorem ne vix.

BIBLIOGRAPHY

- [1] Jeffrey M Bradshaw, Paul Feltovich, and Matthew Johnson. Human-Agent Interaction. In G Boy, editor, *Handbook of HumanMachine Interaction*, chapter 13, pages 293–302. Ashgate, 2011. URL <http://books.google.com/books?hl=en&lr=&id=4opHlu05SNIC&oi=fnd&pg=PA283&dq=Human-agent+interaction&ots=vxrpDdLbSa&sig=07dujtzGjIcBLlZ6FVH33HjrWos>.
- [2] Fitts. Human Engineering for an Effective Air Navigation and Traffic Control System. 1951.
- [3] Robert R Hoffman, Paul J Feltovich, Kenneth M Ford, West Florida, David D Woods, Gary Klein, Klein Associates, Anne Feltovich, Grinnell College, Editors Robert R Hoffman, and Patrick J Hayes. Human-Centered Computing A Rose by Any Other Name ... Would Probably Be Given an Acronym. 2002.
- [4] D Mendonca. Decision support for improvisation in response to extreme events: Learning from the response to the 2001 World Trade Center attack. *Decision Support Systems*, 43(3):952–967, 2007. ISSN 01679236. doi: 10.1016/j.dss.2005.05.025. URL <http://linkinghub.elsevier.com/retrieve/pii/S0167923605000904>.
- [5] Hyacinth S Nwana. Software Agents : An Overview. pages 1–49, 1996.
- [6] Nikos Vlassis. *A Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence*, volume 1. January 2007. ISBN 9781598295269. doi: 10.2200/S00091ED1V01Y200705AIM002. URL <http://www.morganclaypool.com/doi/abs/10.2200/S00091ED1V01Y200705AIM002>.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both L^AT_EX and LyX:

<http://code.google.com/p/classicthesis/>

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

<http://postcards.miede.de/>

DECLARATION

Put your declaration here.

Nottingham, August 2012

Wenchao Jiang