A MODEL OF EMPATHY FOR ARTIFICIAL AGENT TEAMWORK

by

BEHROOZ DALVANDI

B.Sc. Computer (Software) Engineering
Arak University, Arak, Iran.

THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
MATHEMATICAL, COMPUTER, AND PHYSICAL SCIENCES
(COMPUTER SCIENCE)

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

May 2012

Your file Votre référence
ISBN: 978-0-494-94132-4

Our file Notre référence
ISBN: 978-0-494-94132-4

Abstract

This thesis introduces a model of empathy as a basis for helpful behaviour in teams consisting purely of artificial agents that collaborate on practical problem-solving tasks, and investigates whether the performance of such teams can benefit from empathic help between members as the analogy with human teams might suggest. Guided by existing models of natural empathy in psychology and neuroscience, it identifies the potential empathy factors for artificial agents, as well as the mechanisms by which they produce affective and behavioural responses. The performance of empathic agent teams situated in a microworld similar to the Coloured Trails game is studied through simulation experiments, with the model parameters optimized by a genetic algorithm. For low to moderate levels of random disturbance in the environment, empathic help is superior to random help, and it outperforms rational help as rational decision complexity grows, in particular at higher levels of environmental disturbance.

Contents

List of Figures

Acknowledgements

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost I offer my sincerest gratitude to my supervisor, Dr Jernej Polajnar, who has supported me throughout my thesis with his patience and knowledge whilst allowing me the room to work in my own way. I attribute the level of my Masters degree to his encouragement and effort; without him this thesis would not have been completed. One simply could not wish for a better or friendlier supervisor, who cares about all aspects of his students' life besides their study.

I am deeply grateful to Dr Desanka Polajnar, adjunct professor at the Computer Science Department at the University of Northern British Columbia, for her excellent moral and technical support and instructive comments on my thesis over the past two and a half years.

I am greatly indebted to Dr Ken Prkachin, professor at the Department of Psychology at the University of Northern British Columbia, for introducing me to the relevant ideas in psychology and providing guidance in the development of my thesis topic.

My special thanks go to my colleague Omid Alemi, and my cousin Ali Khorami, for their technical advice and friendly support during my work on this thesis.

Chapter 1

Introduction

Humans in everyday life experience situations in which they can benefit from teamwork in order to solve a problem. In such cases the primary requirement for every team member is to have the necessary technical abilities to complete their own tasks. For example, in a team which is formed to design and develop a business website, having people with specific abilities in graphic design, programming, and database management is essential. These technical abilities can be effectively tested and measured. They are sometimes classified as 'visible' skills [Wysocki et al., 1995].

However, there are also other skills that are not easy to test and yet are known to affect team efficiency. One category of such skills, collectively known as emotional intelligence, is increasingly viewed as an important component of team success. Luca and Tarricone [2001] describe emotional intelligence as consisting of five elements that are considered invisible skills: self awareness, self regulation, empathy, motivation, and social skills. Consistent with the long-time experience in engineering, sports, and other domains of human teamwork, their experiments demonstrate the importance of emotional intelligence in general and empathy in particular in leading a student project team to success.

Evidence in neuroscience indicates that emotions play a key role in human decisions. Based on analysis of clinical cases, Damasio [1995] argues that emotions limit the decision space in which logic is being used and that they are intimately linked to reason. Research in neuroscience has shown that brain damage that impairs affective reactions lowers the performance in dealing with time-constrained decisions, resulting in poor judgement of real-life situations. If these findings can be extrapolated to the team level, it is likely that empathy improves teamwork not only because of its comforting effect, but also because it improves the quality of reasoning and the maturity of the resulting decisions.

Given the recent developments in the theory and practice of multiagent systems [Wooldridge, 2009], the studies of teamwork are no longer limited to living systems. As artificial intelligence and robotics progress from laboratory exploration towards mainstream engineering practice, agent-oriented software engineering (AOSE) has become a widely accepted paradigm that may succeed object-oriented software engineering (OOSE) as the dominant software development methodology. With the ascent of networking and distributed computing, the research focus is shifting from individual agents to multiagent systems in general and agent teamwork in particular. There are many studies about teamwork and team-based decision making in artificial agents: how they form a team, how they agree on a common goal, and how they work as team members to achieve that goal (Cohen and Levesque [1990], Levesque et al. [1990], Cohen and Levesque [1991], Wooldridge and Jennings [1994], Grosz and Kraus [1996], Wooldridge and Jennings [1999], Aldewereld et al. [2004], Sycara and Lewis [2004], Brzeziński et al. [2005], Dunin-Keplicz and Verbrugge [2010]). However, the proposed theories are mostly based on the practical reasoning of agents and do not include the representation of emotions.

In computer science, the incorporation of emotions into artificial systems has been

so far mainly concerned with recognizing human emotions and showing emotional expressions and empathy in order to better interact with human users. Research in affective computing has firmly established the significance of computational models of human emotions that enable artificial agents to display empathy for human beings in their mutual interaction [den Broek, 2005]. The future success of many envisioned robotics applications, such as home care for the elderly, depends on empathic agent technology. The use of multiagent systems in simulations of human social interactions also includes the modeling of emotional behaviour. However, Picard [1995] argues that computers must be enabled to use emotional mechanisms in the process of making decisions if we want them to be truly effective.

So far there have been few concrete studies about endowing artificial agents with emotional mechanisms for decision making. One of them is the EBDI model proposed in [Jiang et al., 2007], which extends the well-known Belief-Desire-Intention model (BDI), introduced by Bratman [1987], with an 'emotions' component that affects agents' reasoning. Steunebrink et al. [2010] investigate how emotions can be used to specify constraints on agents' reasoning cycle to reduce non-determinism. Nair et al. [2004] also discuss the possibility of having emotions in pure artificial agent teams, and emphasize the potential importance of emotional mechanisms for efficient and effective teamwork. Nevertheless, to the best of our knowledge, emotions are not yet widely considered as an essential component in interactions between artificial agents.

Given the significance of emotions in human decision making and the role of empathy in human teamwork on one hand, and the increasing practical importance of artificial agents' teamwork on the other hand, several questions arise. Can emotions, and empathy in particular, play an important role in teams consisting of purely artificial agents? If so, how can a suitable notion of empathy be defined for such systems? How is it inspired by and related to the corresponding human emotion? In particular,

are there indications that teams of artificial agents could perform better when their members are endowed with empathy for their teammates? If so, how could such improvements be realized in practice? Such questions have not been investigated much so far and definitely need to be systematically studied. We raise some of them in [Polajnar et al., 2011] and provide a modelling framework for incorporating empathy into artificial agent teamwork. However, the results of that paper do not include the models of factors and mechanisms by which the empathic responses in artificial agents are formed. To the best of our knowledge, such modelling is first undertaken in this thesis.

The central contribution of this thesis is a model of empathy in artificial agents that encompasses the empathy factors that influence the affective response, the combining mechanism for the formation of affective response, and a threshold mechanism that triggers the behavioural response. The specific role of the formulated notion of empathy is to be used in decisions by agents in a team on whether to provide direct help (outside of the general team organization and subtask allocation) to a teammate in distress. The assumption is that the team as a whole works on a practical problem-solving task (as teams of artificial agents typically do), and that agents experience distress when they encounter difficulties in performing their subtasks. Such difficulties are often caused by unpredictable dynamic changes in the environment in which the team is situated. The model is then validated and its properties investigated through simulation experiments in which a team of agents play a cooperative game in a microworld designed for studies of helpful behaviour and its impact on team performance. The microworld includes a disturbance parameter that controls the level of random dynamic change in the environment.

The starting point in the construction of the model is an analysis of empathy factors that is inspired by and based on the Perception-Action Model (PAM) in psy-

chology and neuroscience, introduced by Preston and de Waal [2002]. The authors of PAM identify six factors that can influence the affective responses in natural empathy: depression, similarity, familiarity, learning, past experience, and salience. For each of these factors, we formulate an analogous concept for artificial agents in a team, discuss its potential relevance to team performance, and the requirements that agent design and the operating environment must satisfy in order to make the concept meaningful and relevant. For each factor, we also examine the feasibility of its implementation in the microworld environment used in our simulation experiments. While our model is directly inspired by studies in natural empathy, there is no strict requirement that each empathy factor for artificial agents should faithfully mimic its natural counterpart. Indeed, a designer of artificial empathic agents could be motivated to define empathy factors with no analogue in the living world, but our current scope does not include such possibilities.

The proposed model of empathy in artificial agents is fairly general and gives rise to a variety of stimulating research questions. We discuss some of those possibilities in connection with future work. In the thesis we proceed to show, using a simplified model implemented in a microworld context, that our concept of empathy can indeed serve as a valid trigger for helpful behaviour that leads to better performance of the team. The simplified model uses three of the six empathy factors, realized in terms of microworld concepts, and stipulates that all agents in the team have identical empathy profiles. The parameters of the simplified model are first optimized, using genetic algorithms, to maximize the performance of the team. The experiments show that, for low to moderate levels of disturbance in the environment, a team in which help decisions are based on empathy outperforms a team in which help decisions are random, even if we ensure that the overall rate of positive help decisions is the same for both teams. These results demonstrate that the empathic mechanisms defined in

this thesis are valid triggers for helpful behaviour that can improve the performance of an artificial agent team.

The microworld used for simulation experiments is a variation of the Coloured Trails game [Gal et al., 2010]. It has been developed specifically for the study of helpful behaviour in teamwork and implemented independently at UNBC. The microworld has so far been used in the study of rationally motivated help in artificial agent teams, based on the Mutual Assistance Protocol (MAP) introduced by Nalbandyan [2011] and Polajnar et al. [2012]. This facilitates a comparison between empathic and rational help. Since rational help decisions in the relatively simple microworld do not involve deliberations of realistic complexity, the cost of a rational help decision is modelled as an independent parameter. This precludes realistic performance comparisons between empathic and rational help, but still allows the identification of some general trends. The experiments show, as expected, that rational help is superior when the cost of rational decision is low, and is superseded by empathic help as the growing complexity of rational decisions leads to higher costs. This crossover happens sooner in the case of higher disturbance in the environment, suggesting that empathic help can be more effective than rational help in unpredictable circumstances.

The model of empathy introduced in this thesis complements and strengthens some of our earlier results described in [Polajnar et al., 2011]. In that paper we adapt an existing model of natural empathy model, introduced by Goubert et al. [2005], to explain the formation of affective response in artificial agents from top-down influences (experienced by the subject of empathy) and bottom-up influences (related to the object), and to show how they ultimately lead to behavioural responses of certain types. The current model clarifies the nature of those influences and of the mechanisms involved in the the formation of empathic responses. Our paper also introduces the Empathic Behavioural Response Algorithm, which shows how BDI

agents endowed with empathy can provide different levels of problem-solving help to each other, assisting at the level of beliefs, desires, intentions, plans, or executions. Again, the current model makes its specification more complete by providing concrete mechanisms for the formation of affective and behavioural responses used by the algorithm.

The rest of this thesis is organized as follows. In Chapter 2 we introduce the background concepts and review the work related to our research, from both psychology and computer science points of view. In Chapter 3 we formulate in detail the problems we address in this thesis, clarify how they relate to our earlier published work, and outline our general solution strategy. Chapter 4 introduces our model of empathy for artificial agents. Chapter 5 presents the experiments and results, including their analysis and evaluation. In Chapter 6 we draw our final conclusions together with some suggestions for future work.

Chapter 2

Background and Related Work

We get our original motivation for this research from the study of empathy in psychology and we try to create an empathy-like mechanism for artificial agents. Then, we aim to study the performance of a team of artificial agents endowed with such mechanism for performing helpful behaviour. Therefore, there are three main categories of work we are interested in studying: first, the study of natural emotions and empathy from the psychological point of view; second, the study of emotions and possibly empathy in artificial agents; and third, the study of teamwork and help protocols in artificial agent systems.

As in this research we are mainly concerned with using the notions of emotions and empathy in artificial intelligence we first briefly review the work in that area that is relevant to our purposes (Section 2.1). Next, we introduce the psychology sources and references that we have been using in our research (Section 2.2). Finally we discuss the relevant studies of agent teamwork and helpful behaviour among agents (Section 2.3).

2.1 Emotions and Empathy in Artificial Agents

Computers with emotions had received little attention from researchers (as opposed to science fiction authors) until seventeen years ago, when the work of Rosalind Picard [Picard, 1995, 2000] laid the groundwork for the field of affective computing. Picard analyzes what it would mean for computers to: recognize emotions, express emotions, have emotions, and have emotional intelligence. Citing the thorough and convincing neurological evidence that human emotions are essential to rational thinking [Damasio, 1995], she argues that "computers, if they are to be truly effective at decision making, will have to have emotions or emotion-like mechanisms working in concert with their rule-based systems" [Picard, 2000]. One could further argue that agent teams, if they are to effectively perform complex tasks involving individual and collective decisions, must also rely on suitably defined emotion-like mechanisms. In that context, the significance of empathy in human teamwork suggests a potential significance of empathy-like mechanisms in agent teamwork.

The mainstream development of affective computing has so far focused on human-computer interaction. The emphasis is on the recognition of human emotional state, the synthesis of a proper emotional response, and the expression of that response in a manner recognizable to humans. The recognition of emotional state through perception and analysis of voice, facial expressions, etc., underlies the design of empathic agents [den Broek, 2005]. The study of empathy with a human object is thus central to contemporary affective computing.

The role of emotions in a team of agents is analysed in [Nair et al., 2004] by some of the leading researchers in multiagent teamwork. The authors consider three types of agent teams: (i) teams of simulated humans; (ii) mixed agent-human teams; and (iii) pure agent teams. They emphasize the first two types and present experimental

results that quantify the impact of fear on the performance of a team of human-piloted helicopters in combat. They introduce a scenario in which the pilots can use different paths to the destination, while some of those paths might be protected by enemy troops. Pilots can communicate with each other. The paper considers an example situation in which a pilot feels fear in the voice of another pilot who has gone through a specific path and he notices danger in that path based on it and modifies his decision. Their examples of promising emotional mechanisms are essentially empathy-like. Their brief treatment of pure agent teams affirms the potential importance of emotional mechanisms for effective teamwork.

Jiang et al. [2007] propose an extension for the well-known Beliefs-Desires-Intention (BDI) model Bratman [1987], in order to incorporate emotions in the practical reasoning process of BDI agents. Their work considers primary and secondary emotions (following [Damasio, 1995]) and based on it reformulates the practical reasoning process. They introduce a new 'emotions' component along with the three components of the BDI model (Beliefs, Desires, and Intentions) to form an EBDI model. In their model agents update their emotions toward other agent in each reasoning cycle and practically involve their emotions in the decision making process. Their results suggest that agents capable of having emotions in their reasoning would have a better performance than agents without them.

Memon and Treur [2009] propose a design for social agents capable of understanding other agents in an empathic way. Their paper addresses a way to design agents with mechanisms to understand the other agents' mental state and subsequently generate the same feelings as the observed agents. Their design is mainly concerned with the 'mind reading' aspect of the problem; how an agent can generate corresponding beliefs based on another agent's mental state. However, it does not evaluate the performance of these empathic agents compared to non-empathic ones.

Steunebrink et al. [2010] investigate how emotions can be used to specify constraints on agents' reasoning cycle to reduce non-determinism. They emphasize the point that, although the common 'sense-reason-act' cycle of the agents looks reasonable at the first glance, realistically an agent faces many different choices when making decisions. Therefore, most of the decisions are made non-deterministically. The authors propose a way to shrink the decision space using a representation of emotions. They use OCC (Ortony, Clore and Collins) model in psychology as a reference to model emotions in the agent reasoning cycle. Their work considers four types of emotions: joy, hope, fear and distress which are triggered by different events. Events are either 'actual' or 'prospective'. An event is *actual* if the agent believes it has happened, otherwise it is *prospective.* A desired actual event triggers joy, while a desired prospective event triggers hope. Likewise, an undesired actual event causes distress while an undesired prospective event causes fear. Based on this, the decision space is limited by considering four rules: (i) Plan generation rules are applied only to goals that have triggered hope; (ii) Plan revision rules are applied only to plans that have triggered fear; (iii) Plans that have triggered joy are preferred over other plans; and (iv) Plan execution is interrupted as soon as distress is triggered.

While this does not lead to a complete determinism, it does reduce the choices. The paper does not consider a team of agents, it only discusses the reasoning cycle of one single agent. Also, the paper lacks an experimental result that illustrates how representation of emotions can improve the agents' performance. They emphasize that they do not consider a specific emotional state for agents and emotion types are just used as labels to relate particular cognitive states of an agent.

In [Polajnar et al., 2011] we have raised the question of whether and how empathy between artificial agents can improve their team performance. In that paper we have explored the notion of empathy between artificial agents and argued that it can

have significant impact on the design of robust and resilient agent teams. The paper outlines a basic model of empathy in a team of artificial agents, which is inspired by a model of human empathy discussed in [Goubert et al., 2005]. Also, connecting the Perception-Action Model of human empathy [Preston, 2007] to the reasoning models of artificial agents, we have enhanced the EBDI model [Jiang et al., 2007] to include different components of empathic behavioural response.

One of the most important points we have made in [Polajnar et al., 2011], is the conjecture that empathy has the potential of improving artificial agent teamwork performance by initiating helpful behaviour in a team. Based on a simple simulation, we have illustrated how helpful behaviour improves the performance of a team under specific circumstances related to environmental dynamics and disturbances. We have proposed the Empathic Behavioural Response Algorithm (EBRA) for modelling an agent's activity in favour of another agent (as an empathic response). EBRA is based on the BDI model and formulates an agent's practical reasoning in assistance to another agent that is stuck in its task. In EBRA, the helper agent can assist the agent that is asking for help at five different levels: assisted revision of beliefs, assisted revision of desires, assisted revision of intention(s), assisted revision of plans, and execution of the plan.

2.2 The Relevant Literature in Psychology

As Antonio Damasio argues in [Damasio, 1995], emotions guide behaviour and decision making, and rationality needs emotional input. His theory emphasizes the crucial role of feelings in navigating the endless stream of life's personal decisions [Goleman, 1997], helping to reject immediately the negative courses of action and consequently allowing one to choose from among fewer alternatives.

There is not an absolute agreement in the literature about the exact nature of empathy [Preston and de Waal, 2002]. Essentially, empathy is about one person as *subject* having a sense of knowing the personal experience of another person as *object* that results in affective and then behavioural response of the subject [Preston, 2007]. Affective response in the subject is generating similar feelings as the object has and behavioural responses of the subject might be some actions that are intended to alleviate the object's difficulty.

As we previously mentioned in Chapter 1, Luca and Tarricone [2001] emphasize the role of emotional intelligence in successful human teamwork and name five important elements of emotional intelligence as self awareness, self regulation, empathy, motivation and social skills. The paper also points out that these skills are not as easy to test as technical skills.

Goubert et al. [2005] categorize the four notable characteristics of empathy:

> "First, we contend that empathy is not exclusively for humans. Second, the inferred experience of the other may comprise thoughts, feelings or motives. Third, empathy may manifest itself in various ways. Some of these may be automatic and implicit. Others might be explicit and depend upon the intentional and effortful use of cognitive processes. Fourth, affective responses to facing another person may often, but not always, entail sharing that person's emotional state."

They argue that a human needs an appropriate level of empathy to be capable of performing helpful actions,as lack of empathy causes lack of care for others and over-empathic behaviour causes getting overwhelmed by the object's experience.

They also divide the stimuli of empathy into two main categories: top-down and bottom-up influences. *Top-down influences* are those that are caused by the subject's

own former experience which is similar to what the object is experiencing; *bottom-up influences* are those that are caused by the observation of the object's expressions. The paper emphasizes that these influences finally lead the subject to some affective and possibly behavioural responses.

Preston and de Waal [2002] name five stimuli by which empathy increases: similarity, familiarity, past experience, learning, and salience. They argue that there are two main types of causes for empathy, proximate and ultimate, as it is also stated in [Mayr, 1961]:

> "Proximate causes govern the responses of the individual (and his organs) to immediate factors of the environment while ultimate causes are responsible for the evolution of the particular DNA code of information with which every individual of every species is endowed."

According to [Preston and de Waal, 2002], empathizing is an automatic process unless the subject prevents it for some reason:

> "Empathy specifically states that attended perception of the object's state automatically activates the subject's representations of the state, situation, and object, and that activation of these representations automatically primes or generates the associated autonomic and somatic responses, unless inhibited."

Preston [2007] talks about the perception-action model (PAM) that emphasizes the degree of matching between the subject and object. The subject needs to have representations for the state of the object in order to be able to empathize; the more similarity between the subject's and object's representation of the state, the greater the likelihood of an empathic response:

> "In order to achieve empathy, subjects must be motivated to and capable of attending to the state of the object, they must be able to activate personal representations of a similar state, and to generate an emotional response. Thus impairment in any of these phases will create an impairment of empathy."

The paper mentions that, for example, a depressed person may not be able to perform empathic response, due to an excessive focus on self. So the level of empathy not only depends on the subject's level of understanding of the object's state, but also on the subject's personal situation.

2.3 Helpful Behaviour in Artificial Agent Teams

In this section, we cite papers that study helpful behaviour in a team of artificial agents. The following papers are not concerned with the role of emotions in decision making. Nevertheless, they are related to our current research as they study the possible solutions for including helpful behaviour in an agent team and our aim is also to use empathy as a trigger for initiating helpful behaviour in a team of artificial agents.

Kamar et al. [2009] propose a decision-theoretic mechanism for helpful behaviour and collaborative teamwork. In this mechanism, agents rationally decide about helping other agents. These decisions are based on the believed team utility of the actions and require agents to be able to model how their surrounding world is changing. In this proposed mechanism, agents need to find out whether the considered help action is beneficial to the team to perform a helpful action.

In such a mechanism, it is required for agents to either know or have an estimation

about the values of probability that they can bring about a specific action, plus the cost and the value of that action. This approach relies on introducing the concept of *probabilistic recipe trees (PRT)* that enable agents to represent their beliefs about the recipes that might be chosen by the members of the team to complete a collaborative activity.

That paper proposes two types of helpful behaviour: communication, and adding helpful acts to the group plan. The former is done by an agent helping another agent to update its beliefs, and the latter is done by actually performing an action in favour of another agent. The experimental results provided in the paper demonstrate the superiority of this mechanism compared to purely axiomatic methods which are non-decision theoretic models without probabilistic representation.

Cao et al. [2005] discuss proactive helpful behaviour among (sub)teams of agents. The work mainly focuses on identifying help needs and providing help correspondingly, and considers two types of helpful behaviour: (i) Agent A takes an action in favour of agent B if agent B has failed in that action (backup behaviours), (ii) Agent A helps agent B to achieve conditions required by what agent B is doing (promotion behaviours). The paper proposes a formal model based on shared mental states of agents by which agents can identify each other's help needs and take an action correspondingly. The agents are aware of each other's tasks. Therefore, an agent can monitor another agent's activity and, based on its own understanding of the situation, decide whether it should help or not. The agent A will help agent B if agent A is sure that agent B cannot finish the task, and agent A's intervention will change the situation positively.

In that paper there is no mechanism defined for deciding about asking for help and agents keep checking help needs in regular time periods without being asked for it.

Nevertheless, the experimental results demonstrate the usefulness of having helpful actions incorporated in agents' behaviour.

Nalbandyan [2011] proposes a novel protocol called *Mutual Assistance Protocol* (MAP) for incorporating helpful behaviour into multiagent teamwork. In *MAP*, an agent can use its own local resources and capabilities to assist by performing an action (or providing resources) towards a subtask that has been assigned to another agent. The agents participating in a prospective help act both judge whether the act benefits the team; and the act happens only when both sides have jointly agreed that it does. Each agent's assessment of team benefit is based its evaluation of the team impact of its changing its own local plan.

The thesis compares MAP to unilateral approaches for helpful behaviour, where the decision for performing a help act is made by only one side, either the receiver or the helper. The comparison is done through different simulation experiments using varying levels of mutual awareness in the team, dynamic disturbance in the environment, communication cost, and computation cost. These results demonstrate the superiority of MAP over unilateral decision mechanisms for helpful actions, given that in different situations the beliefs of team members about each other's abilities might not be accurate.

The studies we cited in this section use different approaches for making decision about performing helpful behaviour, but they all use rational reasoning for that purpose.

Chapter 3

Incorporating Empathy into Artificial Agent Teamwork

This chapter describes the research problem addressed in this thesis, presents our early results that establish its conceptual framework, and sets the direction of research for the remaining chapters. Here, we first present an overview of the problem, clarify our motivation for studying it, and specify the general direction of our research (Section 3.1). After that, we briefly review our previous work in [Polajnar et al., 2011], explain how much progress we have made in that paper towards designing a model of empathy for artificial agents, and outline what remains to be done (Section 3.2). Then we introduce some terminology that will be used frequently in the rest of the thesis (Section 3.4). Finally, we draw an outline of our strategy for approaching the formulated research questions (Section 3.5).

3.1 The Problem Overview

The general direction of this research is to investigate whether and how empathy between artificial agents can improve the performance of their teamwork. The existing

computer science studies about empathy (as discussed in Section 2.1) have been mainly in the area of human-computer interaction, where computers are given the ability to empathize with human users in order to better communicate with them. Our work, however, is about modelling empathy within a team consisting entirely of artificial agents (with no human involved) and studying whether and how it can affect their teamwork performance.

Our initial motivation for investigating that problem comes from several sources. First, the studies of emotions and empathy in living systems in psychology have established the positive impacts of emotional mechanisms on decision making under time-constrained conditions (e.g., [Damasio, 1995]). As empathy can trigger an individual to take actions in favour of others, it has the potential to be considered as a mechanism for triggering help in a team. It has been confirmed by experience in human teamwork and documented in studies such as [Luca and Tarricone, 2001] that empathic help can improve team performance. Furthermore, as artificial intelligence and robotics progress from laboratory exploration towards mainstream engineering practice, agent-oriented software engineering (AOSE) becomes a widely accepted paradigm that may succeed object-oriented software engineering (OOSE) as the dominant software development methodology. Finally, with the ascent of networking and distributed computing, the research focus is shifting from individual agents to multiagent systems in general and agent teamwork in particular. Therefore, the mechanisms for facilitating helpful behaviour among a team of agents, possibly based on suitably defined empathic concepts, which could lead to higher team performance, merit more study and research.

We have already raised the question about the possibility of improving the performance of artificial agent teamwork by using empathy in [Polajnar et al., 2011]). In that paper we have outlined the basics of a model of empathy in a team of artificial

agents. In the next two sections, we present some of the main results of that paper, point out some of their limitations and formulate the questions to be explored in the rest of this thesis.

3.2 The Initial Modelling Steps

The model we introduced in [Polajnar et al., 2011] is inspired by the human empathy model proposed in [Goubert et al., 2005], which considers top-down and bottom-up influences in the formation of affective response, and analyses the transition from affective response to behavioural response. Our adaptation of that model is based on the well-known Belief-Desire-Intention (BDI) model introduced by Bratman [1987], its enhancement to include emotion to form an emotional BDI (EBDI) model as introduced by Jiang et al. [2007], and our own general mechanism for EBDI agents to perform empathic behavioural response. Figure 3.1 shows our adaptation.

In [Polajnar et al., 2011] we have discussed the possibility of applying such notions of human empathy, with suitable modifications, to artificial agents. However, we have not identified the specific factors involved in the formation of top-down and bottom-up influences in artificial agents, nor the specific mechanisms involved in the formation of affective response.

In that paper, we have also examined the Perception-Action Model (PAM) of human empathy introduced in [Preston and de Waal, 2002] and discussed how it can be connected to the Belief-Desire-Intention (BDI) model of agent reasoning [Bratman, 1987]. We have formulated the Empathic Behavioural Response Algorithm (EBRA) to model empathic responses in a team of artificial agents. In relation to different components of the BDI model, our model proposes offering help at five different levels: beliefs, desires, intentions, plans, and execution. The analysis does not include

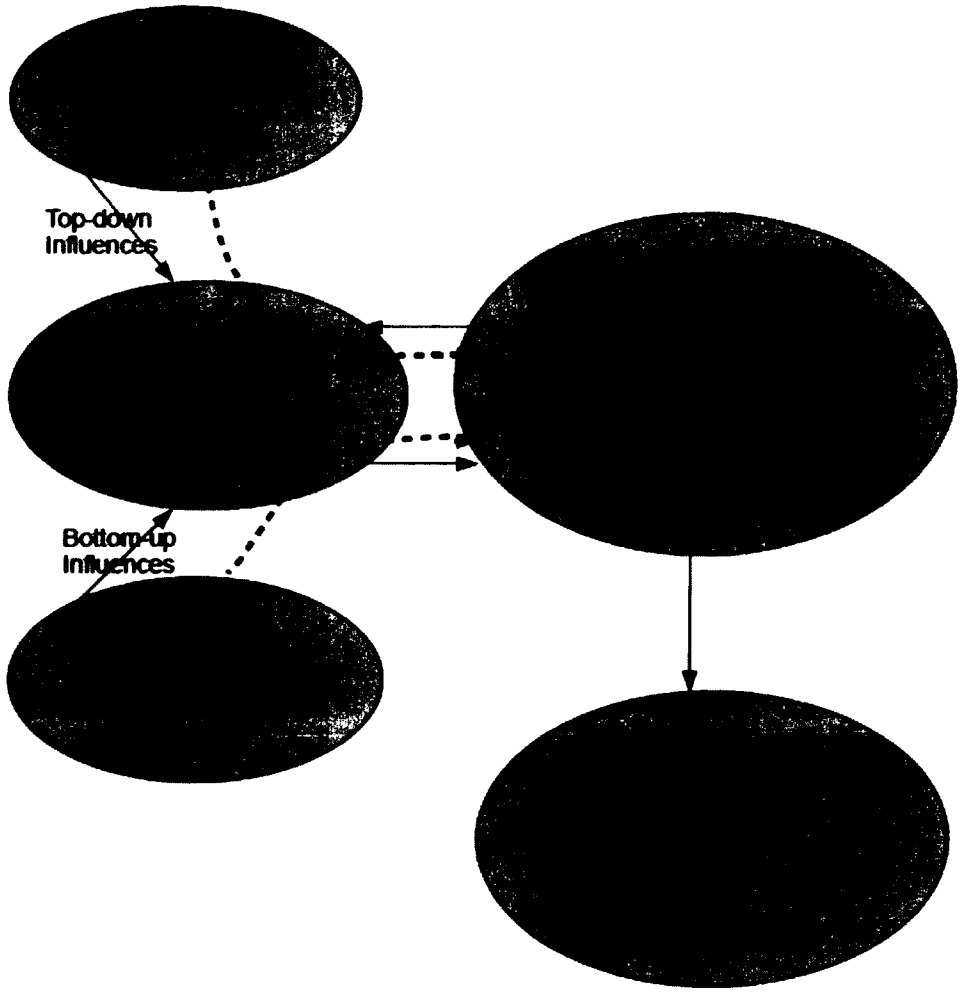

Figure 3.1: Empathy concepts of [Goubert et al., 2005] adapted to BDI agents (Reproduced from [Polajnar et al., 2011])

the specifics of how the affective response is formed, or when and how it triggers each particular level of behavioural response.

Figure 3.2 illustrates the Empathic Behavioural Response Algorithm (EBRA) introduced in [Polajnar et al., 2011]. This algorithm includes direct communication of emotional state representations between the subject and the object. The agents communicate by sending messages. The presented pseudocode uses communication primitives in the style of Communicating Sequential Processes [Hoare, 1985]. The communication operation *O!value* sends *value* to process *O* (the object agent in this

case). The communication operation *O?variable* receives a value from process O and puts it into *variable*. These primitives are synchronous, meaning that the sender process is blocked until the receiver process receives and vice versa. We introduce the * operator to indicate asynchronous execution, meaning that the receiving agent can do other work while waiting and gets interrupted when the message arrives. Following the established standards in agent communication languages (e.g., [FIPA, 2002]), each message includes a *performative* field that indicates the type of speech act it contains. Message with the *express* performative contains an emotional state, while an *inform* message contains a BDI component such as belief, desire, intention, plan, or agent task[1].

The algorithm proceeds as follows. In lines 1–3, Subject initializes its own emotions, beliefs, and intentions. It then performs its own deliberations, not shown here, until it receives a message from Object in line 4. The message expresses a negative emotional state E^- and indicates the task T that causes the concern. In lines 5–7, Subject stores the communicated emotional state, forms the empathy *Emp* based on emotional states of both agents, and updates its own emotional state with the empathy component. Subject now has emotional state that is in part identical to the state of Object; it can now proceed to derive affective and then behavioral responses.

The affective responses correspond to the levels of desire to help, represented by the hierarchy of predicates *B-level*, *D-level*, *I-level*, π*-level*, and *T-level*, whose values depend on the empathy *Emp*, as well as on the understanding of task T. In the current version of the algorithm, true predicates always form an initial segment of

[1]The performatives in agent communication languages (ACLs) are based on Searle's classification of speech acts (see, e.g., [Wooldridge, 2009]). The Searle's category of speech acts that expresses emotional state of the speaker has not been reflected in the performative sets of standard ACLs (such as the one specified by the Foundation for Intelligent Physical Agents [FIPA, 2002]). Our paper introduces the *express* performative for direct communication of emotions between artificial agents.

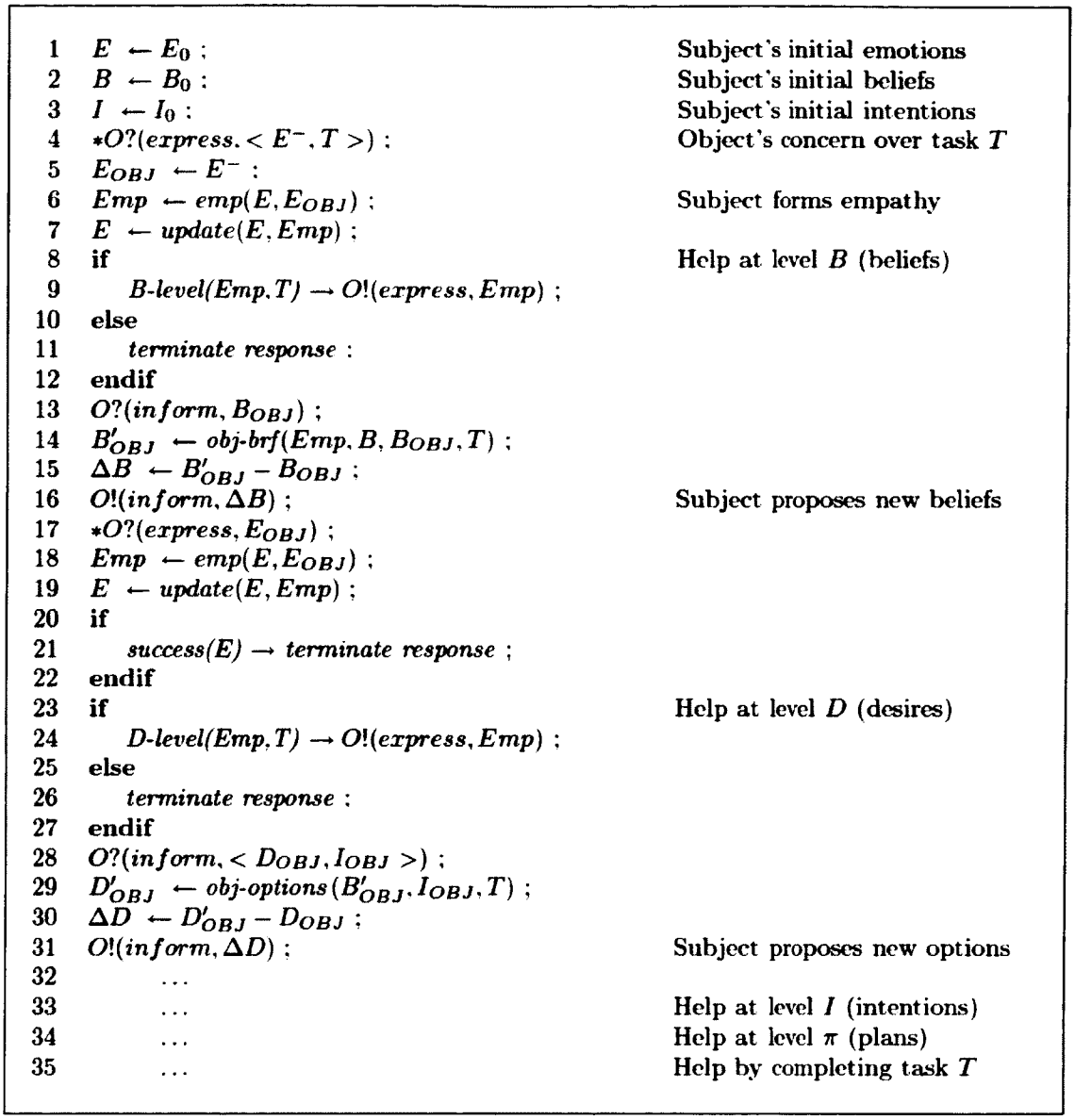

Figure 3.2: The empathic behavioural response algorithm [Polajnar et al., 2011]

this predicate sequence. For instance, if exactly the first two predicates are true, then Subject wishes to help by suggesting new beliefs and options, but will not engage in deliberations to produce intentions. This reflects the idea (inspired by the PAM view of empathy) that Subject and Object pass through the same steps of the practical reasoning process. In practice, if Object can reliably identify the critical step, Subject may skip the preliminaries. For instance, a robot may only need help in lifting a heavy

object in order to execute an otherwise valid and feasible plan.

Lines 8–16 show in detail how the affective response at *B-level* leads to the corresponding behavioral response. If *B-level* is true, an expression of empathy is communicated to O; if not, the behavioral response is terminated. In response, O informs about its beliefs relevant to task T (line 13). Subject then forms its own view of beliefs relevant to O's task (line 14), and informs O about the beliefs that were not in O's belief set (line 16). At this point, Subject switches back to its own work, while Object may try to take advantage of new beliefs and solve the problem.

In lines 17–22, Subject (asynchronously) receives and processes an expression of emotional state from O, which reflects the outcome of the Subject's attempt to help. Subject then updates its own emotional state. If the assistance had been successful, the response is terminated.

The next stage of computation takes place only if Subject wishes to engage in help at the level of proposing options (desires) to Object. If that is the case, Subject sends to Object an expression of empathy (otherwise it terminates the response) in lines 23–27. Object responds by providing the set of its current desires and intentions (line 28). Subject then generates Object's options on its own (line 29). Note that this computation is not fully independent in that Subject uses the intention set supplied by Object, rather than relying on Subject's own deliberations. The Subject sends the generated options that were not in the Object's original set of desires back to Object (line 31), completing its help at the level of desires.

The remaining stages are similar to the ones described so far. Subject first awaits the emotional response from Object. If the outcome was not successful, it next examines whether it wishes to assist with intention generation. If so, Subject deliberates on Object's behalf and communicates back to Object any new intentions that were

not in the received intention set. The handling of plans is similar. Finally, Subject can take over and try to achieve Object's task T.

A general observation on the algorithm is that it reflects the perception-action model of empathy (PAM). The PAM stipulates that Subject forms an emotional state that is similar to the state of Object. In the algorithm this is achieved in a straightforward manner by directly communicating the emotional state of Object and using it to update the emotional state of Subject. The algorithm integrates the hierarchy of affective and behavioural responses and formally represents the interactions between the subject and object of empathy for artificial agents. This provides a basis for theoretical and practical studies of the question raised in the title of the paper, namely whether empathy between artificial agents improves agent teamwork.

In EBRA, as it is shown in the Figure 3.2, the focus is mainly on the interactions leading to behavioural responses; there is no clear representation of affective response, or explanation of how it is formed and how exactly it leads to a behavioural response. In the rest of this chapter, we explain our approach to the modelling of empathic affective response for artificial agents.

3.3 Our Aim in the Rest of this Thesis

Our aim in the current thesis is to study the possibility of using empathy, and affective response in particular, as a trigger for initiating helpful behaviour in a team of artificial agents. One of the advantages of empathy in human teamwork is that it facilitates helpful behaviour among team members, as discussed in Section 2.2. In order to investigate whether a similar advantage can be achieved in artificial agent teamwork, one needs to address four questions. First, can helpful behaviour in a team of artificial agents improve their teamwork performance? Second, what exactly

constitutes empathic affective response in artificial agents and how does that notion fundamentally differ from other mechanisms for triggering help? Third, is the formulated notion of empathy in artificial agents an adequate trigger for helpful behaviour that can improve the performance of the team? Fourth, given that there are already other proposed mechanisms for initiating helpful behaviour in an agent team, why do we need empathic mechanisms to trigger help?

Regarding the first question, a number of authors argue and demonstrate that helpful behaviour has the potential to improve team performance. Cao et al. [2005] present a formal model for proactive assistance among agents in an agent team. Based on their model, agents can dynamically identify if other agents need help and they can provide help by performing a set of actions. Their experiments demonstrate that a team of agents with proactive help behaviour achieves a better performance compared to a team of agents without it. Kamar et al. [2009] propose a decision-theoretic model in which agents make rational decisions about offering help. Their mechanism has a set of rules for reasoning about the cost of help actions as well as their utilities. Their results of experimenting with that mechanism on the Coloured Trails game [Gal et al., 2010] indicate the improvement of team outcomes based on help. Nalbandyan [2011] and Polajnar et al. [2012] introduce a bid-based protocol specifically designed for offering help in teams of artificial agents; the simulations based on a modified version of the Coloured Trails game show that this leads to superior performance compared to the same teamwork scenario with no helpful behaviour.

In [Polajnar et al., 2011] we have also discussed the potential impact of (empathy-triggered) helpful behaviour on agent team performance. Through simple simulations we have illustrated how an artificial agent team, in which agents are capable of mutual assistance, under some circumstances performs better than a team without help mechanisms.

In our experiment, a team of agents $A_1, ..., A_n$ is in charge of processing, at a minimum cost, a sequence of m events of types $\alpha_1, ..., \alpha_n$ that occur in an environment Env. An agent is qualified to process events of type α_i if it has the capability C_i. An agent needs a fixed time quantum q to process an event for which it is qualified, and a much larger fixed time quantum u if unqualified. We assume that each capability has a fixed cost c per unit of time, as in physically embodied agents involving equipment amortization. Each agent A_i perceives and processes the events in its own segment Env_i of the environment (disjoint from other segments). The assumption at the time of team design is that the events occurring in Env_i belong to the same type α_i, with occasional exceptions. The rate of exceptions is modelled by the disturbance probability d, which is not known in advance. The type of each event arising in Env_i is chosen with probability $1 - d$ to be α_i, and with probability d to be a uniformly random pick from the set of all event types. The total processing cost is calculated as the time required by the team to process the entire event sequence, multiplied by the sum of all capability costs per unit time. Our simulation experiment compares three possible static designs of agent roles for this system.

In the "minimalist" design D_{spec}, each agent A_i has the capability set $\{C_i\}$. This is a cheap design with full specialization of agent roles that should work well for low disturbance levels. In contrast, the "maximalist" design D_{univ} has universal agent roles, with each agent A_i having the set of all capabilities $\{C_1, ..., C_n\}$. The capabilities of D_{univ} are n times as costly, but it handles every disturbance level with the maximum effectiveness. Both designs are simple in that they require no coordination among team members. The third design D_{emp} has the same specialized agent roles as D_{spec}, but the agents are prepared to help their teammates when a mismatch occurs. When an agent A_i is about to process a mismatched event of type $\alpha_j, j \neq i$, it will inform the specialist A_j about the expected processing time for its current workload, and A_j

will agree to take over the event if its own expected workload remains lower than A_i's. D_{emp} has the same capability cost as D_{spec}, but requires an additional time quantum γ for each coordination message.

Figure 3.3 presents the results of our simulation given the following configuration: $m = 300, n = 4, c = 0.5, q = 4, u = 54$, and $\gamma = 2$. The diagram shows that, for low to medium disturbance, mutually assisting agents outperform both competitors.

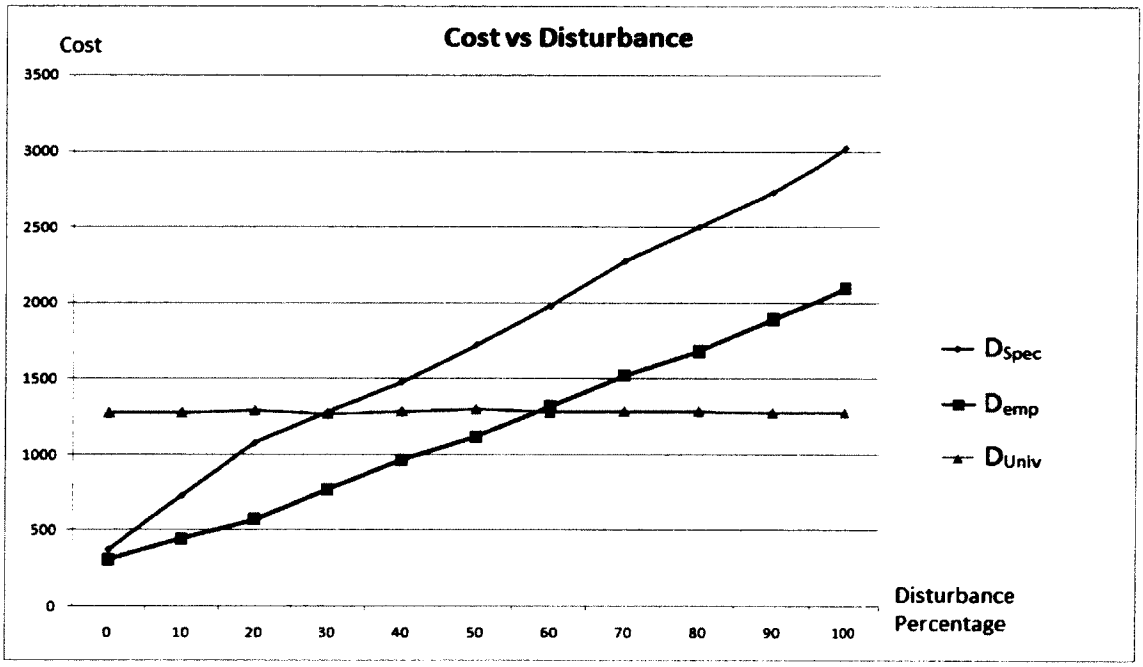

Figure 3.3: The relative cost-effectiveness of the three designs of agent roles in the team as the level of random disturbance in the environment behaviour varies. Role specialization with mutual assistance (D_{emp}) outperforms pure role specialization (D_{spec}) everywhere, and universal roles (D_{univ}) for low-to-medium random disturbances.

The experiment demonstrates how direct mutual assistance between team members can benefit the performance of the team as a whole. However, one should note that, while D_{emp} is intended to represent a team with empathic help, its help decisions are in fact triggered by a rational mechanism: the agents compare their workloads to decide whether the help act would benefit the team. This decision method was

adopted because, at that stage in our research, there was no available model of a genuine empathy trigger for help, and a rational bilateral decision mechanism had to be employed as an approximation. Rational bilateral decision mechanisms for mutual assistance have since been explored by Nalbandyan [2011] and Polajnar et al. [2012], while the genuinely empathic decision mechanisms are explored in this thesis.

While the investigation of helpful behaviour continues, the cited results provide sufficient evidence that helpful behaviour can improve the team performance, which establishes the motivation for our further research. We do not address this issue in the rest of the thesis.

In order to address the second question, which concerns the nature of empathy in artificial agents, we need a model of empathic affective response that is concrete enough to let us derive a formal mechanism for deciding about the behavioural response. Such a model does not exist at present. The modelling framework in [Polajnar et al., 2011] suggests the study of existing models of empathy in the living world, in particular the PAM, as a starting point. The intent is that the resulting empathy-like concepts should lead to an alternative mechanism for triggering helpful behaviour in artificial agent teams that indeed differs fundamentally from the mechanisms based on the calculation of the team utility value of help act, as described in [Kamar et al., 2009] and [Nalbandyan, 2011].

The third question is a critical test for any model of affective response in artificial agents that is introduced with the intent of improving team performance through empathy between team members. In order to be validated as a potentially useful direction for further study, the model does not have to uniformly improve performance, but needs to lead to improvement in some circumstances. Assuming that it does, the next objective is to characterize those circumstances and study how they affect the

improvement levels. Related objective involves the optimization of various parameters of the model relative to the properties of the agent team and the environment in which it operates. This type of study is a necessary prerequisite for comparing the effectiveness of empathy as a help trigger to the triggers based on rational maximization of the team utility value.

The fourth question, concerning the usefulness of empathy as a trigger for help compared to other existing mechanisms for the same purpose, requires further investigation. An essential part of such investigation is to find out whether, and under which circumstances, empathy provides a better mechanism for deciding if a help act should occur than the mechanisms that rely on calculation of team utility values. The empathy-based help mechanism does not need to be generally superior in order to be useful. Indeed, the analogy with human teams leads us to explore if it could be effective in combination with utility-based mechanisms and help overcome some of their limitations.

Analogies with human empathy suggest that there are circumstances under which empathy is a superior trigger for initiating helpful behaviour. As we discussed in Section 2.2, Damasio [1995] argues about the role of emotions in human decision making as a short-cut to a more limited decision space in which logic is being used. We expect that in the context of deliberation on whether to offer help, empathy could perform that role. In time-constrained situations when it is not practical to calculate the expected benefit of a help act, empathy could provide a faster trigger for help.

With respect to the four research questions formulated above, the focus of this thesis is on investigating questions two and three. We assume that previous and ongoing research adequately address question one, by providing enough evidence about the positive impact of helpful behaviour on the performance of agent teams, including

teams of artificial agents, in problem-solving tasks such as the one represented by the Coloured Trails game that forms the basis of subsequent investigations we report in this thesis. The fourth question, concerning the possible superiority of empathy as help trigger remains the primary longer-term objective in this direction of research. It requires a stable and well-tested model of affective response in artificial agents, as well as a degree of optimization of such a model with respect to its effectiveness as help trigger in particular models of agent teams and their operating environment. This optimization step is particularly important because the performance impact is a manifestation of emergent behaviour resulting from the presence of empathy, as opposed to the direct targeting of performance objectives by the utility-based triggers.

In order to study these questions in context, one also needs to identify the properties of artificial agents, their teamwork, and the environment in which the team operates, that make empathic behaviour possible and practically effective. This motivates the development of a comprehensive model of a team of artificial agents, endowed with a suitable notion of empathy along with the contextual properties that it requires, situated in an environment where helpful behaviour triggered by empathic and other mechanisms produces measurable effects. A conceptual framework for such modelling has been provided in [Polajnar et al., 2011], but many of its aspects remain to be defined, studied, and elaborated.

The focus of the current thesis belongs to the general direction outlined above, but has a narrower scope, that fits within the limitations of an MSc research topic. Rather than attempting to create a comprehensive model of agent team and its operating environment that could be varied with respect to many possible properties, we suitably restrict the scope of our modelling task in Chapter 4. In our experimental studies in Chapter 5, we situate our model of empathy into a specific microworld, designed to capture some relevant properties in a simple but still representative form, and then

we perform simulation experiments in that context.

The microworld is based on a variation of the Coloured Trails game [Gal et al., 2010]. The game itself was specifically designed to provide a test-bed for agent interactions; its variation used in our research group so far (e.g., in [Nalbandyan, 2011], [Polajnar et al., 2012]) has been adapted for the study of interaction protocols for helpful behaviour in agent teams; it is expected that a further adaptation will be needed in order to provide a microworld for the study of empathic artificial agents and their teamwork performance[2]. In the game, a team of agents cooperatively address a problem-solving task, in which agents pursue individual goals and keep individual scores, but can also directly assist each other. The agents move on a board of coloured squares, and their individual abilities determine the cost, in terms of resource points, of moving to a neighbouring square of a specific colour. The exact scoring rules are described in Chapter 5. The objective of the game is to maximize the team score, which is the sum of individual scores.

A final comment is in order in characterizing the proposed direction of this research. Our model of empathy in artificial agents is inspired by psychology and refers to the research literature in that discipline for motivating analogies. However, the purpose of our modelling is not to mimic natural empathy in the design of artificial agents, but to formulate an empathy-like concept for artificial agents that could effectively trigger their mutual help when appropriate, with the objective of improving their performance in practical tasks in certain types of situations. Apart from this difference in the objectives, there are two additional differences between studies in psychology and our own research which compel us to more carefully qualify the analogies between natural empathy as observed and analysed in psychology research

[2]The simulator is being developed by my colleague Omid Alemi, who has made it available for our experiments. Its detailed description is to appear in [Alemi, 2012].

and artificial empathy as explored in this thesis.

First, our study focusses on agent *teamwork*; it examines whether and how empathic interactions between individuals can enhance the performance of the team as a whole in practical problem solving. The same question is meaningful for a human team, for instance in the context of an engineering project. Subjective experiences of project participants suggest a positive impact, and some experimental studies support the same conclusion [Luca and Tarricone, 2001]. The fact that empathy has developed through evolution of the living world [de Waal, 2005] also suggests possible benefits to the species when facing practical problems of survival. However, the prevailing emphasis of experimental studies of empathy in psychology remains on the nature of individual interactions rather than qualitative and quantitative aspects of their collective impact.

Second, experimental psychologists typically observe interactions of live subjects, such as human or animal adults or infants, in situations involving pain or distress, that may be stimulated by using, for example, electric shock [Masserman et al., 1964] or tape-recorded crying of a person [Martin and Clark, 1982]. We contend that practical problem-solving activities can involve distress (e.g., when facing challenges in one's studies or work), and provide a context for empathic response, in which the subject offers practical assistance to the object. Human situations of this type provide close analogies for artificial agent interactions that we intend to explore (and are indeed reflected rather literally in the EBRA algorithm). However, they remain outside of the scope of most experimental studies of empathic behaviour in psychology.

In summary, our research is motivated by analogies with empathy in human teamwork directed towards practical problem solving, while most of scientific knowledge about natural empathy is derived from experimentation in other, significantly different

contexts. This discrepancy obliges us to particular caution, in addition to the usual concerns about reasoning based on analogies between natural and artificial systems.

3.4 A Note on Terminology

In the rest of this thesis, we frequently use the terms "empathic" and "rational" to refer to different types of agents in our experimentation. Therefore, we need to clarify what we mean by each of these terms before we discuss further modelling and experimentation.

In general, agents in this thesis perform their tasks by executing plans derived through rational deliberation and planning, motivated by the interest of the team. However, when it comes to offering direct help to teammates, outside of the general team organization, they rely on mechanisms that can be classified as either *empathic help* or *rational help*. Since our focus is on the study of helpful behaviour, we use the term *empathic agent* to describe an agent with rational deliberation about its mainstream activities and empathic help, and the term *rational agent* to describe an agent with rational deliberation and rational help. Unless specified otherwise, in the rest of this thesis our teams are composed of agents of the same type, and we extend the terminology to speak about *empathic teams* vs *rational teams*.

3.5 The Solution Strategy

In this section we outline our general strategy for approaching the remaining research problems in this thesis. It relies on a gradual development of a line of models, as summarized in Figure 3.4.

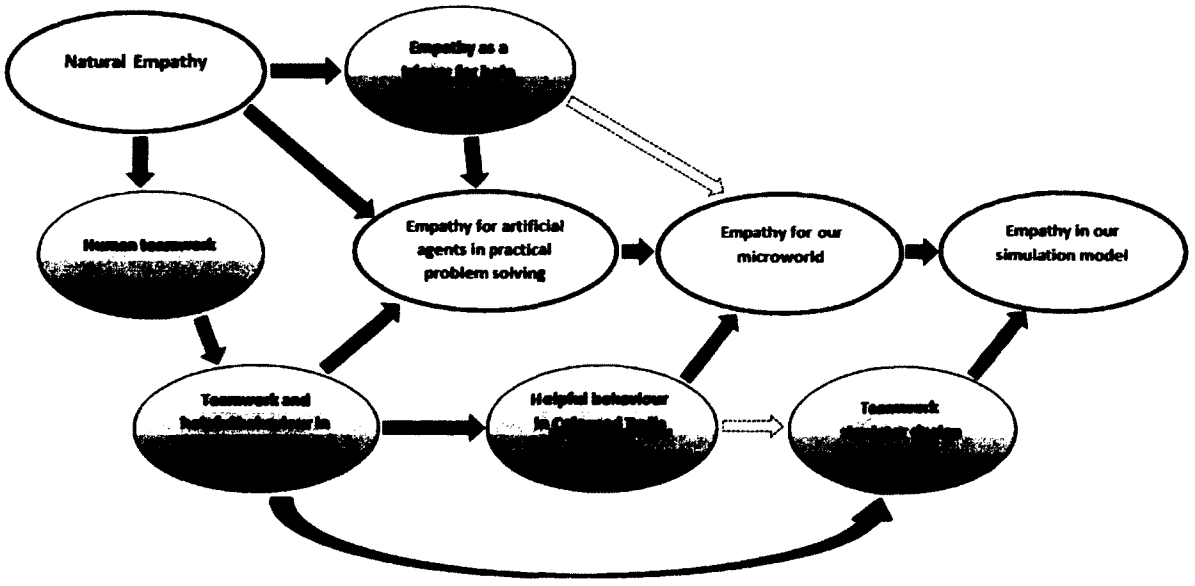

Figure 3.4: The general solution strategy

Our aim is to create a model of empathy for artificial agent teamwork. The practical motivation is to study the possible usefulness of empathy as a trigger for help in a team of artificial agents that jointly work on a problem-solving task. We adopt the concept of empathy from psychology (which studies empathy in humans and animals) and, after creating an analogous concept for artificial agents, we proceed to experimentally study its impact on the performance of an agent team. Therefore, we need to find a way for the transition from studying *natural empathy* to *measuring the performance of a team of artificial agents endowed with empathy.* Figure 3.4 shows the required steps for making such a transition. In Chapter 4 we explain in detail how we practically implement those steps.

In Figure 3.4, the white ovals represent the main steps that we take in the development of key concepts, starting with natural empathy and ending with empathy in our simulation model; the grey ones represent other studies and information that we use in order to be able to move forward in our direction of research. Solid arrows show direct influences, and dotted arrows show implicit influences between items.

In order to be able to practically measure the performance of a team of empathic agents we mainly need two things: first, a suitable design of artificial agents endowed with empathic mechanisms, and second, a test-bed in which we can situate a team of our empathic agents into a suitable environment, involve them in some practical problem-solving tasks, vary the relevant parameters, and measure the performance of the team. We describe the details of the modelling of our empathic agents and the test-bed we use for experimentation in Chapter 4. The experiment setups, results, and interpretations are given in Chapter 5.

Chapter 4

A Model of Empathy between Artificial Agents

In this chapter we develop a model of empathy between artificial agents. As a first step, we clarify the scope of our work and specify our exact targets (Section 4.1). Then we formulate the empathy factors for artificial agents, based on an analysis of the known empathy factors in living systems (Section 4.2). Finally we discuss and formulate the mechanisms for the formation of affective response and the triggering of behavioural response (Section 4.3).

4.1 The Scope of Modelling

The design of a complete model of empathy for artificial agents involves several mutually related decisions:

1. One needs to identify the *empathy factors* that influence the empathic behaviour in artificial agents; they may be either analogous to the factors that are known to influence natural empathy, or specifically designed for artificial agents without

analogies in the natural world.

2. One needs to define a *combining mechanism* by which the influences of individual empathy factors are aggregated to form the affective response, and the *triggering criteria* that determine when a behavioural response occurs and what it is.

3. One needs to identify the *properties of agent design* that are required as support for the defined notion of empathy.

4. One needs to determine the required *properties of the environment* in which the agents operate that make the empathic behaviour meaningful.

In order to make this modelling task feasible in our context, we proceed as follows: we restrict our domain of study to agent teamwork aimed at practical problem solving; we restrict our goals to the improvement of performance of such teamwork; and we design our empathy factors as direct analogues of some of the natural empathy factors. These restrictions are used to construct a simple but representative model of empathic agent team that operates in a specifically designed microworld environment. Our quantitative studies are then based on simulation experiments performed only in the microworld context.

In order to address the first task formulated above, we have investigated the literature in psychology that identifies different natural empathy factors. As we restrict our domain of study to artificial agent teamwork in practical problem solving tasks, we need to create concepts similar to natural empathy factors but specifically designed for artificial agents involved in such tasks, and decide how to represent those concepts in the microworld context. Our selection, design, and representation of empathy factors for artificial agents are partly driven by the intuitive expectations that they may contribute to the goal of improving the performance of the team.

In regard to the design of a combining mechanism that determines how different empathy factors participate in the formation of affective response, it should be noted that our agents are supposed to use empathy as a means for making decisions about performing help acts. We should design a function that maps the numerical values representing the strengths of the individual empathy factors into a numerical value representing the strength of affective response. After that, we need a threshold which determines how strong that affective response needs to be in order to make the agent perform a behavioural response. The behavioural response in our domain is a help act in favour of another agent's contribution to the team's objectives.

The design of empathic agents must meet a set of specific requirements that make the selected empathy factors meaningful. The agent design must include the mechanisms for observation and introspection that allow the agent to accurately determine the strength of each empathy factor in the concrete problem-solving context, and to correctly map them into the affective response. For example, if the familiarity between the subject and the object is an empathy factor, then the agents must be able to retain information related to their prior interactions. Furthermore, the possible helpful behaviours for agents must be clearly designed as the behavioural responses. The agent design must also support a suitable protocol for interactions between the subject and the object.

The environment in which empathic agents operate must meet some requirements as well. It must provide a setting in which agent teams can perform practical problem solving tasks with measurable outcomes. In order to enable the agents to request and offer help, the environment must provide the appropriate means for communication between agents. These are the minimum requirements that enable one to situate an empathic team in the environment and examine its performance. In addition, our aim is to model situations in which we expect that empathic help could lead to better

teamwork. To this end, we want the environment to allow us to generate and observe such situations, for example by controlling the level of dynamic change that impacts the predictability of the outcomes of specific plans and strategies. These requirements drive the design of the microworld that we use to determine if empathy-triggered help can improve artificial agent teamwork.

4.2 The Modelling of Empathy Factors

As reviewed in Section 2.2, Preston and de Waal [2002] and Preston [2007] name six parameters that influence affective response in humans and animals (where the distressed individual is known as object and the observer as subject): *depression* (whether or not the subject is self-distressed), *familiarity* (subject's former experience with the object), *similarity* (how similar the subject and the object are), *learning* (through implicit or explicit teaching), *past experience* (subject's former experience with a similar task), and *salience* (how strong the object's distress signal is).

In order to derive their analogues for artificial agents, we analyse each empathy factor from several points of view: its definition in psychology, its possible definition in artificial agent teams concerned with practical problem solving, its interpretation in the context of our microworld, its potential usefulness in improving the efficiency of an agent team, and the basic requirements that the agent and the environment designs must meet in order to allow for its meaningful implementation.

Emotional State[1]

The notion in psychology: Preston [2007] argues that "individuals with depression would have an empathy impairment due to an excessive focus

[1]In psychology literature, this item is often exemplified by the affective state of 'depression'. In this research, we use the word 'emotional state' in order to better relate the way we intend to use it in our model.

on the self, precluding the necessary interest in and attention to the state of the object". This implies that a distressed person is not very much motivated to show empathic behaviour to other distressed persons, as that person's main priority would be doing something for oneself.

The notion in artificial agent teamwork: In an artificial agent engaged in a problem solving task, emotional state can be considered as an internal value that reflects that agent's sense of personal progress. Depending on the exact context of the system in which the agent is acting, this sense of personal progress can represent how close the agent is to solving a problem, or what is the agent's situation compared to other agents in the system.

The notion in the microworld: Emotional state in our microworld can be considered as a value that is composed by different parameters: how far the agent is from the goal square, how many resources exist for the agent, the agent's relative progress compared to the rest of the team, etc.

Potential usefulness to the team: This factor is useful for preventing a lagging agent from spending its resources to attend other agents' subtasks or an advanced agent from staying idle or over-progressing in its own direction of work while it can help other agents.

Requirements: Agents must have personal emotional states that reflect their understanding of their personal progress within the team. Agents who are lagging behind their schedule *feel bad* while agents in normal situation *feel OK* and agents that are ahead of their work schedule *feel great.* This personal emotional state can affect their willingness for paying attention to the other agents' state.

Similarity

The notion in psychology: Preston and de Waal [2002] define similarity as "perceived overlap between subject and object, e.g., species, personality, age, gender". Martin and Clark [1982] study the empathic behaviour of human infants in response to the tape-recorded crying of different objects, and mention that infants keep crying in response to the crying of other infants, while they do not show any specific reaction to the crying of older children and chimpanzees.

The notion in artificial agent teamwork: In artificial agent teamwork, this is a value that reflects the degree of similarity between two agents. It can be based on their structures, their roles in the team, or the subtasks they are handling, for example.

The notion in the microworld: In the microworld we mainly determine the level of similarity regarding two agents' capabilities. The question that arises here is that whether it is a good idea to have an agent helping a similar agent. One could say that if an agent is stuck in some subtask, probably a similar agent would be stuck in it too, and it takes a *different* agent to overcome the problem. While the full investigation about this remains to be done as a part of the thesis, another idea is that in some situations, an agent may be able to handle most parts of the subtask but cannot finish it because of a small issue. An agent that is similar enough to this agent but is also capable of handling that tricky part is more likely to be helpful, compared to an agent that is totally different and may face issues in some other parts of the subtask that could be done by the first agent.

Potential usefulness to the team: a subject that has a similar structure, is handling similar tasks, has the same goal, or has any other kind of

similarity, may have the potential to be more helpful. Just like a student can help another student in school work much better than a soccer player can do. When two individuals are similar, their representation of the situation is more similar as well.

Requirements: Agents need to have parameters by which they identify themselves and compare themselves to other agents in order to determine how similar they are. Such parameters can be, for instance, capabilities, roles, etc.

Familiarity

The notion in psychology: Preston and de Waal [2002] define familiarity as the subject's previous experience with the object. Stinson and Ickes [1992] summarize their experimental results by concluding that "male friends were found to be more accurate than male strangers in inferring each other's thoughts and feelings". They mention more interaction and information exchange, more similar personalities and more detailed knowledge about each other as the primary reasons for this fact.

The notion in artificial agent teamwork: Familiarity is defined regarding two agents in the team and it is a value that reflects the level of prior interactions between them.

The notion in the microworld: In the current version of our microworld the interactions between agents are limited to requesting and offering help. These interactions do not seem sophisticated enough to reflect the potential usefulness of familiarity between agents. Therefore, taking advantage of this parameter seems to require a more complex test-bed.

Potential usefulness to the team: This factor is useful because former

experiences, help the subject to know the object more and better, and subsequently makes it more able to help. In fact this former experience helps the subject to diagnose the object's problem much faster and propose a suitable solution easier, by limiting the solution space and having less trial and error approach.

Requirements: Agents need to store a history of their interactions with other agents. For each interaction, they need to save different information, such as object's identity, object's problem, agent's owns response, and the successful solution (if exists).

Teaching[2]

The notion in psychology: Preston and de Waal [2002] define this as implicit or explicit teaching.

The notion in artificial agent teamwork: Teaching consists of implicit and explicit teaching. In implicit teaching, agents learn with time and it affects their further decisions (based on their observations of past events, decisions and their consequences). In explicit teaching, some specific decisions for some specific situations can be hard-wired in the agents' decision making mechanism.

The notion in the microworld: In our microworld agents may learn from the consequence of their helpful actions in past games. For example, if their helpful action has led to team success, their motivation for help in the next games will be increased.

[2]In psychology literature, this item is named "learning". However, as the word "learning" in artificial intelligence refers to techniques such as machine learning, here we use the word "teaching" to prevent confusing the two concepts in psychology and artificial intelligence.

Potential usefulness to the team: In case of explicit teaching, this can be useful by having some helpful actions implemented in the design time in an agent. In case of implicit teaching, this can be useful when agents learn along the time about some specific situations in which it is better to intervene, even if other empathy parameters do not encourage it.

Requirements: For teaching, especially implicit teaching, agents need to be capable of learning. They need to store information about the interactions and the impact of their behaviour on the situations. Later on, they must be able to reason about the result of their intervening (or not intervening) and the advantages or disadvantages of it.

Past Experience

The notion in psychology: Preston and de Waal [2002] define this factor as former experience with situation of distress. Masserman et al. [1964] experimentally studied the reactions of monkeys in a situation where they need to give an electrical shock to their con-specifics in order to obtain food. They mention that monkeys with the past experience of being shocked were more likely to accept self-starving instead of shocking other monkeys and getting food.

The notion in artificial agent teamwork: Past experience is defined in relation to the agent and the action it is about to do. It is a value that reflects the frequency of prior experience with the same or a similar subtask.

The notion in the microworld: In our microworld past experience can be defined in terms of an agent's former moves to squares with specific colours. The number of times an agent has moved to, for example, green squares, is considered as the amount of that agent's experience with the action of moving to a green square.

Potential usefulness to the team: Past experience with a similar problem helps an agent to have a better understanding of the current problem and leads to finding a solution faster, as the agent has already deliberated about that subtask or a similar one before. Therefore, it may lead to less resource consumption for reasoning about the same or a similar situation.

Requirements: There must be parameters by which agents can identify subtasks. If such parameters exist, then an agent can compare the current subtask with another subtask it has had past experience with, and determine how similar they are. Such parameters could be course of actions for handling a specific subtask, etc. Furthermore, agents must be able to gain experience with time, which enables them to spend less resources on a subtask they have faced previously.

Salience

The notion in psychology: Preston and de Waal [2002] define salience as "strength of perceptual signal, e.g., louder, closer, more realistic, etc." Sagi and Hoffman [1976] summarize their experiment results on human infants by mentioning that "infants exposed to the newborn cry cried significantly more often than those exposed to silence and those exposed to a synthetic newborn cry of the same intensity."

The notion in artificial agent teamwork: Salience is a value that reflects the level of an agent's distress with a subtask when it is asking for help. Since artificial agents in a team may have identical emotional structures, and can use messages with the *express* performative to communicate emotions, they can obtain information about the emotional states of others through communication rather than perception.

The notion in the microworld: In the microworld, for an agent that is requesting help, salience can be measured based on the agent's remaining resources, its distance to the goal square, its relative progress compared to the rest of the team, etc.

Potential usefulness to the team: Salience is useful in determining how urgently a distressed object needs help. It is also useful in prioritizing the helpful actions when there are different distressed object's around that are all asking for help. Subject can distinguish which object is more distressed (and is accordingly in a worse situation).

Requirements: There must be parameters by which an agent as object can specify the level of distress. Based on such parameters, other agents can determine the strength of distress signal. Besides, a reliable communication channel is needed for the object to inform the subject(s) about its level of distress.

It should be noted that each of these empathy factors would require further study with respect to its possible implementation in a particular context. Based on our general analysis so far, we regard them all as potentially useful components of a model of empathy for artificial agents, but as we will see in the next chapter, we practically model only a subset of them in our microworld.

Our analysis has also identified the required properties of the agents and their environment that are needed in order to support each empathy factor. The basic microworld model may need to be modified and enhanced depending on the specific selection of empathy factors that one decides to implement.

The above analysis provides the basic answers to questions (1), (3), and (4) formulated in Section 4.1. The remaining question (2) is addressed next.

4.3 The Modelling of Affective and Behavioural Responses

In this section we consider how the selected empathy factors participate in the formation of the affective response, and how the affective response leads to a behavioural response. This is a research question that merits additional study and requires further analysis of the empathy factors and their mutual relationships.

The implementation of empathy factors in a concrete multiagent system is related to the class of problem-solving tasks that the system is intended to solve; we have seen instances of this when adapting the general empathy factors to the microworld context in the last section. As a consequence, the combining mechanism for the formation of affective response and the triggering criteria for behavioural response will also be problem-specific to a degree. However, we contend that in the case of empathy such specificity is lower than in the case of calculation of team utilities employed by the rational decision methods. An empathic decision should in general depend less on the fine problem-solving details, compared to a rational decision. The empathic decision mechanisms should be comparatively more general and less computationally complex. These effects should be more pronounced in systems of realistic complexity than in our highly simplified microworld.

For our purposes in this thesis we shall adopt a relatively straight forward approach in which the combining mechanism for the formation of affective response is the weighted average, and the criterion for triggering the behavioural response is the comparison to a fixed threshold. Formally, let $A_1, ..., A_n, n > 1$ be a team of agents and let $F_{i1}, ..., F_{ik_i}$ be non-negative real numbers representing the strength of the K_i selected empathy factors as perceived by the subject agent A_i. The strength of the *affective response* $\mathcal{A}_i$ of the agent A_i is then determined as

$$\mathcal{A}_i = \frac{\Sigma_{j=1}^{k_i} W_{ij} F_{ij}}{\Sigma_{j=1}^{k_i} W_{ij}} \tag{4.1}$$

where the weights W_{ij} are positive reals. The *behavioural response*, which in our case is a help act, is triggered whenever

$$\mathcal{A}_i > \theta_i \tag{4.2}$$

where θ_i is a non-negative real constant representing the triggering threshold.

In general, the formal model above allows each agent A_i to have its own selection of empathy factors, its own weights representing how its affective response is formed, and its own threshold representing its general level of empathic sensitivity. Noting that empathic interactions in teams composed of heterogeneous empathic agents represent an intriguing research topic for future work, we restrict our studies in this thesis to homogeneous teams in which all members have identical empathic properties. Accordingly, we use the same values $W_1, ..., W_k$, and θ for every agent in the team.

The next step is to select the concrete values of $W_1, ..., W_k$, and θ in specific context. Given our objective of enhancing the team performance, these values should be determined through an optimization process that seeks to maximize the team performance in the context of a concrete environment and problem-solving task. Since empathy in living systems has been formed by natural evolution, it is intuitively appealing to apply evolutionary optimization techniques for this purpose. In Chapter 5, we use genetic algorithms [Mitchell, 1998], with the fitness values represented by the team performance scores as measured in simulation experiments.

As a final observation, we note that the empathy model developed in this chapter is not fundamentally restricted to teamwork situations, but could be employed to implement empathic interactions between artificial agents in multiagent systems in general.

Chapter 5

Experiments and Results

In this chapter we describe the simulation experiments that we have performed and analyse their results. We introduce a very basic implementation of empathic agents in a highly simplified microworld, which still provides enough structure to enable us to demonstrate that our notion of empathy is a valid help trigger in artificial agent teamwork, and to provide an abstract comparison between empathic and rational help in the presence of varying disturbance in the environment. First of all, in Section 5.1 we review the preliminaries and terminology about the teamwork simulator and the microworld we use, and describe the implementation of empathic agents. In Section 5.2, we then use genetic algorithms to determine suitable values of the weights of the individual empathy factors in the formation of affective response and the triggering threshold of the behavioural response. Section 5.3 demonstrates that, for low to moderate disturbance in the environment, empathic help leads to better team performance than random help, which shows that, even in a very simple simulation model, our notion of empathy is a valid trigger for mutual help in the team. In Section 5.4 we compare the performance of a team consisting of empathic agents vs. a team consisting of agents that rely on rationally motivated mutual assistance, as the disturbance in the environment and the cost of rational decision vary. Finally,

in Section 5.5 we analyse the significance of the results in a wider context and discuss some of their implications for future work.

5.1 The Simulation Environment

Before we proceed to discuss our experiments and their results, we need to clarify our terminology, describe the structure of the experiments and the environment in which they are performed, specify the experimental conditions, and present the implementation of empathic agents.

5.1.1 The Teamwork Simulator

Our experiments use a teamwork simulator developed by my colleague, Mr. Omid Alemi [Alemi, 2012], which allows us to simulate in parallel the behaviour of several agent teams that operate in identical microworld environment configurations.

In simulation experiments the agents are located in an environment in which they try to reach individual goals in order to complete the task assigned to the team. The team task is formulated as a game in which agents individually score points, but have the objective of maximizing the total team score. The game proceeds in discrete *rounds* in which each agent can make a single move. Between successive rounds the agents in the team can exchange any number of messages in order to coordinate their actions. This communication is simulated as a sequence of synchronous communication cycles, each consisting of a *send* phase and a *receive* phase.

After a number of rounds, all of the agents will either achieve their goals or run out of resources, and also exhaust the possibility of progress through mutual assistance; at that point the *match* is completed. A specified number of matches, one after

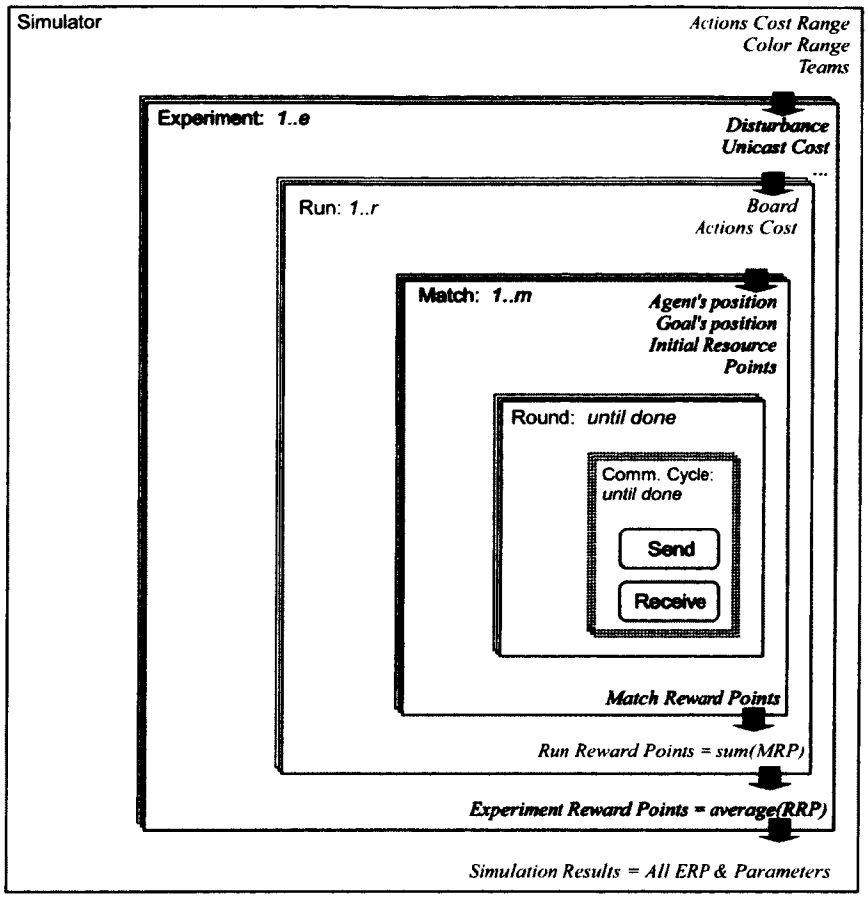

Figure 5.1: The structure of experiments in Teamwork Simulator

another, during which agents may retain memory of prior matches, constitute a *run*. Finally, an *experiment* consists of a number of runs (determined by the statistical aspects of experiment design) where we average the results over those runs to have a more reliable and accurate analysis. The structure of experiments in the teamwork simulator is depicted in Figure 5.1. Some of the terminology in the figure is specific to the microworld and explained in the next subsection.

5.1.2 The Microworld Configuration

The microworld used in this thesis consists of the core design presented in [Polajnar et al., 2012] and slight extensions required by empathic agents. In that paper, the microworld was used for experiments with rational teams. Leaving out the components that are not relevant to our work, we quote the relevant descriptions directly from that paper.

> "The test bed for simulation experiments is a variation of the Coloured Trails game [Gal et al., 2010]. It has been developed specifically for the study of helpful behaviour in teamwork and implemented independently. The players are software agents $A_1, \ldots, A_n$, $n > 1$, situated on a rectangular board divided into coloured squares. The game proceeds in synchronous *rounds*. Each agent can move to a neighbouring square in each round. Each move represents the execution of an action. The types of actions $\alpha_1, \ldots, \alpha_m$ are represented by the available colors, and their costs to individual agents by the $n \times m$ matrix *cost* of positive integer values. "

The game proceeds as follows:

> "At the start of the game, each agent A_i is assigned its initial location on the board, a unique goal with a specified location and amount g_i of *reward points*, and a budget $r_i = d_i a$ of *resource points*, where d_i is the shortest distance (i.e., number of squares) from the agent's initial location to its goal, and a a positive integer constant. Whenever A_i moves to a field of color α_j, it pays $cost_{ij}$ from its resource budget; if the budget is insufficient, the agent is blocked. Each agent chooses its own path to the goal, which represents the choice of its own local plan. The paths can

intersect; it is legal for multiple agents to be on the same square at the same time. The game ends when no agent can make a move (because it has either reached the goal or lacks the resources)."

The team's objective in the game is to maximize the team score, which is computed as follows:

"All agents remain in the game until the end, when their *individual scores* are calculated as follows: if A_i has reached the goal, its score is the goal achievement reward g_i plus any remaining resource points (as a savings bonus); if A_i has failed to reach the goal, its score is $d_i'a'$, where d_i' is the number of moves A_i has completed, and a' is a positive integer constant representing the reward for each move. The *team score* is the sum of all individual scores."

The level of random change in the simulated environment is controlled by an additional parameter:

"As a representation of environment dynamics, the colour of any square can be replaced, after each round, by a uniformly random choice from the color set. The change occurs with a fixed probability D, called the *level of disturbance.*"

Helpful behaviour is modelled as follows:

"In this presentation, the model includes only *action help.* The requester A_i faces a move to a square of color α_k, charged at $cost_{ik}$; if A_j agrees to help, A_i moves at no cost to itself, with the $cost_{jk}$ charged to A_j. Protocol interactions involve explicit computation and communication costs, and

> the help act has a fixed overhead cost. While the specific decision criteria and protocols for help transactions may vary, the general intent of such transactions is to advance the performance of the team as represented by the team score."

Empathic agents keep records of the actions they have taken in each run. They memorize how many times they have performed a specific action and they use that information later when deciding about a help act involving that action. We explain the details in Section 5.1.3

The settings for our experiments are also similar to the settings presented in [Polajnar et al., 2012]:

> "We simulate eight-agent teams on a 10×10 board with six colours. Each goal reward is 2000 points. The cost vector for each agent includes three high-cost entries, randomly chosen from the set $\{300, 400, 450, 500\}$, and three low-cost entries from $\{10, 40, 70, 100\}$. Thus each agent's capabilities are high for three colors, and low for the other three. The threshold cost of next action that triggers help deliberation is 300. The reward for accomplishing each step on the chosen path is 100 reward points; The initial allocation of resources for each step towards the goal is 200 points. The overhead cost of a help act is 30 points."

For every experiment, we record the team scores for each of the teams, averaged over 3000 simulation runs. In every team, at the beginning of the game each agent selects the lowest-cost path (based on the initial board state) among all of the shortest paths to its goal square, and commits to it for the rest of the game.

5.1.3 The Implementation of Empathic Agents

We shall model empathic agents based on three empathy factors that we have adapted from the study of empathy is psychology: *emotional state*, *past experience*, and *salience*. This subset is chosen based on the potential usefulness (as discussed in Section 4.2) and also the feasibility of implementing them within the context of our microworld. In this section we describe how these empathy factors are implemented in the microworld and how empathic agents use them to make a decision when they are asked for help.

We model the *emotional state* of an empathic agent as follows:

$$Emotional\ State = \frac{Remaining\ Resource\ Points}{Estimated\ Cost\ of\ the\ Remaining\ Path} \tag{5.1}$$

The *Estimated Cost of the Remaining Path* is always a positive number, except when the goal has been reached; at that point, *Emotional State* is defined to equal a fixed constant (1000 in our experiments).

As agents perform different actions and gain experience, the resource points they need to spend for a specific action decrease to some extent. After some point, gaining experience does not reduce the cost any longer. Later, when an agent is asked for help regarding a specific action, it takes its level of experience with that action into account in order to determine its willingness to help.

The influence of past experience is modelled as follows. For each agent A_i the cost of action α decreases by δ each time A_i performs α, until the cost of α reaches a given floor. In the experiments, $\delta = 20$ and the floor value is 40.

We model salience as:

$$Salience = \frac{1}{Emotional\ State} \tag{5.2}$$

When *emotional state* is non-zero; otherwise *Salience* equals a fixed constant (1000).

Each help request initiated by an agent contains some information: the agent's identity, the current action for which the agent needs help, and the agent's "salience". When another agent receives a help request, it will use the information in the help request message to determine the strength of its affective response. In order to distinguish this particular implementation from the general concept of affective response, we instead use the term *willingness to help (WTH).*

After receiving a help request, an agent computes its own level of willingness to help (WTH) based on the received salience, its own emotional state, and its own past experience with the requested action. Based on the Formula 4.1, WTH is computed as:

$$WTH = \frac{W_e E + W_s S + W_p P}{W_e + W_s + W_p} \tag{5.3}$$

where E is emotional state, P is past experience, and S is salience, while W_e, W_p, and W_s are their respective weights. The weights are positive reals ($W_e, W_p, W_s > 0$).

Whenever a decision has to be made, WTH is compared to a suitably chosen threshold in order to determine if the affective response is strong enough to lead to a help act.

A help request is broadcast to all team members. Also, an agent may receive multiple requests for help in the same round of the game. The protocol for resolving these conflicts is outlined next.

The empathic agents in our microworld work based on a bidding-like system. In each round of the game the following scenario happens:

1. Agents calculate their emotional state based on the remaining path to the goal and their remaining resource points.

2. Agents decide whether or not they need to ask for help, based on their emotional state.

3. Those that need help broadcast a help request.

4. Those that receive help requests will ignore them if they themselves need help; otherwise, they will calculate their "willingness to help" (WTH), which represents the strength of affective response for each help request based on its empathy factors.

5. Those that have calculated their WTH for help requests from different agents, will choose the request with highest WTH that exceeds threshold, and offer help to the corresponding agent.

6. Those that receive help offers will accept the offer with the highest WTH.

7. Agents proceed with performing their actions (moving to a neighbouring square, helping another agent, or just doing nothing, depending on the criteria).

5.2 Optimizing Performance of Empathic Team

We have chosen to optimize the influence level of empathic parameters using Genetic Algorithms. Genetic Algorithms (GA) is a heuristic technique in artificial intelligence that is useful for finding solutions for optimization and search problems; it simulates the process of natural evolution. We briefly review the technique and our reasons for using it in Section 5.2.1. In the optimization process we use the Matlab GA Toolbox connected to our teamwork simulator, as explained in Section 5.2.2.

5.2.1 The Role of Genetic Algorithms in Our Research

In this section we give a brief description of genetic algorithms and then we discuss our reasons for choosing GA as the appropriate optimization technique regarding the nature of our problem. Characterizing various aspects of our problem, we can explain how genetic algorithms can be an suitable candidate to be used for optimizing the influence level of empathy parameters.

In GA, each candidate solution for a problem is called a "chromosome" and it is represented as a string of values. Those values can be of type bits, integers, real numbers, etc. There is a function called the "fitness function" which can evaluate how good or bad a candidate solution is. The process of optimizing a function using genetic algorithm starts with creating a random population of candidate solutions (*initialization*). In each generation, every individual within a population is evaluated by the fitness function and then the *good* ones are chosen to reproduce the next generation of solutions (*selection*). Each new generation is used to produce the following generation; this *reproduction* is carried out using the *crossover* and *mutation* operators (Crossover is a process of taking more than one parent chromosome and producing a child chromosome from them. Mutation alters one or more gene values from its initial state within one chromosome, to maintain genetic diversity). This process goes on until either the specified satisfactory solution has been found or there has been a specific number of consecutive "stall generations" that have produced no progress (*Termination*).

A problem can be optimized using GA if we can represent its possible solutions as chromosomes along with suitable notions of crossover and mutation, and if we can define a fitness function that is able to evaluate those solutions. The approach appears to be very compatible with the structure of our problem, for several reasons.

First, we can simply represent a candidate solution as a chromosome that contains four positive real numbers, representing the weights of empathy factors and the threshold; $< W_E, W_P, W_S, \theta >$. We can look at teamwork simulation as the fitness function which takes a candidate solution, applies it in an experiment, and returns the team score that represents how good that solution is.

Second, genetic algorithms are very useful for problems that have a vast solution space (like our problem in optimizing empathy parameters). The reason for this advantage is that, unlike most optimization techniques that search from a single point, a genetic algorithm starts with a whole population of solution candidates [Busetti, 2000].

Third, genetic algorithms are very suitable for approaching problems for which we have no idea where to start! The problem of optimizing empathy factors in our microworld seems to be one of them, as we do not have any basis for speculating about a *good* candidate solution. The GA approach is also helpful in such cases because it does not need to know anything about the problem it is going to solve. It starts with generating many random solutions as a population, theoretically keeping its door open to all the different solutions.

Fourth, the proper use of mutation helps a genetic algorithm to avoid getting trapped in a local minimum which makes it superior to gradient methods, for instance. However, in this thesis we primarily use the method to find a "good enough" solution that will let us validate the use of empathy as a help trigger. For concrete quantitative comparisons with other help triggers we would need to conduct the optimization process with special care towards avoiding a local minimum.

5.2.2 Optimizing the Effects of Empathic Help

Our aim is to determine the values of the weights of the empathy factors and the triggering threshold, $< W_E, W_P, W_S, \theta >$, which together lead to optimum performance of the empathic agent team.

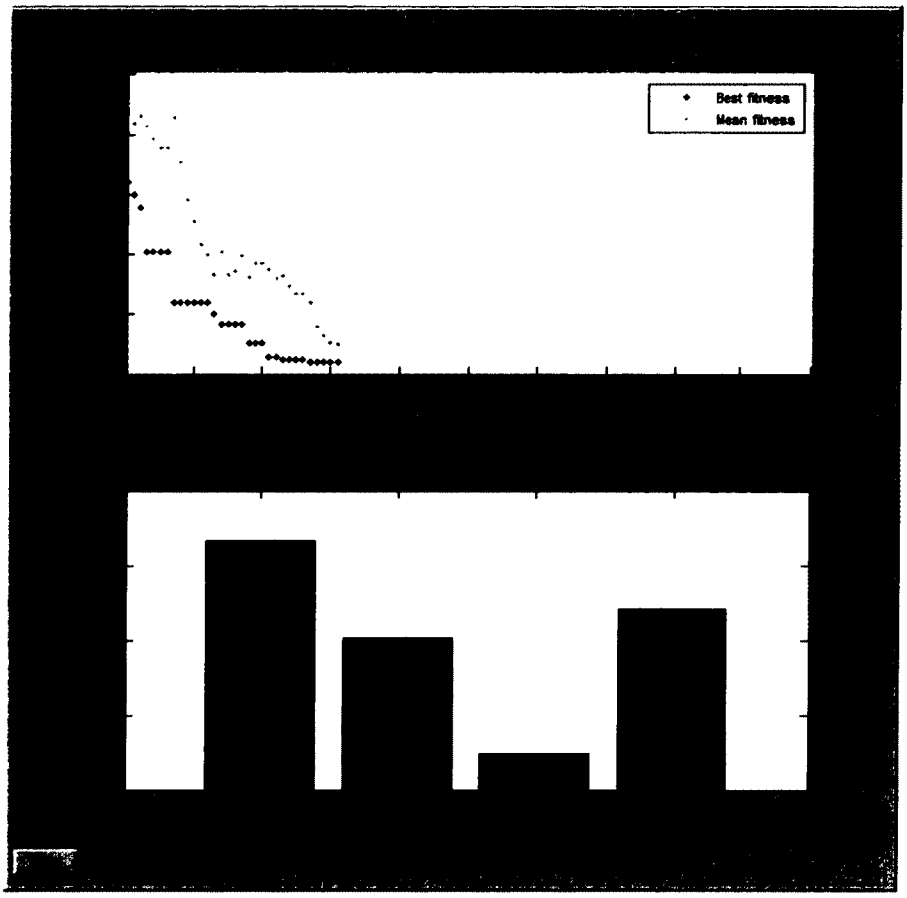

Figure 5.2: The genetic algorithm optimization of $< W_E, W_P, W_S, \theta >$

We have used the GA Toolbox in Matlab together with the teamwork simulator for performing the optimization. The GA Toolbox in Matlab provides a comprehensive and flexible set of features and tools for optimizing a function. In our case, we have used the simulator as the *fitness function* that takes the chromosome values from Matlab, performs the simulations and returns the negative value of the team score as the fitness value. (The negative value is returned because the optimization method

always seeks the minimum.) Matlab then reads this value, creates a new generation of chromosomes, and sends them to the simulator. This cycle continues until Matlab finds out which values produce the best results.

The results of our optimization of the values $< W_E, W_P, W_S, \theta >$ are shown in Figure 5.2. The process starts with 30 initial populations, uses the *rank* fitness scaling function, and the stochastic uniform selection function. The Matlab GA tool stops regenerating after facing 50 stall generations, as shown in Figure 5.2(a). The optimization process has produced the values $W_E = 1.012, W_P = 0.24, W_S = 1.208$, and $\theta = 1.668$ as shown in Figure 5.2(b).

Having optimized the values of the four variables, we can use them in our simulation experiments to study the behaviour of empathic agents under different circumstances.

5.3 The Validation of Empathy as a Help Trigger

Having designed our agents, structured our microworld, and situated the agents in the microworld, we can now examine the question of whether or not empathy is an eligible mechanism for triggering help in a team of artificial agents. We formulated this question in Chapter 3 as one of our main concerns in the current thesis.

We investigate the question experimentally, by comparing the performance of a team of empathic agents with the performance of a team in which agents provide help randomly. To make that experiment fair, we let agents in both teams use the same procedure when asking for help. When it comes to offering help, the agents in the second team make random decisions based on a fixed probability value P_{help}.

Figure 5.3 shows the results of series of experiments with different values of the

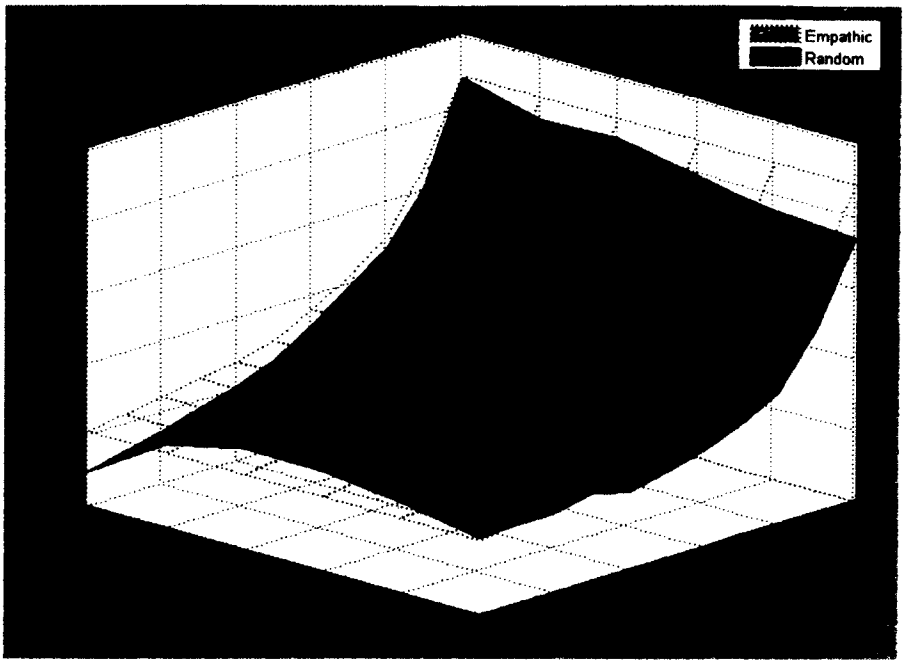

Figure 5.3: The performance of empathic team vs random-helping team

help probability P_{help} (Both the random help rate P_{help} and the disturbance d are represented as percentages). In every case the empathic agent team outperforms the random-helping team for low to moderate levels of disturbance in the environment. This demonstrates that, for low to moderate disturbance levels, empathy does provide a valid trigger for help. If the disturbance is so high that the behaviour of the environment becomes highly unpredictable, the impact of empathy factors becomes irrelevant and the bias they introduce apparently becomes counterproductive.

We can modify this experiment further to make it even more fair! We can first measure the *help act percentage* in the empathic agent team $R_{help}(d)$, defined as the ratio of help acts over help requests for different disturbance levels d (Figure 5.4), and then let the random-helping agents use the same help rates as probabilities of help, $P_{help}(d) = R_{help}(d)$. The resulting performance comparison between empathic agents, and our new 'guided random-helping agents' is shown in Figure 5.5. It reinforces our previous conclusion that for low to moderate disturbance levels empathy-triggered

help is better than random help, which means that empathy represents a valid help trigger in teams of artificial agents.

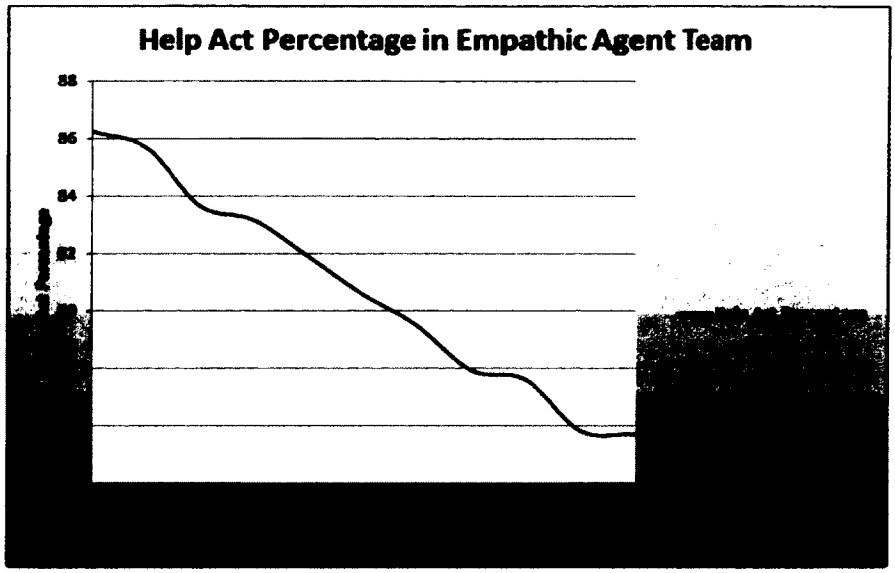

Figure 5.4: Help acts percentage in the empathic agent team

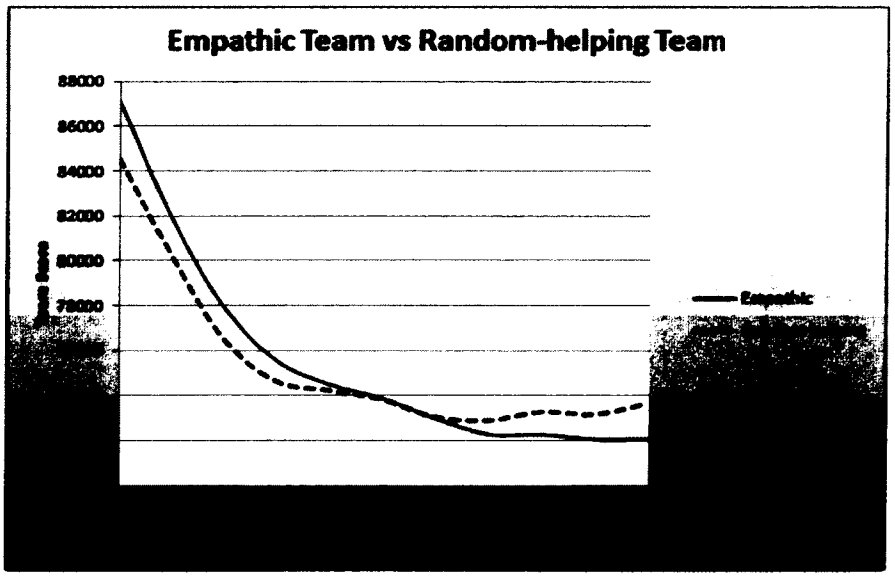

Figure 5.5: The performance of empathic team vs guided random-helping team

A final observation concerns the decrease in the observed help rate among empathic agents in Figure 5.4 as the disturbance grows. The explanation is based on two effects.

First, one of the empathy factors, namely the emotional state E of the subject, is negatively affected by the growing disturbance, and lowers the willingness to help as defined by Formula 5.3. Second, as the disturbance grows, the help requests become more frequent (as we illustrate later in Figure 5.9), which diminishes the capacity of the team to serve them with a high rate of acceptance.

5.4 A Comparison of Empathic and Rational Help

In this section we perform some experiments to compare the performance of empathic agents vs. agents endowed with a different help triggering mechanism. We formulated the question of whether or not empathy can be superior to rational help triggers in Chapter 3, and indicated that it requires further systematic investigation beyond the scope of this thesis. Such a systematic investigation is likely to require a test-bed in which rational deliberations about whether to help involve potentially high levels of computational complexity. This view is based on the conjecture that empathic mechanisms would outperform rational ones when the cost of rational decision is high. It would be interesting to investigate such trade-offs in problem-specific settings, with agent teams addressing concrete practically relevant tasks. Our current microworld model does not support such requirements.

The conjecture that empathy can provide a superior help trigger when rationally motivated help decisions are computationally complex is supported by the intuition coming from analogies with psychology, where emotions provide shortcuts to decisions in complex situations under time constraints. In this section we describe a simple experiment in support of this conjecture, situated in the microworld context. In the experiment, we compare the performance of a team of empathic agents versus a team of rational agents. The rational agent team is an *Action MAP* team as introduced and

studied in [Nalbandyan, 2011] and [Polajnar et al., 2012]. The agents in the team use a rational decision method, called the Mutual Assistance Protocol (MAP) to decide whether an agent should perform an *action* on behalf of another. Both the empathic team and the action map team use a bilateral interaction to decide whether a help act should take place and the nature of help is also the same (performing an action), which facilitates the comparisons. In addition, Action MAP teams had already been implemented and studied using the Team Simulator in the same basic microworld environment, and their implementations were made available for our experiments.

In an Action MAP agent team, a distributed joint decision is made about whether to perform a help act or not. The agent that needs help in performing an action α sends to other agents a message that contains its estimation of the *team benefit* if α is removed from its local plan. An agent that receives this message estimates the *team loss* if α is added to its local plan. The difference between the benefit and loss is the *net team impact*; if this value is positive, the recipient offers help to the requester. The requester then picks the help offer with the highest expected net team impact.

In the microworld implementation of an Action MAP team, we simulate the concept of decision *complexity* in a highly simplified way, by introducing a new parameter in our microworld, called the *rational decision cost*. This parameter is the cost associated with estimating the team benefit or team loss of a help act for rational agents. The more complex the decision space gets, the higher the rational decision cost is.

Again, for making the experiments as fair as possible, the procedure of asking for help is the same for both empathic and rational agents. They just differ in making decisions about performing help acts.

Figures 5.6 and 5.7 show the performance comparison of empathic help team against rational help team, for varying disturbance in the environment, with two

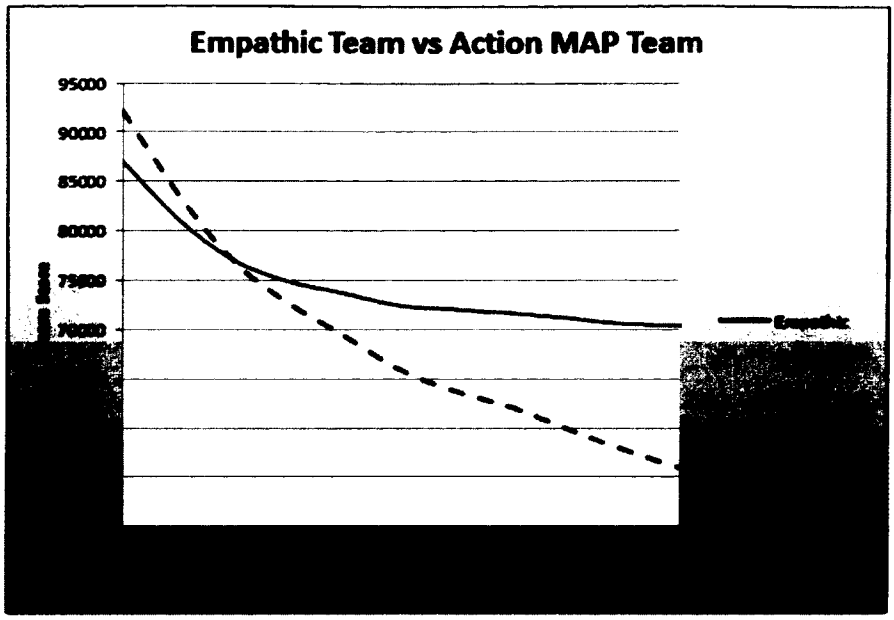

Figure 5.6: Empathic vs Action MAP team scores for high rational decision cost (40)

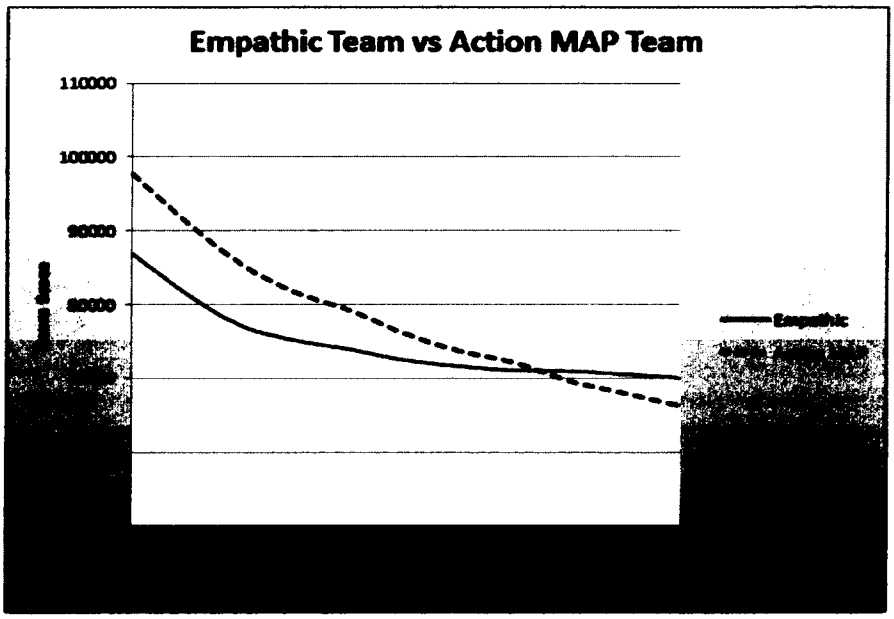

Figure 5.7: Empathic vs Action MAP team scores for low rational decision cost (20)

different values for rational decision cost: 20 and 40. As the rational decision cost increases, the performance of the agent team using rational help mechanisms decreases.

Figure 5.8 shows the performance comparison of empathic team vs Action MAP

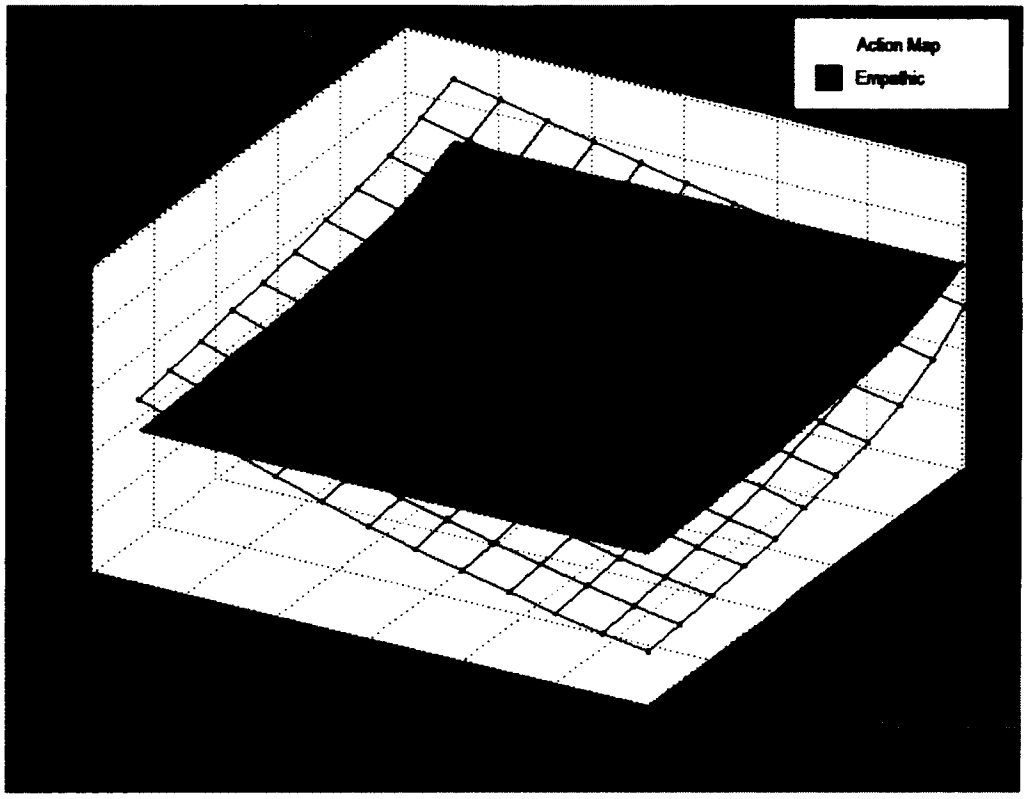

Figure 5.8: The performance of empathic team vs Action MAP team

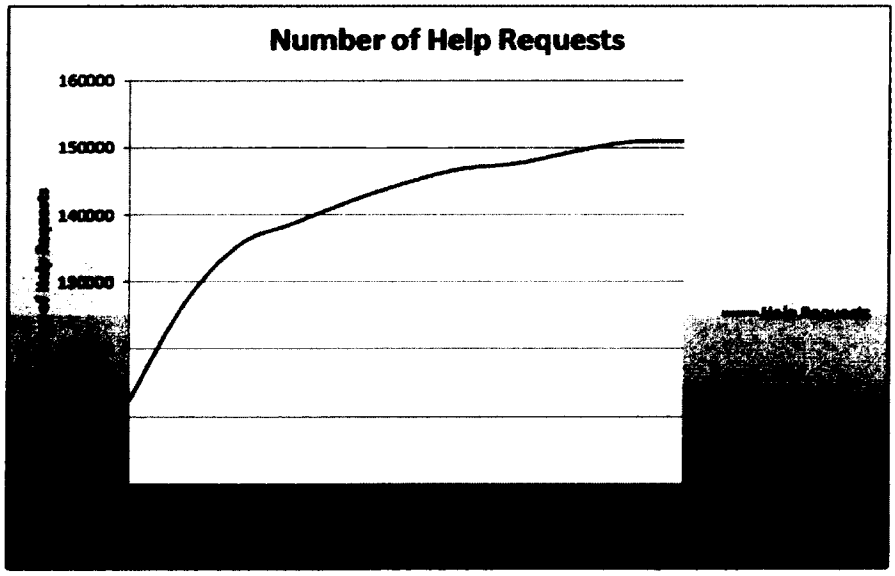

Figure 5.9: The number of help requests in a team

team in a three-dimensional graph where the disturbance and the rational decision cost vary.

The graph clearly shows that empathic agents perform better when the complexity increases in the system. For higher disturbance levels, empathic agents start to outperform rational agents faster, compared to the lower disturbance. This is because in higher disturbance levels, the number of help requests increases (as shown in figure 5.9) which itself makes the rational agents perform more calculations, as they need to analyse the presumable outcome of each request.

5.5 Analysis and Evaluation

Now, we go through the experiments we presented in this chapter once more and analyse the obtained results.

5.5.1 Empathy as a Trigger for Help

In Section 5.3 we observed that empathic agents perform better than various types of random helping agents as long as the disturbance in the environment is not too high. After disturbance reaches more than around fifty percent, random helping agents start to outperform empathic agents.

This crossover occurs because in high levels of disturbance, almost everything in the environment is happening on a random basis. It means that there is no logic behind the sequence of events in the surrounding world. And when there is no logic in the sequence of events, there cannot be any logic for facing those events either. No reasoning (whether rational or emotional) can predict anything when everything happens just randomly. Therefore, possibly the best strategy for facing a random sequence of events is acting at random.

However, for lower disturbance levels (where we have focused our attention), em-

pathic agents perform better. This result means that empathy, as modelled in our experiments, can be an acceptable mechanism for initiating help in a team of artificial agents.

5.5.2 Empathy vs. Rational Mechanisms

In Section 5.4 we compared the performance of empathic teams against Action MAP teams that are capable of making rational decisions about performing a help act. The results indicated that when the decision complexity (rational decision cost) goes up, empathic agents outperform rational agents, and when the decision complexity decreases, rational agents perform better. We also observed that for growing levels of complexity, empathic agents start to outperform rational agents sooner when the disturbance is higher.

Rational agents target team performance improvement as a direct goal when making decisions. On the other hand, empathic agents do not directly care about team performance when they need to make a decision about offering help. In an empathic team, performance improvement is an emergent behaviour, or in other words, a side effect of the agents' decisions. Therefore, it is expected that the accuracy of the decisions made by rational agents is higher than those made by empathic agents, regarding the outcome of a possible help act in relation to the team performance. However, the amount of resources the rational agents need to spend for calculating the outcome of an action depends on the complexity of the decision space; but such a dependency does not exist in empathic agents.

Having this analysis, it is fairly clear why empathic agents perform better in higher complexities and rational agents outperform them in lower complexities. When decision complexity is low, rational agents can make more efficient decisions by calculating

the outcome of their actions more accurately while not spending a lot of resources on it. But when the decision complexity increases, no matter how accurate their decisions are, the team costs may become prohibitive. In such cases, an empathic mechanism that does not directly deal with complexity seems to be a better choice.

5.5.3 Combining Rational and Empathic Help

In this subsection, we talk about the possibility of combining rational and emotional help mechanisms into a single mechanism that can take advantage of both of them. We also discuss the possible advantages of having agents of different types within one team. So far in our experiments, agents in a team are identical; meaning that, for example, the agents in a team are either all empathic or all rational. Also, in empathic teams, the weights of empathy factors and the threshold are the same for all agents; which means that all agents have the exact same type and level of empathy.

Observations of real teamwork examples in humans, however, confirm the fact that human teams normally consist of people with different personalities. Some individuals are more emotional, some are more rational, and those who are more emotional, are not necessarily equal in their type and level of emotions. But usually this diversity is needed for a human team in order to succeed. This observation can bring up the idea that artificial agent teams can also perform better if they consist of agents with different decision making mechanisms; or basically, different *personalities.*

In Section 5.4 we observed that rational and empathic decision making mechanisms, are in fact complementary mechanisms and we can not simply pick one of them as the *better* one. But depending on the level of complexity together with the level of disturbance in the environment, both rational and emotional mechanisms have the potential to outperform each other.

Those experimental results reinforce the idea that a team of *hybrid* agents, capable of making decisions based on both rational and emotional mechanisms, could achieve a higher performance compared to either empathic or rational teams. As we saw in 5.4, at low levels of complexity rational teams perform better, while after increasing the decision complexity, empathic teams achieve higher scores. In a very quick analysis, we can say that if we had a team of agents that are capable of making rational decisions when complexity is low, and empathic decisions when complexity is high, this team would probably gain a higher performance than both an exclusively rational team and an exclusively empathic team.

The points made in this section, generate a few different ideas and open some questions about having stronger agent teams that can possibly achieve a higher performance. First, would it improve the performance of an empathic agent team if the level of threshold or the empathy factors' weights varied for different agents? Second, would a team consisting of both pure empathic and pure rational agents, assuming that suitable protocols enable them to interact, perform better than a team consisting of identical agent types? Third, would a team consisting of agents capable of both empathic and rational decision making perform better than other types of teams mentioned above? And finally, what are the complexities of modelling such diverse agent teams, given that different agent types use different protocols?

Each one of these questions requires a lot of investigation and can be discussed as a separate thesis topic. However, analogies in the study of human teamwork suggest that modelling each of those speculative agent teams could lead to both a better understanding of the underlying phenomena and possibly better performance under some circumstances.

Chapter 6

Conclusions and Future Work

We have introduced a model of empathy as a basis for helpful behavior in teams consisting purely of artificial agents that collaborate on practical problem-solving tasks.

Guided by existing models of natural empathy in psychology and neuroscience, in particular the Perception-Action Model, we have identified the potential empathy factors for artificial agents, as well as the mechanisms by which they might produce affective and behavioral responses. The empathy model is fairly general and allows the agents in the team to have individual empathic profiles.

We have then investigated whether the performance of such teams can benefit from empathic help between members as the analogy with human teams might suggest.

For that purpose, we have situated a team of empathic agents, endowed with a simplified version of our general empathy model, and having identical empathy profiles, into a microworld similar to the Colored Trails game, developed within our research group to support studies of helpful behavior in agent teamwork, and examined the team's performance through simulation experiments. As a preliminary step, we have optimized the parameters of the empathy model using a genetic algorithm, with

the teamwork simulator providing the team performance score as the fitness value of each candidate solution.

The experiments show that, for low to moderate levels of disturbance in the environment, a team in which help decisions are based on empathy outperforms a team in which help decisions are random, even if we ensure that the overall rate of positive help decisions is the same for both teams. These results demonstrate that the empathic mechanisms defined in this thesis are valid triggers for helpful behavior that can improve the performance of an artificial agent team.

We have also performed experiments in which the performance of team with an empathic help mechanism is compared to a team with a rational help mechanism, based on the Mutual Assistance Protocol. Since rational help decisions in the relatively simple microworld do not involve deliberations of realistic complexity, the cost of a rational help decision is modelled as an independent parameter. This precludes realistic performance comparisons between empathic and rational help, but still allows the identification of some general trends. The experiments have shown that rational help is superior when the cost of rational decision is low, and is superseded by empathic help as the growing complexity of rational decisions leads to higher costs. That result is consistent with the study of natural emotions and empathy in psychology, which confirms the positive role of emotions in decision making when the decision space is too big and complex. The crossover happens sooner in the case of higher disturbance in the environment, suggesting that empathic help can be more effective than rational help in unpredictable circumstances.

The model of empathy introduced in this thesis complements and strengthens some of our earlier published results, which had provided a framework for the incorporation of empathy into artificial agent teamwork. In this thesis we have revisited those

results, and pointed out the need to complete them by developing a suitable model of empathy factors and the formation of empathic responses. Our empathy model in this thesis fills that void. One of the results that has been made more complete in this manner is the Empathic Behavioral Response Algorithm, which shows how BDI agents endowed with empathy can provide different levels of problem-solving help to each other, assisting at the level of beliefs, desires, intentions, plans, or executions.

Regarding possible future work in the direction of this research, several ideas are worth investigating.

In our discussion of experimental results we indicated that, in order to obtain realistic performance comparisons between empathic and rational teams, we need a test bed that practically models the real computational complexity. An interesting topic in that direction is to design a microworld that has that property and yet remains sufficiently simple for effective experimentation. Such a microworld test bed could allow one to draw more certain conclusions about the role of emotional (and in particular empathic) mechanisms when the cost associated with rational decisions becomes too high.

Another possible direction of future work is designing a microworld in which all of the empathy factors presented in this thesis can be modelled and practically implemented. The decison to choose only a subset of the empathy factors for experimental study was due in part to implementation feasibility issues in the current microworld. Since our general analysis indicates that all six empathy factors are potentially useful, it would be interesting to investigate whether by modelling the rest of those factors one can improve the efficiency of empathic agent teams.

In our experiments in this thesis, all of the empathic agents within a team have identical empathy profiles, as determined by the weights of the empathy factors and

the threshold representing the empathic sensitivity of the agent. However, since human teams consist of individuals with different types and levels of emotions, an appealing topic of research is to investigate whether having a similar emotional diversity within a team of empathic agents improves their performance, compared to a team of identical empathic agents. Since our general model of empathy already supports such diversity, the task mainly involves the design of a suitable simulation environment and experimentation strategy.

And finally, a very interesting topic for research is the idea of *hybrid agents* that we introduced in Subsection 5.5.3. Such agents that are capable of providing help both using empathic and rational decision mechanisms, seem to have the potential to outperform both pure empathic and pure rational agents. However, providing the sufficient infrastructure for situating hybrid agents and supporting their more complex communication protocols requires further research and study.

Bibliography

Huib Aldewereld, Wiebe van der Hoek, and John jules Meyer. Rational Teams: Logical Aspects of Multi-Agent Systems. Technical report, Utrecht University, 2004.

Omid Alemi. A framework for study and development of interaction protocols for agent teamwork. Master's thesis, University of Northern British Columbia, 2012. Forthcoming.

Michael Bratman. *Intention, Plans, and Practical Reason.* Harvard University Press, Cambridge Mass., 1987.

Jacek Brzeziński, Piotr Dunin-Kęplicz, and Barbara Dunin-Kęplicz. Collectively Cognitive Agents in Cooperative Teams. In *Engineering Societies in the Agents World V*, pages 191–208. Springer, 2005.

Franco Busetti. Genetic algorithms overview. `http://www.aiinfinance.com/gaweb.pdf`, 2000. Retrieved on April 15th, 2012.

Sen Cao, Richard A. Volz, Thomas R. Ioerger, and Michael S. Miller. On proactive helping behaviors in teamwork. In *International Conference on Arti?cial Intelligence*, pages 974–980, 2005.

Philip R. Cohen and Hector J. Levesque. Intention is Choice with Commitment. *Artificial Intelligence*, pages 213–261, 1990.

Philip R. Cohen and Hector J. Levesque. Teamwork. *Special Issue on Cognitive Science and Artificial Intelligence*, pages 487–512, 1991.

Antonio R. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain.* HarperCollins Canada, November 1995.

Frans de Waal. The evolution of empathy. `http://greatergood.berkeley.edu/article/item/the_evolution_of_empathy`, 2005. Retrieved on January 14th, 2012.

Egon den Broek. Empathic agent technology (EAT). In L. Johnson, D. Richards, E. Sklar, and U. Wilensky, editors, *Proceedings of the AAMAS-05 Agent-Based Systems for Human Learning (ABSHL) workshop*, pages 59–67, Utrecht - The Netherlands, 2005.

Barbara Maria Dunin-Keplicz and Rineke Verbrugge. *Teamwork in Multi-Agent Systems: A Formal Approach.* Wiley, 2010.

FIPA. *FIPA ACL Message Structure Specification.* Foundation for Intelligent Physical Agents, 2002. URL `http://www.fipa.org/specs/fipa00061/SC00061G.html`.

Ya'akov Gal, Barbara Grosz, Sarit Kraus, Avi Pfeffer, and Stuart Shieber. Agent decision-making in open mixed networks. *Artificial Intelligence*, 174(18):1460 – 1480, 2010.

Daniel Goleman. *Emotional Intelligence: Why It Can Matter More Than IQ.* Bantam, 1st edition, June 1997.

L Goubert, K D Craig, T Vervoort, S Morley, M J L Sullivan, A C de C Williams, A Cano, and G Crombez. Facing others in pain: The effects of empathy. *Pain*, 118 (3):285–288, December 2005. ISSN 0304-3959.

Barbara J. Grosz and Sarit Kraus. Collaborative Plans for Complex Group Action. *Artificial Intelligence*, 86:269–357, 1996.

C.A.R Hoare. *Communicating Sequential Processes.* Prentice-Hall International Series in Computer Science. Prentice Hall, 1985. ISBN 0-13-153271-5.

Hong Jiang, Jose M. Vidal, and Michael N. Huhns. EBDI: An architecture for emotional agents. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '07, pages 11:1–11:3, New York, NY, USA, 2007. ACM.

Ece Kamar, Ya'akov Gal, and Barbara J. Grosz. Incorporating helpful behavior into collaborative planning. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '09, pages 875–882, Richland, SC, 2009. International Foundation for Autonomous Agents and Multiagent Systems.

Hector Levesque, Philip Cohen, and Jose Nunes. On Acting Together. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 94–99, 1990.

Joe Luca and Pina Tarricone. Does emotional intelligence affect successful teamwork? In *Meeting at the Crossroads (Presented at the 18th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education)*, Melbourne, December 2001. Ascilite.

Grace B. Martin and Russell D. Clark. Distress crying in neonates: Species and peer specificity. *Developmental Psychology*, 18(1):3 – 9, 1982.

Jules Masserman, Stanley Wechkin, and William Terris. Altruisic behaviour in rhesus monkeys. *The American Journal of Psychiatry*, 121:584–585, December 1964.

Ernst Mayr. Cause and effect in biology. *Science*, 134(3489):1501 –1506, November 1961.

Zulfiqar Memon and Jan Treur. Designing social agents with empathic understanding. In *Proceedings of the 1st International Conference on Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems*, ICCCI '09, pages 279–293, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 978-3-642-04440-3.

Melanie Mitchell. *Introduction to Genetic Algorithms.* A Bradford Book, third printing edition, February 1998. ISBN 0262631857.

Ranjit Nair, Milind Tambe, and Stacy Marsella. The role of emotions in multiagent teamwork. In Michael A. Arbib Jean Marc Fellous, editor, *Who Needs Emotions? The Brain Meets the Robot.* Oxford University Press, 2004.

Narek Nalbandyan. A Mutual Assistance Protocol for Agent Teamwork. Master's thesis, The University of Northern British Columbia, Prince George, BC, September 2011.

Rosalind Picard. Affective computing. Technical report, MIT Media Laboratory Perceptual Computing Section No. 321, 1995.

Rosalind Picard. *Affective Computing.* The MIT Press, 1st edition, July 2000.

Jernej Polajnar, Behrooz Dalvandi, and Desanka Polajnar. Does empathy between artificial agents improve agent teamwork? In Y.Wang, A.Celikyilmaz, W.Kinsner, W.Pedrycz, H.Leung, and L.A.Zadeh, editors, *Proceedings of the 10th IEEE I.C. on Cognitive Informatics and Cognitive Computing [ICCI*CC'11]*, Banff, Alberta, Canada, August 2011. IEEE Computer Society.

Jernej Polajnar, Narek Nalbandyan, Omid Alemi, and Desanka Polajnar. An Interaction Protocol for Mutual Assistance in Agent Teamwork. In *Proceedings of the Sixth*

International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2012), 2012. Forthcoming.

Stephanie Preston. A perception-action model for empathy. In T. F. D. Farrow and P. W. R. Woodruff, editors, *Empathy in Mental Illness*, chapter 23, pages 428–446. Cambridge University Press, 2007.

Stephanie D Preston and Frans B M de Waal. Empathy: Its ultimate and proximate bases. *The Behavioral and Brain Sciences*, 25(1):1–20; discussion 20–71, February 2002.

Abraham Sagi and Martin Hoffman. Empathic distress in the newborn. *Developmental Psychology*, 12(2):175–176, 1976.

Bas Steunebrink, Mehdi Dastani, and John-Jules Meyer. Emotions to control agent deliberation. *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1*, pages 973–980, 2010. ACM ID: 1838337.

Linda Stinson and William Ickes. Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, 62(5): 787–797, May 1992.

Katia Sycara and Michael Lewis. Integrating Intelligent Agents into Human Teams. In *Team Cognition: Understanding the Factors that Drive Process and Performance.* American Psychological Association, 2004.

Michael Wooldridge. *An Introduction to MultiAgent Systems, 2nd Edition.* Wiley, 2009.

Michael Wooldridge and Nicholas Jennings. Formalizing the Cooperative Problem Solving Process. In *Proceedings of the Thirteenth International Workshop on Distributed Artificial Intelligence*, pages 403–417, 1994.

Michael Wooldridge and Nicholas R. Jennings. The Cooperative Problem Solving Process. *Logic and Computation*, 9(4):563–592, 1999.

Robert Wysocki, Robert Beck, and David Crane. *Effective Project Management: How to Plan, Manage, and Deliver Projects on Time and Within Budget.* John Wiley & Sons Inc, September 1995.