From Rational to Emotional Agents

by

Hong Jiang

Bachelor of Engineering
Wuhan University, 1997

Master of Science
Wuhan University, 2001

Master of Engineering
University of South Carolina, 2005

Submitted in Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy in the

Department of Computer Science and Engineering

College of Engineering and Computing

University of South Carolina

2007

Major Professor

Chairman, Examining Committee

Committee Member

Committee Member

Committee Member

Dean of The Graduate School

Dedication

To my family ...

for their love and support.

especially my husband Kunming, for taking care of our children such that I have time working on my research.

to my children: Alex and Esther, wonderful gifts from God, who can patiently and quietly listen to a paper instead of a story when they were still infants.

To my parents ...

who always encourage me and show me support when I was in the lurch.

Acknowledgments

First, I want to thank God for giving me chance to study here, helping me growing and giving me idea in research.

I would like to thank each of the members of my dissertation committee for their time, and commitment as well. And I would like to thank Dr. Matthews for giving me chance to teach and give me financial support for a long time, which helped keep me afloat.

I especially wish to acknowledge Dr. Josè Vidal and Dr. Michael Huhns. Dr. Vidal guided me through the detailed aspects in the dissertation, gave me financial support and encouraged me to attend valuable conferences, which stimulated my brain storm in research. Dr. Huhns provided the early inspiration for this work, and gave me continual encouragement and care. His good sense and knowledge in this area, and his serious attitude to research also teach me how to be a responsible scholar.

I appreciate Dr. Larry Stephens for a lot of help in my teaching, Dr. Marco Valtorta for pointing out typo problems in my proposal, and Dr. Joseph E Johnson for the financial support which helped my family got through a period of hard time.

Finally, I would be remiss if I did not recognize the encouragement and support I have received ever from the department, the faculties, the stuffs, and my students.

Abstract

To date, most research on multiagent systems has focused on rational utility-maximizing agents. However, theories show that emotions have a strong effect on human's physical states, motivations, beliefs, and desires. The details have not been explicated clearly so far. In artificial intelligence, emotions have begun to receive more attention, but mostly in human-robot/computer interaction. The research on applying emotions to agents' decision-making is still very limited.

Can agents be intelligent without emotions? We believe that, whether for human-like or non-human-like agents, the effect of emotions on decision-making cannot be ignored, since agents with high emotional quotients (EQs) can be built to have better performance in complex dynamic environments than purely rational agents.

This research focuses on the effects of emotions on decision-making. Taking into account the incompleteness of emotion theories and emotional differences among individuals, I describe EBDI, a common architecture for emotional agents, which specifies a separate emotion mechanism within an agent, instead of trying to model emotion mechanisms to reflect the reasoning process specifically, like most researchers have done. It reflects the practical reasoning process, and one can select and apply part of an emotion theory into the architecture as needed. Sample agents in Tileworld are presented and the results show that an EBDI agent can have better performance than traditional BDI agents.

To apply EBDI in negotiation, a plug-in is designed, which modifies the OCC model, a standard model for emotion synthesis, to generate emotions. Considering

the possibility of incorporating emotions into negotiation, I generate EWOD (Emotional Worth-Oriented Domain), which requires numerical emotions. Thus, a mapping from 22 OCC emotions to 3-dimension numerical PAD emotions is given. Finally, I describe how PAD emotions affect the negotiation strategy and provide an evaluation which shows that it can be used to implement emotional agents that mimic human emotions during negotiation. Thus we can design high EQ agents for negotiation according to specific design purposes. Since negotiation is used widely in many different domains, this research, based on a general process of negotiation, can also be widely applied to other areas.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Most of the great classical philosophers, such as Plato [2], Aristotle [123], Spinoza [34], Descartes [37], Hobbes [136], and Hume [138], had recognizable theories of emotion. Though many philosophers of mind and psychologists in the twentieth-century have tended to neglect them, in recent years emotions have once again become the focus of vigorous interest in philosophy, as well as in other branches of cognitive science [32]. Theories and research show that emotions have a relationship to bodily states, motivations, beliefs, and desires.

Meanwhile, traditionally, most of the research into agents has focused on the development of rational utility-maximizing agents. This research assumes that decisions derive from an analysis of the future outcomes of various options and alternatives. Thus, following questions come out:

- Do agents need emotions? Or can agents be intelligent without emotions?
- Is it possible to incorporate emotions into agents?
- How to incorporate emotions into agents?
- Is it possible to build high EQ (emotional quotients) agents which have better performance than rational agents?

Around above questions, I describe the motivation and overview of my dissertation as follows.

1.1 Motivation

Since Wright [184] and Picard [135] placed emotion into computational theory, emotions have received increasing attention in several AI-related fields, however most prominently in human-robot/computer interaction, which focus on how to express or sense emotions. The influence of emotions on decision-making is largely ignored. Recently, there are a few projects working on commonsense reasoning, however research on applying emotions to agents' decision making is still very limited.

This is technically reasonable. The difficulties involve follows:

- The emotion theory is not complete. As in [43], Ekman reveals the central issues in emotion research and theory in the words of many of the leading scientists working in the field today. Davidson [29] gives a comprehensive road-map to the burgeoning area of affective sciences, and brings together the various strands of inquiry and the latest research in the scientific study of the relationship between the mechanisms of the brain and the psychology of mind. Thus, emotion theory still has some room for development and to build an emotion model to reflect above relationship completely is impossible so far.

- The detailed relationship between emotions and reasoning process is complex and difficult to represent in a simple model. Though emotion theories show that emotions do have some relation to bodily states, motivation, beliefs and desires, the details of the relationships are still not very clear. There are many researchers trying to model emotions, to show how emotions could affect human's behavior, and how environment changes could affect human's emotions. However, there is no standard emotion model to represent such relationships.

- The effects of emotions on human behavior differ individually. For example, one may be timid and another bold. One may be prone to attacks of rage, another peaceable.

On the other hand, to model emotions' effects on decision making or reasoning is important. If we hope to build agents that behave like humans then we must incorporate emotions into our design. As well as expressing or sensing emotions, the internal mechanism of emotions should also be considered, which is the core and involves how emotions affect the decision-making and how emotions are updated.

Also, perfectly rational agents make decision based only on the information that directly related to their intention or goal. As described in Figure 1.1, a behavior is rational generally because there is some relation between the reason and the behavior. If there is no direct relation between the reasons and the behavior, we think the behavior is not rational, so we might describe it as an emotional behavior. Thus, if there is no direct relation between the reasons and the behavior, such reasons are often ignored by rational agent, since it is not rational. However, humans are not perfectly rational and often let their emotions, even those unrelated to the current situation, affect their decisions. Emotional behaviors are different from rational ones, but they are not in complete conflict. By adding emotions between the behavior and the unrelated reason, many things become easy to explain. For example, suppose an agent A gets a gift from a friend B today that makes him very happy. When people are in a happy mood they are more willing to help others. C asks A for help. Usually A rejects C, but today A gives C the help that C requests. There is no direct relation between the fact that A gets a gift from B and that A helps C, but by adding emotions we can explain it. Usually the effect of emotions is rational, in other words, there is often a reason why people are happy or sad. On the other hand, from emotion to behavior, there are also some rules to follow.

Emotions do have some effect on people's behavior. However, these effects are

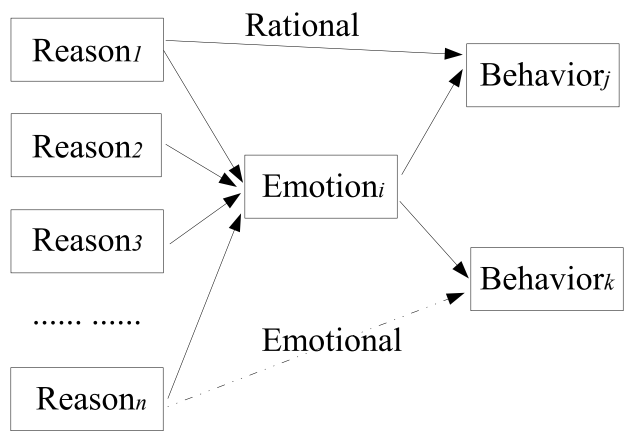


Figure 1.1: Rational and Emotional Behaviors Description

usually ignored in multiagent system. Correspondingly, some features which seems not directly related are also ignored. For some important features sometime people try hard to set up kind of complex relation to make it rational. By adding emotions in, we can ignore the relations set-up for some not directly related features, but still count the part of affection in.

Meanwhile, can agents be intelligent without emotions? We believe that, even if we do not want to build human-like agents, the affect of emotions on decision-making still can not be ignored. As mentioned in some neurological studies, patients with brain lesions that prevent them from processing emotions also have trouble to make decision. So, emotions can help decision-making. For example, emotions can be helpful as they serve as an efficient way to prioritize an agent's multiple goals. In this way they can reduce the computational load of an otherwise rational agent.

Therefore, it is possible to build high EQ agents which have better performance in complex dynamic environment than regular rational agents.

1.2 Overview

In the dissertation, first the related research background is given in Chapter 2. Among it, the emotion theories are outlined first as in Section 2.1. Then Section 2.2 gives a survey of the current research involving applying emotion theories in AI area, and the main research directions are listed. Next, the applied problem domain — negotiation — is described in Section 2.3.

As discussed in above chapter, emotions are as important to AI as to human beings, and research is needed especially to model emotions' effects on decision making or reasoning. Recently, the "Architectures for Commonsense Reasoning" project at the MIT Media Lab and the "Reasoning and Cognition" project at The University of Birmingham School of Computer Science have focused on the relationship of emotions and reasoning, though this kind of research is still very limited.

Our research focuses on emotions' effects on decision-making as well. However, we address the same problem domain differently. Taking into account the incompleteness of current emotion theories and emotional differences among individual persons, here I model EBDI, a common architecture for agents with emotion status, which specifies a separate emotion mechanism within an agent, instead of trying to model emotion mechanisms to reflect the reasoning process specifically, like most researchers have done. As described in Chapter 3, this common architecture can reflect the practical reasoning process, and incorporate specific emotion models and change a traditional rational agent to an emotional agent for some specific application. Thus, we can then select and apply part of an emotion theory into the architecture as needed.

Therefore, this common architecture for agent has following features:

- It contains emotion status, the description method of which is not limited. That is, how to represent the emotion status depends on the specific application. For example, for negotiation, we need a numerical measurement for emotion status, since agents negotiate base on some utility function, the result of which is a number; for some other application, we may use first-order logic, multidimensional logic [59], or others.

- It contains the function of emotion update mechanism, and involves emotion in decision process. The function for emotion update mechanism should keep some flexibility such that we can apply part of emotion theory in it according to specific purpose. Since emotion is not only the interest of computer science, it is also interest of philosophy, cognitive science, neuroscience and behavioral economics, all the research results are helpful in building agents. This common architecture is expected to be flexible enough to involve any old or new research result in. On the other hands, the agents may differ and purposes are differ, so we do not have to model agent that applies a complete emotion mechanism, which is complex and may not exist so far. Thus, we desire this common architecture can apply some emotion mechanism, the details of which we do not care.

- It reflects the practical decision process of human being. This is important, since the practical decision process in human being is natural, harmonious and efficient.

- It is practically applicable. Or, we need to put it to work instead of a concept only. So Sample agents in Tileworld are given, and the results show that an EBDI agent can have better performance than traditional BDI agents. Furthermore, we need to apply this common architecture of agent to some specific problem domain, and define method to record emotion status, and apply emotion theories to update emotion and affect decision process. It will be better to apply it in different domains, and apply different emotion mechanisms. However, in the dissertation, we only apply it to negotiation, and other applications can be in the future work.

- It is compatible with currently widely used agent architecture. For traditional rational agents, there are already widely accepted theories and successful applications. Here, we do not intend to rebuild a complete new emotional agent, instead, we hope it is compatible with those theories for rational agents. Thus, the new emotional agent will be partially rational and partially emotional, just like us human beings. However, we can adjust the weight of the rational part or emotional part according to specific needs.

Then I apply this model in the domain of negotiation, which has been a subject of central interest in DAI (Distributed Artificial Intelligence) and multi-agent area, as it has been in economics and political science, and has wide applications. The negotiating procedures have included the exchange of Partial Global Plans [40, 41], the communication of information intended to alter other agents' goals [167, 168], and the use of incremental suggestions leading to joint plans of action [83]. So, this domain need us to handle multi-level issues, among them, reasoning process is a key issue, agents may have different strategies, they try to affect others and also be affected by others. The main concerns here are:

- During the negotiation process, each agent has its own strategy, which involves reasoning and decision making process;

- Since each agent may have different strategy, we can then design agent with different strategy, some with emotion status, some without emotions, such that we can compare the behaver of the agents with or without emotions;

- Negotiation involves communication process. If we can apply the model of emotional agent to negotiation, then it has big possibility to apply it to other issues in multiagent system.

To apply this architecture in negotiation, I first design a plug-in for EBDI architecture, which modifies OCC model, a standard model for emotion synthesis, to generate emotions, as in Chapter 4. This chapter solves the problem that how an agent generates emotions and how the emotions are updated. Then I analyze the possibility to incorporate emotions into negotiation and generate EWOD (Emotional Worth-Oriented Domain), which require emotions to be numerical. It is described in Chapter 5. Thus, a mapping from 22 OCC emotions to 3-dimension numerical PAD emotions is given in Chapter 6. Finally, how these 3-dimension emotions affect the negotiation strategy is described and an evaluation is given as in Chapter 7, which solves the problem how the different emotion status affect the decision result during the negotiation process. It also shows that it can be used to implement agents with various emotional states that mimic human emotions during negotiation. Thus we can potentially design an agent for negotiation with a high EQ agent according to specific applications and purposes. Since negotiation is already used widely to solve many problems in different domains, and this research is based on a general process of negotiation, the research results can also be widely used in other areas.

Chapter 2

Related Research Background

2.1 Emotion Theories

2.1.1 History

From ancient, medieval history and renaissance, there are many classical philosophers had recognizable theories of emotion. For Plato in the Republic [2], there seemed to have been three basic components of the human mind: the reasoning, the desiring, and the emotive parts. Aristotle [150] describe emotions as first and foremost responses found in the embodied animal to the outside world, which are largely passive states, located within a general metaphysical landscape contrasting active and passive, form and matter, and actuality and potentiality. Compared with Aristotle's moderation, the Stoics seem pretty intolerant of the emotion, stressing their cognitive, eudaimonistic, and moral failings, while recommending their elimination. Stoic doctrines were largely transmitted to early modern philosophers through the writings of Cicero and Seneca [151]. Galen [55] adopted many Stoic physical, metaphysical, epistemological and ethical views on the Pathe. But he also drew off an independent Hippocratic tradition for treating the humors and the physiology of the emotions and produced an influential account of the "spirits." Although the emotions were not a central topic for the Epicureans, the presentation of their views on pleasure and the good life through Diogenes Laertes [88], Lucretius [161], and even such critics as Cicero [23] were important enough to early modern philosophers. On Augustine's view [7], the will is active, identified as a "movement of the soul, under no compulsion, either toward getting or not losing something" and the will simply incorporates the passions into its attractive, hedonistic operations. For Aquinas [75], passions are acts, or movements, of the sensitive appetitive power, which are caused by external objects; passions of the soul can also be identified with certain bodily changes, including contraction or expansion of the "spirits," changes in the distribution of bodily temperature, and particularly alterations in the movements of the heart. Niccoló Machiavelli [96] take a different approach in considering how to characterize humans, particularly geographically specific groups of humans, in terms of their emotional dispositions and the patterns of behavior so motivated. Justus Lipsius [155] adopted a typically Stoic approach to the passions and "affects", identifying them as false opinions that we "must never stop attempting to conquer". Later on, scholars start to treat emotions systematically, such as Francesco Suarez [151].

Early modern discussions of the emotions are deeply indebted to earlier sources, though, which were also soundly rejected by some of the most famous philosophers, starting with Descartes [150]. For example, certain of Stoicism's doctrines were explicitly criticized, including the view that the passions are erroneous judgments. Descartes [37] established that emotions were due to the overall nature of the character of the individual – called Cartesian affect theory. Other main individual Philosophers during this period are Hobbes, Malebranche, Spinoza, Shaftesbury, Hutcheson and Hume. Hobbes [136] explains emotions or "passions", as appetites or aversions of particular things. Malebranche [97] identifies seven moments that together make up the structure of the passions: judgment, impulse of the will, accompanying sensation, some bodily changes, a sensible "emotion" of the soul, some disturbances in the brain, and inner delight. He also believes that the communication of the emotions is crucial to social organization and cohesion. For Spinoza [34, 33], emotions are not lodged in a separate body in conflict with the soul, since soul and body are aspects of a single reality; but emotions, as affections of the soul, make the difference between the best and the worst lives, as they either increase the soul's power to act, or diminish that power. [150] Shaftesbury and Hutcheson's practical concerns with the emotions focused on distinct questions of moral philosophy. [62] Hume's notorious dictum that reason is and ought to be the slave of the passions also placed the emotions at the very center of character and agency.

In the twentieth-century, many of the philosophers and psychologists tended to neglect emotions – perhaps because the sheer variety of phenomena covered by the word "emotion" and its closest neighbors tends to discourage tidy theory. In recent years, emotions have once again become the focus of vigorous interest in philosophy as well as in other branches of cognitive science. In view of the proliferation of increasingly fruitful exchanges between research of different stripes, it is no longer useful to speak of the philosophy of emotion in isolation from the approaches of other disciplines, particularly psychology, neurology and evolutionary biology [32].

2.1.2 What is Emotion?

Emotion is complex, and the term has no single universally accepted definition. Kleinginna [77] mentions that there are as many as 92 different definitions in the literature. For example, emotions are described as conscious states [90], cognitive processes [164], psychosocially constructed, dramatized feeling [101], or mental states that arise spontaneously, rather than through conscious effort. Some other descriptions involve adaptive dispositions, evaluative judgments, or even social facts or dynamical processes [32]. Emotions are physical expressions, often involuntary, related to feelings, perceptions or beliefs about elements, objects or relations between them, in reality or in the imagination. Thus, the study of emotions is part of psychology, neuroscience, and, more recently, artificial intelligence.

The simplest theory of emotions, and perhaps the theory most representative of common sense, is that emotions are simply a class of feelings, differentiated from sensation and proprioception by their experienced quality. William James [69] proposed a variant of this view – the "James-Lange" (after James and Carl G. Lange) theory of emotion, according to which emotions are specifically feelings caused by changes in physiological conditions relating to the autonomic and motor functions. However, James did not give an adequate account of the differences between emotions, which was first voiced by Walter Cannon [20]. Cannon claimed that the visceral reactions characteristic of distinct emotions such as fear and anger are identical, and so these reactions cannot be what allow us to tell emotions apart. The same conclusion is usually drawn from an oft-cited experiment performed by Stanley Schacter and Jerome Singer [148]. Subsequent research has shown that a limited number of emotions do have significantly different bodily profiles [91, 130]. However, bodily changes and the feelings accompanying these changes get us only part way toward an adequate taxonomy. Rorty [140] believes that every emotion has a formal object if it has any object, which is a property implicitly ascribed by the emotion to its target, focus or propositional object, in virtue of which the emotion can be seen as intelligible. Antonio Damasio [28] points out that emotions involves a capacity for the brain to monitor the body's past and hypothetical responses, both in the autonomic and the voluntary systems. While, it falls short of fully explicating the intentional nature of emotion.

The most parsimonious type of cognitivist theory [165, 120, 121] follows the Stoics in identifying emotions with judgments. For example, my anger at someone simply is the judgment that I have been wronged by that person. Other cognitivist theories introduce further elements into their analysis. Emotions have been described as sets of beliefs and desires [99], affect-laden judgments [95], and as complexes of beliefs, desires, and feelings [122]. Deigh [36] has objected that the view of emotions as propositional attitudes has the effect of excluding animals and infants lacking language. Others have argued that if emotions always involve the standard propositional attitudes, namely belief and desire, then an account of the rationality of emotions will collapse into an account of what it is for those standard propositional attitudes to be rational: but emotional rationality is not reducible to the rationality of beliefs or desires [31, 45, 61]. Furthermore, several theorists insist that experiences of emotion have content beyond any propositional content [61, 182].

A crucial mandate of cognitivist theories is to avert the charge that emotions are merely “subjective.” According to [165], emotions may mislead us into “hasty” or “emotional” judgments. On the other hand, the lack of perceptual capacities can be a crippling handicap in one’s attempt to negotiate the world: similarly a lack of adequate emotional responses can hinder our attempts to view the world correctly and act correctly in it [121].

To secure the connection between emotion and cognition, some take the view that emotions are a kind of perception. While de Sousa [31] and Rorty [140] take another way and view the role of emotions as providing the framework for cognitions of the more conventional kind. Some philosophers suggest that the directive power which emotions exert over perception is partly a function of their essentially dramatic or narrative structure [141].

So as the subject of scientific research, emotion has multiple dimensions: behavioral, physiological, subjective, and cognitive. Sloman and others explain that the need to face a changing and unpredictable world makes emotions necessary for any intelligent system (natural or artificial) with multiple motives and limited capacities and resources [164].

2.1.3 Emotions and Reasoning

Current research on the neural circuitry of emotion suggests that emotion makes up an essential part of human decision-making, including long-term planning. Early in Aristotle work, the Nichomachean Ethics [4] is concerned with the place of the emotion within the economy of acting according to our habits and desires as moderated by reason, whereas the Rhetoric [5] concerns the arousal and management of emotion in the context of producing persuasion. In both cases, the emotions are treated as susceptible to rational influence and voluntary action. So Aristotle's assessment of the emotions is mixed: they can be cultivated by reason and figure in the good life; they can also disrupt our reason and action, and be used for nefarious ends. On Damasio's view [27], emotions are passions for reasoning, which makes the brain system to be enmeshed in need of reasoning. However, emotion was sometimes regarded as the antithesis of reason. This is reflected in common phrases like appeal to emotion or your emotions have taken over. Emotions can be undesired to the individual feeling them; he or she may wish to control but often cannot. Later on, some state that there is no empirical support for any generalization suggesting the antithesis between reason and emotion: indeed, anger or fear can often be thought of as a systematic response to observed facts. In any case, it is clear that the relation between logic and argument on the one hand and emotion on the other, is one which merits careful study.

For rationality, the clearest notions associated are coherence and consistency in the sphere of belief, and optimizing outcomes in the sphere of action. But these notions are mainly critical ones, and they would be not suffice to guide an organism toward any particular course of action by themselves. Because the number of goals logically possible to posit at any particular time is virtually infinite; the number of possible strategies that might be employed in pursuit of them is a lot; and moreover, in considering possible strategies, the number of consequences of any one strategy is again infinite, so that unless some drastic pre-selection can be effected among the alternatives their evaluation could never be completed. This gives rise to what is known among cognitive scientists as the "Frame Problem" [103]: in deciding among any range of possible actions, most of the consequences of each must be eliminated a priori, i.e. without time being wasted on their consideration.

This may be due to our capacity for emotions, since emotions constitute one of the chief mechanisms whereby attention is constrained and directed [102]. This allows them to frame our decisions in two important ways. First, they define the parameters taken into account in any particular deliberation. Second, in the process of rational deliberation itself, they render salient only a tiny proportion of the available alternatives and of the conceivably relevant facts. The suggestion relabeled as "Search hypothesis of emotion", has been criticized and elaborated by Evans [49], who argues convincingly that it needs to be buttressed by a positive theory of what the emotional mechanisms actually are which are capable of effecting this task.

In a more pervasive way, the capacity to experience emotion seems to be indispensable to the conduct of a rational life over time. Antonio Damasio [27] has amassed an impressive body of neurological evidence suggesting that emotions do have this sort of function in everyday reasoning. Subjects in his studies who, because of injuries sustained to the prefrontal and somatosensory cortices of the brain, had a diminished capacity to experience emotion, were severely hindered in their ability to make intelligent practical decisions. In these ways, then, emotions would be all important to rationality even if they could not themselves be deemed rational or irrational. Thus can emotions themselves be assessed for their rationality? It is enough to note that there is no logical reason why judgments of reasonableness or irrationality in relation to emotions need be regarded as any more subjective than any other judgments of rationality in human affairs [32]. Certain philosophers have argued that emotions are more like actions [147, 165]. However, if this is true, and emotions are to some extent under our voluntary control, then emotions will also be assessable for their strategic rationality. If a person is not aware that a substitution has taken place, then she will be self-deceived about her emotions.

Then how one conceives of the nature of emotional rationality will depend on one's theory of what the emotions are. Cognitivist and appraisal theories will say that a reasonable emotion is one whose constituent propositional attitudes or appraisals are reasonable. Theories which take emotions to be perceptions of objective values will claim that the target of an appropriate emotion possesses the value which the emotion presents it as having. Narrative theories will consider an emotion appropriate if its dramatic structure adequately resembles that of its eliciting situation. All these suppose that the relevant notion of rationality is an epistemic one, and that what appropriate emotions succeed in achieving is some sort of representational adequacy.

2.2 Emotions in AI

Over the last several years, emotions have received increasing attention in many AI-related fields. Many concerns are mentioned when researchers take emotions into agents. The main issues concerned are the following:

- Some researchers consider it necessary to incorporate human aspects such as personality and emotion in order to make agents more engaging and believable so that they can better play a role in various interactive systems involving simulation [129]. Entertainment is one obvious application area for such simulation systems, another is education and training. For example, a simulator that was able to realistically model emotional reactions of people could be used in training programs for staff who need to be trained to deal with the public.

- Some people believe that emotions play a functional role in the behavior of humans and animals, particularly behavior as part of complex social systems

[172]. a successful modeling of emotion will enable us to come closer to the goal of building software agents which approach humans in their flexibility and ability to be adaptable and survive in complex, changing and unpredictable environments. For example, as in systems the Woggles of Oz-world [12], emotion modifies the physical behavior of agents: a happy agent moves faster, and more bouncily, while a sad agent is slower and flatter in its movements.

- Emotions can effect an agent's goals, hence affecting their actions. Emotional effects on goals can be via reordering or re-prioritizing, existing goals, or by introducing completely new goals. The goals' success or failure can affect emotional states. An agent which experiences a goal failure may feel unhappy while one experiencing goal success may feel glad. Dyer [42] develops a comprehensive lexicon of emotional states based on goal success and failure.

- Frijda [53] postulates emotions as processes which safeguard long-term persistent goals or concerns of the agents, such as survival, a desire for stimulation or a wish to avoid cold and damp.

- Toda [172] postulates emotions as processes which affect the rational system of the agent, and which are based on basic urges: emergency urges, biological urges, cognitive urges and social urges. Emotions are seen as varying in intensity where the intensity level is an important factor in determining the effect on the rational processing of the agent.

- Rational agents often are thought as self-interest, that is, they always want to maximize their own wealth or other material goals. However, practically, people may sometimes choose to spend their wealth to punish others who have harmed them, reward those who have helped, or to make outcomes fairer [18].

- Damasio [27] finds that people with relatively minor emotional impairments

> have trouble making decisions and, when they do, they often make disastrous ones. Other research shows that what appears to be deliberative decision making may actually be driven by gut-level emotions or drives, then rationalized as a thoughtful decision [176]. Bechara [13] also mentions that most theories assume that decisions derive from an assessment of the future outcomes of various options and alternatives through some type of cost-benefit analysis. The influence of emotions on decision-making is largely ignored. The studies of decision-making in neurological patients who can no longer process emotional information normally suggest that people make judgments not only by evaluating the consequences and their probability of occurring, but also and even sometimes primarily at a gut or emotional level.

Following sections will show details along three main research directions in this area.

2.2.1 Affective Computing

The early interests of involving emotions in AI is triggered by Rosalind Picard's book – "Affective Computing" [135]. Its proponents believe computers should be designed to recognize, express, and influence emotion in users. Though Picard is known as the godmother of this field, the research in the computational theories of emotion are started early in 1970's, and it has been particularly attractive to psychiatrists and psychoanalysts for a long time.

The computational theories were broached early by a couple of psychoanalysts turned hackers Peterfreund [133], Shank and Colby [157] and played an important role in the theoretical elaborations of John Bowlby's work on the mechanisms and psychological consequences of early separation and loss [15]. These works attempted to model Freudian concepts of the dynamics of conscious and unconscious mental life in computational terms. Colby even constructed a simulation of a paranoid patient, "Parry", which famously fooled some psychiatrists. The key idea was to set up second-order parameters that acted on the first-order modules of perception, belief and desire, thus regulating or disrupting the operation of perceptual and action programs. From the sidelines, de Sousa [31] suggested that connectionist systems or analog models stand a better chance of modeling emotion than those based on classical von Neuman-type digital computation, but that suggestion has not gone anywhere. From the point of view of computational theory, the prevailing wind, backed by both evolutionary speculation and neurological findings on control systems and relatively independent affect-programs, has tended to favor modular conceptions of emotion rather than holistic ones [22].

Still, some philosophers and computer scientists have continued to be interested in integrating computing theory with emotions [32]. Aaron Sloman has elaborated the sort of ideas that were embryonic in Shank and Colby into a more sophisticated computational theory of the mind in which emotions are virtual machines, playing a crucial role in a complex hierarchic architecture in which they control, monitor, schedule and sometimes disrupt other control modules [184]. Rosalind Picard [135] lays out the evidence for the view that computers will need emotions to be truly intelligent, and in particular to interact intelligently with humans. She also adverts to the role of emotions in evaluation and the pruning of search spaces. But she is as much or more concerned to provide an emotional theory of computation than to elaborate a computational theory of emotions. Lastly, a forthcoming book by Marvin Minsky [115] bears the promising title of "The emotion machine". Eliott [44], Cãnamero [19], Marsella and Gratch [100] work on "computational models of human emotion", which typically are studied in simulations in artificial environments.

2.2.2 Human-Robot/Computer Interaction

Though there are many research involving emotions in AI area, while most are prominently in human-robot/computer interaction, where emotional receptiveness – ability to perceive and interpret emotional expressions of others, and expressivity – ability to express emotions in a way that can be perceived and interpreted by others, are crucial [149]. The main focus in most of the employed agents is on the display of emotions such as animated facial expressions and/or on their recognition such as in speech signals.

Over the recent past, researchers have been particularly interested in endowing artificial agents with emotional expressivity to improve their "believability" and to make them more "life-like" [12, 66, 128]. Such believable virtual and robotic agents and human-like synthetic characters are of particular interest, with applications ranging from the entertainment industry, to training and tutoring systems [63, 158, 24], as well as in the design of user-interfaces [124, 67, 169].

There is also an increasing number of examples of robotic agents that are based on emotional control [113, 17, 6], most of which are intended for human-robot interaction.

2.2.3 Architectures and Commonsense Reasoning

While achieving believable emotion display and reliable emotion recognition are important goals in the context of designing virtual and robotic agents for human - robot/computer interaction, the more general question about what possible roles emotions could have in an agent architecture and in what circumstances they might be useful for the control of agents and possibly even better than other, non-emotional control mechanisms, has received very little attention. However, it becomes more and more important.

Actually, from very early on, architectures with emotional components have been proposed for simple and complex agents [171, 160, 42], and several others. Pfeifer

[134] has a discussion of the early models, which highlights how the essential issues have not really changed in the past 15 years.

Padgham [129] believes that emotions and personality interact with goal-oriented behavior and describes some simplifications to build an initial interactive environment for experimentation with animated agents that simulate personality alongside rational goal-oriented behavior. Morgado [117] presents an agent model where emotion and cognition are conceived as two integrated aspects of intelligent behavior. They show affective-emotional mechanisms that support the adaptation to changing environments and a controlled use of resources. Meyer [112] extends the KARO (Knowledge, Abilities, Results and Opportunities) framework – supplies a range of expressive modal logics for describing the behavior of intelligent agents [68], and use logic in reasoning about the emotional or affective states that an agent can reside in.

The MIT Media Lab is developing a theory about the architecture of commonsense thinking. The design is described most fully in Marvin Minsky's forthcoming book *The Emotion Machine* [115], where an emotional state is looked as a different way of thinking. The related projects about the architecture of commonsense thinking include:

- Architectures for Common Sense Reasoning: It is try to solve problem about how to build reasoning system as resourceful and adaptive as people, and focus on developing new types of reasoning technologies and cognitive architectures that support great procedural and representational diversity.

- The Panalogy Reasoning Engine: It is developed to be one instance of the Emotion Machine Architecture that places a special emphasize on reflective analogical reasoning using multiple representations [163].

- Reasoning and Cognition Project at The University of Birmingham School of Computer Science: It covers research on architectures for accounting for human mental states and processes as well as recreating them in computer programs. It also includes research on automated reasoning with applications to mathematical knowledge management and computer algebra, and try to investigate whether the ability to have emotional states is an accident of animal evolution or an inevitable consequence of design requirements and constraints by analyzing architectures for human mental states and processes.

Juan D. Velásquez [174] in *When Robots Weep: Emotional Memories and Decision - Making* describes an agent architecture that integrates emotions, drives, and behaviors, and that focuses on modeling some of the aspects of emotions as fundamental components within the process of decision-making. He also shows how the mechanisms of primary emotions can be used as building blocks for the acquisition of emotional memories that serve as biasing mechanisms during the process of making decisions and selecting actions. The architecture has been implemented into an object-oriented framework that has been successfully used to develop and control several synthetic agents and which is currently being used as the control system for an emotional pet robot. Pereira [132] presents a Emotional-BDI architecture including internal representations for agent's capabilities and resources. However, this paper does not represent the difference between emotional agents and normal rational agents. The capabilities and resources themselves are independent of emotions, as such, they cannot reflect the relationship between emotions and beliefs or how emotions influence agents' decision making. Another related research is [131], which enhance the standard BDI model using the OCC(Ortony, Clore, Collins) model of emotion [125] in a framework that can support large numbers of combatants.

2.3 Negotiation

Negotiation has been a subject of central interest in DAI (Distributed Artificial Intelligence), as it has been in economics and political science [139]. As defined in [177], negotiation is a process by which a joint decision is reached by two or more agents, each trying to reach an individual goal or objective. The negotiating procedures have included the exchange of Partial Global Plans [40, 41], the communication of information intended to alter other agents' goals [167, 168], and the use of incremental suggestions leading to joint plans of action [83]. The agents in a negotiation face an interesting problem [175]: They want to maximize their own utility but they also face the risk of a break-down in negotiation, or expiration of a deadline for agreement. Thus, each agent must decide whether the current deal is good enough or whether it should ask for more and risk agreement failure.

We can use negotiations to resolve conflicts in a wide variety of multi-agent domains [70], such as conflicts over the usage of joint resources or task assignment, conflicts concerning document allocation in multi-server environments and conflicts between a buyer and a seller in electronic commerce. Also, negotiation can be an effective method for finding the one global course of action which maximizes utility without having to send all the local knowledge bases to a central location for consideration in a problem where each agent has different local knowledge. As such, automated negotiation provides a distributed method of aggregating distributed knowledge [30].

To build an autonomous agent which is capable of flexible and sophisticated negotiation, the main questions to be considered are [79]:

- What negotiation protocol will be used?
- What reasoning model, decision making procedures and strategies will the agents employ?

2.3.1 Negotiation Problems

Bargaining Problem

The bargaining problem [126, 127] is a specific version of the negotiation problem that has been studied in game theory, which was first proposed in [118]. Formally, we assume that each agent i has a utility function u_i defined over the set of all possible deals Δ.

$$u_i : \Delta \rightarrow \Re$$

And there is a special deal δ^- which is the no-deal deal. Without loss of generality we will assume that for all agents $u_i(\delta^-) = 0$ so that the agents will prefer no-deal than accepting any deal with negative utility. The problem then is finding a protocol f which will lead the agents to the best deal.

Task Allocation Problem

The task allocation problem is a common problem in multiagent system, which wants to decide how to re-allocate a set of tasks among a set of agents [175]. Formally, in this problem there is a set of task T, a set of agents, and a cost function

$$c_i : s \rightarrow \Re$$

which tells us the cost that agent i incurs in carrying out tasks $s \subseteq T$. In some simplified versions of the problem we assume that all agents share the same cost function. The agents start out each one with a set of tasks such that all tasks are distributed. We can think of this initial allocation as δ^- since, if negotiations break down then each agent is responsible for the tasks it was originally assigned. Similarly, every allocation of tasks to agents is a possible deal δ where $s_i(\delta)$ is the set of tasks allocated to i under deal δ. The problem we then face is how to design a negotiation protocol such that the agents can negotiate task re-allocations and arrive at a final re-allocation of the tasks.

Complex Deal

Real world deals are known to be composed of many different items [175] such as price, warranty, delivery time, color, etc. For example, two agents negotiating over the purchase of a car will need to agree on the price to be paid, the color of the car, the number of years in the warranty, the value of the tradein car, the type of financing and interest rate provided, and many other possible parameters. These dimensions could be used explicitly during the negotiation when, for example, an agent claims that it cannot pay more than $5,000$ for the car or that it really likes red cars. These dimensions inevitably lead to an explosion in the space of possible deals. Dynamic strategy for a complex world also explores multiple dimensions of negotiation [178]. A land assembly, for example, illuminates how linked negotiations to acquire separate parcels must be linked to a larger strategy.

More formally we define a multi-dimensional deal as composed of a set of multi-dimensional deal variables $x_1, x_2, ..., x_n$ with domains $D_1, D_2, ..., D_n$, respectively. For example, one variable could correspond to price and its domain could be all integer numbers in the range 0 to 100. We can also re-define the agents utility functions so that they explicitly deal with each variable. For example, instead of an opaque $u_i(\delta)$, we could have a more expressive:

$$u_i(\delta) = c_1 u_1^i(x_1) + c_2 u_2^i(x_2) + \cdots + c_n u_n^i(x_n)$$

or some other combination of the individual variables. The negotiation problem remains that of finding a deal that corresponds to a chosen solution concept, such as the utilitarian deal or the Nash bargaining deal.

2.3.2 Negotiation Approaches

Here, we will give a short survey of two main approaches to negotiations in the social sciences and various negotiation approaches in Distributed Artificial Intelligence (DAI). We will then demonstrate the application of one of the approaches to multiagent systems, which we are interested in.

Negotiation Approaches in the Social Sciences

There are two main approaches to the development of theories relating to negotiation in social sciences.

The first approach is the formal theory of bargaining, which is known as "Nash's Bargaining Problem" [94] or "Nash's Model of Bargaining" [143], constituting a formal, game-theoretic approach that provides clear analysis of various situations and precise results concerning the strategy a negotiator should choose. Classic game theory [118, 186, 65, 143, 94] talks about agents reaching "deals," which are defined as vectors of utilities. A bargaining problem is described as in previous section. Nash [118, 119] shows that under some rational behavior and symmetry assumptions, players will reach an agreement on a deal that will be *individual rational, pareto optimal*, and will maximize the product of the players' utility. Zeuthen [186] considered the two-players bargaining problem as a one-player decision process, and evaluated how much risk each player would be willing to take when he decides not to concede. The player who is least willing to risk will be the one who will make the next concession. Harsanyi [65] showed that if both players use the Zeuthen strategy they will converge to a Nash solution. Rubinstein and Osborne [144, 145, 126] discussed an alternative approach about it as well. Above all, this approach can only be applied to situations satisfying very restricted assumptions. In particular, this approach assumes that the agents are acting rationally, have large computation capabilities and follow strict negotiation protocols.

The second approach, comprises informal theories which attempt to identify possible general beneficial strategies for a negotiator. The works based on this approach advise a negotiator how to behave in order to reach beneficial results in a negotiation [139, 51, 38, 74, 73, 64]. These negotiation guides do not presuppose the strong restrictions and assumptions presented in the game-theoretic models. They can be used in domains where people interact with each other and with automated systems, and situations where automated systems interact in environments without predefined regulations. These informal models can serve as guides for the development of negotiation heuristics [81] or as a basis for the development of a logical negotiation model [82]. However, applying these methods to automated systems is more difficult than using the first approach, since there are neither formal theories nor strategies that can be used.

Other work includes [52, 98], which focuses on the organizational aspects of societies of agents, and [56, 57] by Gasser, which explored the social aspects of agent knowledge and action in multiagent systems. Gasser's approach exploits a predesigned social layer for multiagent system that, social mechanisms can dynamically emerge and "communities of programs" can generate, modify and codify their own local languages of interaction. It may be effective when agents are interacting in unstructured domains, or in domains where their structure is continuously changing.

Negotiation Models in DAI

Negotiations were used in DAI both in Distributed Problem Solving (DPS) where the agents are cooperative and in Multiagent Systems (MA) where the agents are self-interested [79]. In DPS, much of the work focused on the implementation and analysis of data fusion experiments, where systems of distributed sensors absorb and interpret data, ultimately arriving at a group conclusion [39, 35, 87]. There are several works using negotiation for distributed planning and distributed search for possible solutions for hard problems. For example, Conry [25] suggests multi-stage negotiation to solve distributed constraint satisfaction problems when no central planner exists. Moehlman and Lesser [116] use negotiation as a tool for distributed planning: each agent has certain important constraints, and it tries to find a feasible solution using a negotiation process. They applied this approach in the Phoenix fireman array. Lander and Lesser [89, 26] use a negotiation search, which is a multi-stage negotiation as a means of cooperation while searching and solving conflicts among the agents.

Rosenschein and Zlotkin [142] identified three distinct domains for the multiagent environments, where negotiation is applicable and found a different strategy for each domain:

1. Task-Oriented Domain: Finding ways in which agents can negotiate to come to an agreement, and allocating their tasks in a way that is beneficial to everyone;

2. State-Oriented Domain: Finding actions which change the state of the "world" and serve the agents' goals;

3. Worth-Oriented Domain: Same as 2 above, but, in this domain, the decision is taken according to the maximum utility the agents gain from the states.

Sycara [166] presented a model of negotiation that combines case-based reasoning and optimization of multi-attribute utilities, where agents try to influence the goals and intentions of their opponents. Ephrati and Rosenschein [46, 47, 48] used the Clarke Tax voting procedure as a consensus mechanism, in essence to avoid the need for classical negotiation. Kraus [83, 85, 78] explored negotiation in which the negotiation time itself is an issue. Kraus and Lehmann [81] developed an automated Diplomacy player that negotiates and plays well in actual games against human players. Sierra et al. Sierra [159] present a model of negotiation for autonomous agents to reach agreements about the provision of service by one agent to another. Their model defines a range of strategies and tactics, distilled from intuition about good behavioral practice in human negotiation, that agents can employ to generate offers and evaluate proposals. Zeng and Sycara [185] consider negotiation in a marketing environment with a learning process in which the buyer and the seller update their beliefs about the opponent's reservation price [1] using the Bayesian rule. Sandholm and Lesser [146] discuss issues arising in automated negotiation among self-interested agents whose rationality is bounded by computational complexity, such as levels of commitment.

2.3.3 Strategic Negotiation

The strategic-negotiation model [80, 79] is based on Rubinstein's model of alternating offers [144] and draws upon Rosenschein and Zlotkin's worth-oriented domain [142]. It has many applications, which includes negotiations about data allocation [9], resource allocation, task distribution, pollution reduction and hostage release.

Formally, in the strategic model there are N agents, $Agents = \{A_1, ..., A_N\}$. The agents need to reach an agreement on a given issue. It is assumed that the agents can take actions in the negotiation only at certain times in the set $T = 0, 1, 2...$ that are determined in advance and are known to the agents. In each period $t \in T$ of the negotiation, if the negotiation has not terminated earlier, an agent whose turn it is to make an offer at time t, will suggest a possible agreement (with respect to the specific negotiation issue), and each of the other agents may either accept the offer (choose Yes), reject it (choose No), or opt out of the negotiation (choose Opt). If an offer is accepted by all the agents (i.e., all of them choose Yes), then the negotiation ends, and this offer is implemented. If at least one of the agents opts out of the

[1] In microeconomics, the Reservation Price is the maximum price a buyer is willing to buy a good or service, or the minimum price a seller is willing to sell a good or service. Reservation prices vary for the buyer according to their disposable income, their desire for the good, and the prices of, and their information about substitute goods.

negotiation, then the negotiation ends and a conflictual outcome results. If no agent has chosen "Opt," but at least one of the agents has rejected the offer, the negotiation proceeds to period $t + 1$, and the next agent makes a counteroffer, the other agents respond, and so on. It assumes that an agent responding to an offer is not informed of the other responses during the current negotiation period. This protocol is called a *simultaneous response* protocol . $j(t)$ will denote the agent that makes an offer at time period t.

There are no rules which bind the agents to any specific strategy in the strategic-negotiation model . Thus there are no assumptions about the offers the agents make during the negotiation. In particular, the agents are not bound to any previous offers that have been made. After an offer is rejected, an agent whose turn it is to suggest a new offer can decide whether to make the same offer again, or to propose a new offer. The protocol only provides a framework for the negotiation process and specifies the termination condition, but there is no limit on the number of periods.

A fair and reasonable method for deciding on the order in which agents will make offers is to arrange them randomly in a specific order before the negotiation begins [14]. That is, the agents will be denoted randomly $A_1, .., A_N$. At each time t, $j(t)$ will be A_i where i is equal to $(t \ mod \ N) + 1$. The set of possible agreements is denoted S. An outcome of the negotiation may be that an agreement $s \in S$ will be reached at time $t \in T$. This outcome is denoted by a pair (s, t). When one of the agents opts out of the negotiations at time period $t \in T$, the outcome is denoted (Opt, t).

The agents' *time preferences* and the preferences between agreements an opting out are the driving force of the model. They will influence the outcome of the negotiation. In particular, agents will not reach an agreement which is not at least as good as opting out for all of them. Otherwise, the agent which prefers opting out over the agreement, will opt out.

Negotiation Strategies

An agent's negotiation strategy specifies for the agent what to do next, for each sequence of offers $s_0, s_2, s_3, ..., s_t$. In other words, for the agent whose turn it is to make an offer, it specifies which offer to make next. That is, it indicates to the agent which offer to make at $t + 1$, if in periods 0 until t the offers $s_0, ..., s_t$ had been made and were rejected by at least one of the agents, but none of them has opted out. Similarly, in time periods when it is the agent's turn to respond to an offer, the strategy specifies whether to accept the offer, reject it or opt out of the negotiation. A strategy profile is a collection of strategies, one for each agent [127].

Subgame Perfect Equilibria

The main question in this model is how a rational agent will choose its negotiation strategy. A useful notion is the Nash Equilibrium [119, 94]. Formally, a strategy profile $F = f_1, ..., f_N$ is a Nash equilibrium of a model of alternating offers,if each agent A_i does not have a different strategy yielding an outcome that it prefers to that generated when it chooses f_i,given that every other agent A_j chooses f_j. Briefly, no agent can profitably deviate, given the actions of the other agents.

Then if all the agents use the strategies specified for them in the strategy profile of the Nash equilibrium, then no agent has a motivation to deviate and use another strategy. However, the use of Nash equilibrium in a model of alternating-offers leads to an absurd Nash equilibria [170]: an agent may use a threat that would not be carried out if the agent were put in the position to do so, since the threat move would give the agent lower payoff than it would get by not taking the threatened action. This is because Nash equilibrium strategies may be in equilibrium only in the beginning of the negotiation, but may be unstable in intermediate stages. The concept of subgame perfect equilibrium (SPE) [127], which is a stronger concept, is defined in the following definition and will be used in order to analyze the negotiation.

Formally, a strategy profile is a *subgame perfect equilibrium* of a model of alternating offers if the strategy profile induced in every subgame is a Nash equilibrium of that subgame. This means that at any step of the negotiation process, no matter what the history is, no agent has a motivation to deviate and use any strategy other than that defined in the strategy profile. In situations of incomplete information there is no proper subgame. The *sequential equilibrium* [86], which takes the beliefs of the agents into consideration, can be used in the incomplete information situations.

In summary, the strategic negotiation model provides a unified solution to a wide range of problems. It is appropriate for dynamic real-world domains. In addition to the application of the strategic-negotiation model to data allocation problems in information servers, it was applied to resource allocation and task distribution problems, and the pollution allocation problem [80]. In all these domains the strategic-negotiation model provides the negotiators with ways to reach mutually beneficial agreements without delay. The application of the strategic-negotiation model to human high pressure crisis negotiations was also studied [180, 84].

2.3.4 Evaluation of Negotiation

Kraus in [79] mentions that evaluating the results of multi-agent negotiation is not an easy task. It is hard to say which solution is the best. Since the agents are assumed to be self-interested, when saying, for example, a "negotiation was successful" the question "successful for whom?" must be asked, since each agent is concerned only about its own benefits or losses from the resolution of the negotiation. Nevertheless, there are certain parameters that can be used to evaluate negotiations.

Negotiation Time: Negotiations which end without delay are preferred over negotiations which are time-consuming. It is assumed that a delay in reaching an agreement causes an increase in the cost of communication and computation time spent on the negotiation. We want to prevent the agents from spending too much time on negotiations resulting in deviation from their timetables for satisfying their goals.

Efficiency: An efficient outcome of the negotiations is preferred. In other words, an outcome that increases the number of agents which will be satisfied by the negotiation results and the agents' satisfaction levels from the negotiation results.

Thus, it is preferred that the agents reach Pareto optimal agreements[2] In addition, if there is an agreement that is better for all the agents than opting out, then it is preferred that the negotiations will end with an agreement.

Simplicity: Negotiation processes that are simple and efficient are better than complex processes. Being a "simple strategy" means that it is feasible to build it into an automated agent. A "simple strategy" also presumes that an agent will be able to compute the strategy in a reasonable amount of time.

Stability: A set of negotiation strategies are stable if, given that all the other agents included in the set are following their strategies, it is beneficial to an agent to follow its strategy too. Negotiation protocols which have stable strategies are more useful in multiagent environments than protocols which are unstable. If there are stable strategies, we can recommend to all agent designers to build the relevant strategies into their agents. No designer will benefit by building agents that use any other strategy.

Money transfer: Money transfer may be used to resolve conflicts. For example, a server may "sell" a data item to another server when relocating this item. This can be done by providing the agents with a monetary system and with a mechanism for secure payments. Since maintaining such a monetary system requires resources and efforts, negotiation protocols that do not require money transfers are preferred.

[2]An agreement is Pareto optimal if there is no other agreement that dominates it, i.e., there is no other agreement that is better for some of the agents and not worse for the others.

Chapter 3

EBDI: A Generic Architecture for Emotional Agent

This chapter presents EBDI, a generic architecture for emotional agents which implements both practical reasoning and emotional mechanisms. By taking into account the incompleteness of emotion theories and individual difference, I separate specific emotion mechanisms with in the agent, instead of trying to model emotion mechanisms to reflect some specific reasoning process, like most researchers do. EBDI can be used to build high EQ agents by selecting and implementing specific emotion theories into the architecture as needed. We show a sample EBDI agent in Tileworld, and the test results show that this EBDI architecture is applicable and that the emotional agent has better performance than rational agents.

Thus, the desirable features for this architecture are:

- It contains emotion status, the description method of which is not limited and depends on the specific application.
- It contains the function of emotion update mechanism, and also have emotion's effect on decision process. The function for emotion update mechanism should keep some flexibility such that we can apply part of emotion theory in it according to specific purpose.

- It reflects the practical decision process of human being, since this way is natural, harmonious and efficient.

- It is practically applicable, not a concept only.

- It is compatible with currently widely used agent architecture to take advantage of the traditional theories for rational agents.

The basic idea here is to choose a widely used agent architecture as a base, and expand it. There are some other researchers who have tried to expand traditional agents by adding emotion to them, however none of them can satisfy above descriptions. The closest candidate is [132], which present a Emotional-BDI architecture including internal representations for agents' capabilities and resources. Unfortunately, this paper does not clearly represent the difference between emotional agents and normal rational agents. The capabilities and resources of these agents are independent of emotions, as such, they cannot reflect the relationship between emotions and belief, and how emotions influence agents' decision making.

The traditional agent architecture chosen here is Belief-Desire-Intention (BDI) model, and the main concern is described in Section 3.1.

Taking into account both primary emotions and secondary emotions [27] into decision making process, we use primary emotions as the first filter for adjusting the priority of beliefs, such that the agents can speed up decision making. Meanwhile, we can also solve the problem of an agent's bounded computational resources. Secondary emotions can be then used to refine the decision when time permits. Instead of considering beliefs from perception only as in the standard BDI model, we add possible belief candidates directly from communication and contemplation. This is reasonable, because we obtain information not from perception only, we also get information from communication and our thinking. For agents in dynamic complex environment, communication and individual reasoning sometimes are more important way to obtain information, when the perception from environment is limited. The detailed EBDI architecture and interpreter are described as in Section 3.2. An example implementation is given in Section 3.3.

3.1 Why Based on BDI Model

Typically, there are four types of traditional agent architectures [177]:

- Logic based agents—in which decision making is realized through logical deduction;
- Reactive agents—in which decision making is implemented in some form of direct mapping from situation to action;
- Belief-desire-intention agents—in which decision making depends upon the manipulation of data structures representing the beliefs, desires, and intentions of the agent;
- Layered architectures—in which decision making is realized via various software layers, each of which is more or less explicitly reasoning about the environment based at different levels of abstraction.

For the logic based agents, decision making is predicated on the assumption of calculative rationality—the assumption that the world will not change in any significant way while the agent is deciding what to do, and that an action which is rational when decision making begins is still rational when it ends. However, most current multiagent systems can hardly guarantee a static and deterministic environment. The problems associated with representing and reasoning about complex, dynamic, possibly physical environments are unsolved.

Reactive agents make decisions based on local information, so they must have sufficient information available in their local environment for them to determine an acceptable action, and it is difficult to see how such decision making could take into account non-local information. On the other hand, for reactive agents, overall behavior emerges from the interaction of the component behaviors when the agent is placed in its environment, which suggests that the relationship between individual behaviors, environment, and overall behavior is not understandable, such that it is very hard to engineer agents to fulfill specific tasks.

The layered architectures are very general. The main problem is that while they are arguably a pragmatic solution, they lack the conceptual and semantic clarity of un-layered approaches. Another issue is that of interactions among layers, if each layer is an independent activity producing process, then it is necessary to consider all possible ways that the layers can interact with one another.

The original Belief - Desire - Intention model was developed by the philosopher Michael Bratman [16]. The BDI architectures reflect human's practical reasoning process, which is part of the reason why we choose it as a basic rational model for agent. It has shown to be a very successful architecture and it is attractive for several reasons: first, it has widely accepted philosophical roots; second, there are logical frameworks for modeling and reasoning about BDI agents; third, there are a considerable set of software systems which employ the architecture's concepts, such as PRS system [58]. The main problems about this architecture are: how to efficiently implement these functions; how to reach the balance between being committed to and over-committed to one's intentions. As they stand, BDI architectures ignore the influence of emotions in decision-making, while emotions do have influence on beliefs [54]). By adding the idea of primary and secondary emotions [27] we can then filter the decision making process of an agent.

3.2 EBDI Architecture

To incorporate emotions into agents, we need solve three problems: (1)How to measure or present emotions? (2)How do emotions affect the decision making process? (3)How to update the status of emotions? Most of the time, the details of the solutions depend on specific applications. Thus, our EBDI architecture combines these three concerns into a BDI architecture based on human's practical reasoning process, while leaving the details open to the designers.

3.2.1 Main Components and Functions

According to [183], practical reasoning involves two important processes: deciding what state of affairs we want to achieve – deliberation, and deciding how we want to achieve this state of affairs – means-ends reasoning. In the EBDI architecture we still follow this processes. We divide the process into four components: Emotion, Belief, Desire and Intention, and we connect these four components through some main functions.

Since no real agent has unlimited memory and can deliberate indefinitely, we assume that *resource bounds* are applicable to all above items, which means there are fixed amount of memory and a fixed processor available to carry out its computations. It also means that means-ends reasoning and deliberation must be carried out in a fixed, finite number of processor cycles, with a fixed, finite amount of memory space. For our architecture we assume that all emotions, beliefs, desires and intentions are stored according to some priority. For all processes, the emotion, belief, desire or intention with the highest priority is considered first. During information updating, if the resource bound is reached, the one with the lowest priority will be removed or replaced.

Emotion: There is no standard definition for emotions. Kleinginna [77] mentions that there are as many as 92 different definitions in the literature. Here we use the one that defines emotions as conscious states [90]. For this component, we did not limit the representation method of emotions, which can be stored as first-order logic, multidimensional logic [59], some numerical measurement method as PAD emotion scales [108], or something else.

Belief: Belief is usually defined as a conviction to the truth of a proposition. Beliefs can be acquired through perception, contemplation or communication. In the psychological sense, belief is a representational mental state that takes the form of a propositional attitude. Knowledge is often defined as justified true belief, in which the belief must be considered to correspond to reality and must be derived from valid evidence and arguments. However, this definition has been challenged by the Gettier problem [60] which suggests that justified true belief does not provide a complete picture of knowledge. Still, we believe the component of belief in the original BDI model is enough to cover the idea of resources added by David Pereira [132], that is, the resources mentioned can actually be looked as a kind of beliefs.

Practically, beliefs are subjective for humans. The original BDI model gets its beliefs from the *see* function, which senses the environment objectively. In our architecture, beliefs are influenced by the agent's emotional status and, instead of acquiring beliefs through perception only, we also add alternative methods to acquire beliefs through contemplation and communication.

Desire: Desires point to the options that are available to the agent, or the possible courses of actions available to the agent. Desires are obtained through an *option generation function*, on the basis of its current beliefs and intentions.

Intention: Intentions play a crucial role in the practical reasoning process, because they tend to lead to action. Wooldridge [183] summarizes four important roles in practical reasoning; here we modify them to five after taking emotions into account:

- *Intentions drive means-ends reasoning.* Like a human, once an agent has formed an intention, it will attempt to achieve the intention and decide how to achieve it; if one particular course of action fails to achieve an intention, then the agent will typically try others.

- *Intentions persist.* An agent will not give up its intentions until it believes that it has successfully achieved them, it is impossible to achieve them, or the reason for the intention is no longer present.

- *Current emotions influence the determination of intentions.* BDI agents determine their intentions based on their desires and beliefs. If the available options are equally reasonable, then a human making the decision might rely on emotions. Some research [18] points out that deliberative decision making may actually be driven by emotions, based on the observation that people with relatively minor emotional impairments have trouble making decisions. In our architecture, emotions set the priority of desires and help decide intentions.

- *Intentions constrain future deliberation.* In other words, an agent will not entertain options that are inconsistent with its current intentions.

- *Intentions influence emotions upon which, together with beliefs, future practical reasoning is based.* Once an agent adopts an intention it can plan for the future on the assumption that it will achieve the intention. Thus, if there is a belief that the agent cannot benefit from the intention, the agent will feel unhappy, and this emotion will also influence future beliefs and reasoning. BDI agents avoid *intention-belief inconsistency* [16] —the status of having an intention to bring about φ while believing that the agent will not bring about φ, because it is irrational. In contrast with the original BDI model, our model does not avoid intention-belief inconsistency completely, since intention did not influence beliefs directly, but indirectly through emotions. For example, when agent i has an intention to bring about φ, it is possible to have a belief that φ will not be true in our model: assume that this belief is obtained through a message from j, and i has negative emotion about j, such that i does not care about the belief very much. Later on, if such emotion becomes very strong, then i may remove this belief with the limitation of the resource bound; or if i's emotion about j changes to some positive one, and the belief becomes subjectively important to i, then the intention might be canceled. Thus, using emotions as a tool to balance intention and belief, such inconsistencies can be solved naturally.

More specifically, intentions in our architecture represent the agent's current focus—those states of affairs that it has committed to trying to bring about, and are affected by current emotional status together with current desires and working intentions.

We now formally define an EBDI architecture. Let E be the set of all possible emotions; B be the set of all possible beliefs, D be the set of all possible desires, and I be the set of all possible intentions. The state of an EBDI agent at any given moment is given by its current set of emotions, beliefs, desires, and intentions. These components are connected via the following functions:

Belief Revision Functions: Beliefs can be acquired from perception, contemplation or communication, unlike the original BDI model which uses only perception. We define three belief revision functions which map input from these three areas into beliefs. The input from perception is treated the same as in the BDI architecture. Input from communication is treated similarly but we take into consideration the identity of the sender. The input from contemplation comes from the agent's beliefs themselves and from deliberation. As with human beings, the beliefs related to the current intentions will be given higher priority and the rest will be given lower priority or ignored. For convenience, we combine effects of emotions and intentions together since they involve common issues about priority.

The three belief revision functions are defined as follows:

Belief Revision Function through Perception (*brf-see*) generates belief candidates from the environment:

$$brf\text{-}see : Env \rightarrow B_p$$

where Env denotes the environment, $B_p \subseteq B$ is the set of possible belief candidates from perception.

Belief Revision Function through Communication (*brf-msg*) generates belief candidates from the content of communication messages:

$$brf\text{-}msg : Content \rightarrow B_m$$

where $Content$ denotes the content of possible communication messages, B_m the set of possible belief candidates from message, and $B_m \subseteq B$.

Belief Revision Function through Contemplation ($brf\text{-}in$) takes into consideration the current emotion status[1] and intentions, and revises the current beliefs based on previous beliefs and the set of belief candidates from the environment and communication messages:

$$brf\text{-}in : E \times I \times (B \cup B_p \cup B_m) \rightarrow B$$

Emotion Update Functions: We take into account both primary emotions and secondary emotions [27], so we have two corresponding update functions.

Primary emotions are those that we feel first, as a first response to a situation. Thus, if we are threatened, we may feel fear. When we hear of a death, we may feel sadness. They appear before conscious thought and are thus instinctive or reactive functions of the human brain. When time is limited and we do not have enough time to think about something clearly, primary emotions become extremely useful in

[1]Frijda [54] shows emotions' influence on beliefs.

decision making. In agents, we can use primary emotions to speed up decision making similarly. The primary emotion update function (*euf1*) can be defined as:

$$\mathit{euf1} : E \times I \times (B_p \cup B_m) \rightarrow E$$

Secondary emotions appear after primary emotions, which may be caused directly by primary emotions, or come from more complex chains of thinking. For example where the fear of a threat turns to anger that fuels the body for a fight reaction. For agents, the secondary emotions come from the result of further deliberation and can replace the primary emotions. They are used to refine the decision making if time permits. The secondary emotion update function (*euf2*) is defined as:

$$\mathit{euf2} : E \times I \times B \rightarrow E$$

Option Generate Function: This function is similar to the one in BDI model. The option generate function (*options*) is defined as:

$$\mathit{options} : B \times I \rightarrow D$$

Filter Function: This function is also similar to the one in BDI model, however we add emotions, which are used to find the best option(s). The filter function (*filter*) is defined as:

$$\mathit{filter} : E \times B \times D \times I \rightarrow I$$

Plan Function: The functions above complete the process of deliberation and generate the best option(s)—intention(s), which can be some actions or states of mind. These intentions drive means-end reasoning, which can be represented by the plan function (*plan*):

$$\mathit{plan} : I \times Ac \rightarrow \pi$$

where Ac is the set of possible actions that the agent can do, and π denotes a plan which is a sequence of actions

$$\pi = (\alpha_1, \cdots, \alpha_n)$$

where α_i is an action, and $\alpha_i \in Ac$.

Plan Execution Function: This function is used to execute the sequence of actions produced by plan function. It is represented as

$$execute : \pi \to Env$$

During the execution, if π is empty, or the current intention succeeds, or the agent finds that the current intention is impossible to achieve, then this function will be terminated.

3.2.2 Architecture

Figure 3.1 shows our EBDI architecture. Figure 3.2 shows the interpreter that is used in our EBDI architecture. Its functions are described in the previous section.

We can summarize the execution cycle as follows:

1. When there is new information from the environment via sensor or communication messages, the EBDI agent generates belief candidates;
2. These belief candidates together with current intentions trigger emotion updating, that is, the agent obtains its first feeling about the information;
3. Based on the new emotion status and the new information, together with current intentions as a guide, the agent re-evaluates its beliefs;
4. From the beliefs and intentions, the agent generates desires;

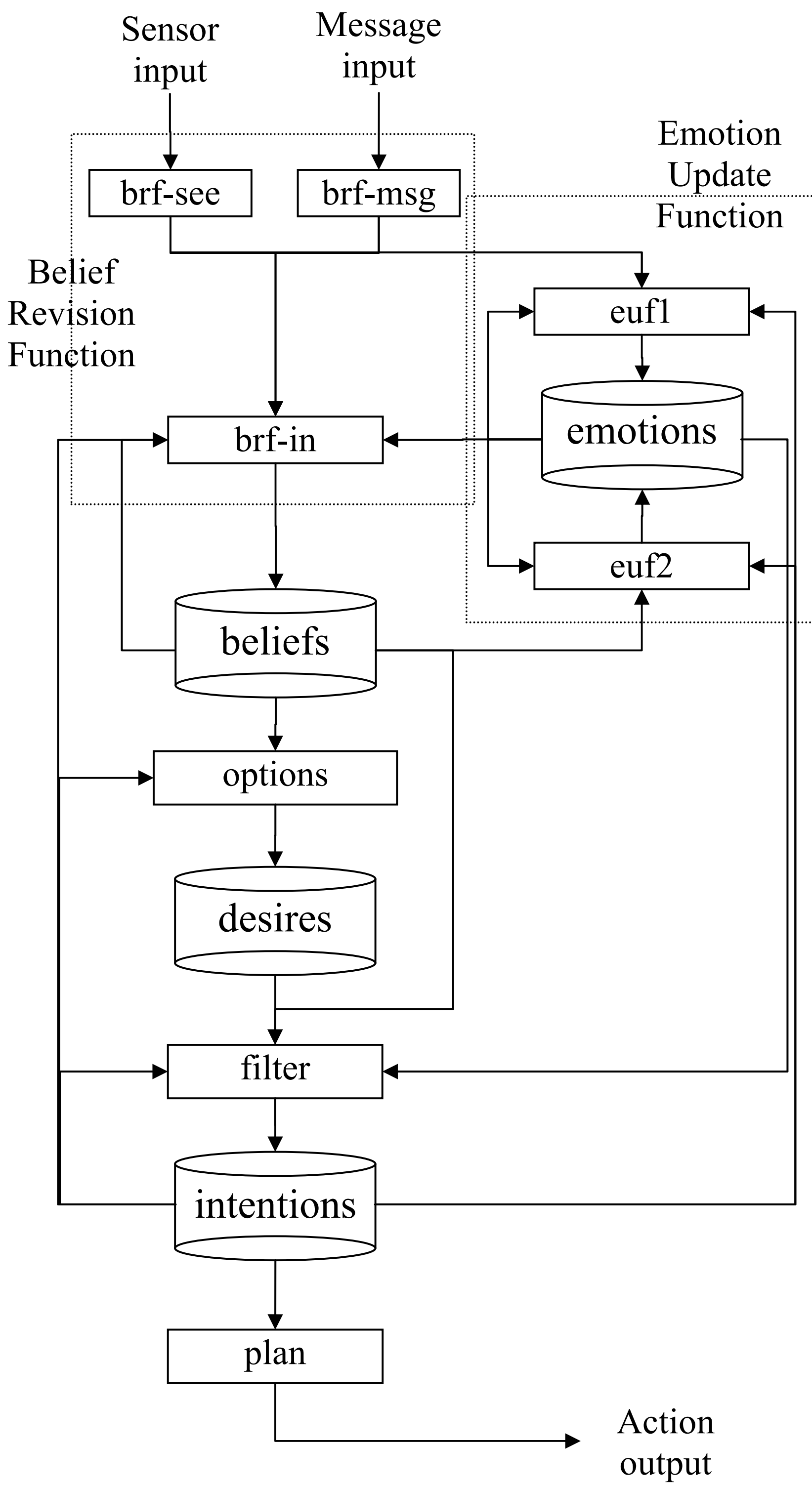


Figure 3.1: Schematic diagram of EBDI architecture.

EBDI-MAIN-LOOP

```
1  E ← E_0;                 ▷ E_0 are initial emotions
2  B ← B_0;                 ▷ B_0 are initial beliefs
3  I ← I_0;                 ▷ I_0 are initial intentions
4  while true
5       do B_p ← brf-see(Env);
6          B_m ← brf-msg(Content);
7          E ← euf1(E, I, B_p ∪ B_m);
8          B ← brf-in(E, I, B ∪ B_p ∪ B_m);
9          D ← options(B, I);
10         I ← filter(E, B, D, I);
11         E' ← E
12         E ← euf2(E, I, B);
13         if time permits and E ≠ E'
14            then B ← brf-in(E, I, B);
15                 D ← options(B, I);
16                 I ← filter(E, B, D, I);
17         π ← plan(I, Ac);
18         execute(π)
```

Figure 3.2: Pseudo-code of an EBDI agent's main loop.

5. Under influence of the emotions, the agent chooses the best options or intentions based on current beliefs, desires and intentions. Notice that, since intentions persist, the current working intention always has the highest priority unless it is already achieved or is found to be impossible to achieve, or the reason for this intention is no longer present.

6. From this deliberation result, the secondary emotions are triggered, and this updating is based on current intentions, beliefs and previous emotions.

7. If there is no time for deeper consideration or emotion status is not changed, the agent will directly generate a detailed plan and execute it. Otherwise, the agent begins a deeper consideration and refines the decision making. It will reconsider if current beliefs are suitable, as in line 14, and reconsider the desires and intentions, as in line 15 and 16. After this reconsideration, the agent then

generates a plan and executes it.

Our EBDI agent architecture thus manages to integrate emotions into the standard processing loop of a BDI agent.

3.3 EBDI Agent in Tileworld

The Tileworld system is an experimental environment for evaluating agent architectures [137]. We chose Tileworld as a platform for experimentally investigating the behavior of various metalevel reasoning strategies, since we can assess the success of alternative agent strategies in different environments by tuning the environmental parameters. We simulate the Tileworld in NetLogo [179], as in Figure 3.3. In it a number of agents must find and push tiles so they cover any open holes. Instead of consisting of only one simulated robot agent and a simulated environment, as in [137], we design several agents and show how to apply our EBDI model to one agent in a specific Tileworld, where both the agents and the environment are highly parameterized.

A Tileworld with EBDI agent is described as follows:

Environment: It is both dynamic and unpredictable. There are holes and tiles, which appear with some probability at a randomly selected location, live for awhile and then disappear. We design the size of the grid be 35×35 in the simulation.

Agents: The task for agents is to push the tiles to cover as many holes as possible. Each agent i tries to obtain the highest utility of its own $u_i = \textit{num-holes-filled}_i$, where $\textit{num-holes-filled}_i$ denotes the number of holes filled by agent i in the environment.

We make the following assumptions for agents:

- At each time step, an agent can only change its facing direction once or move one step.

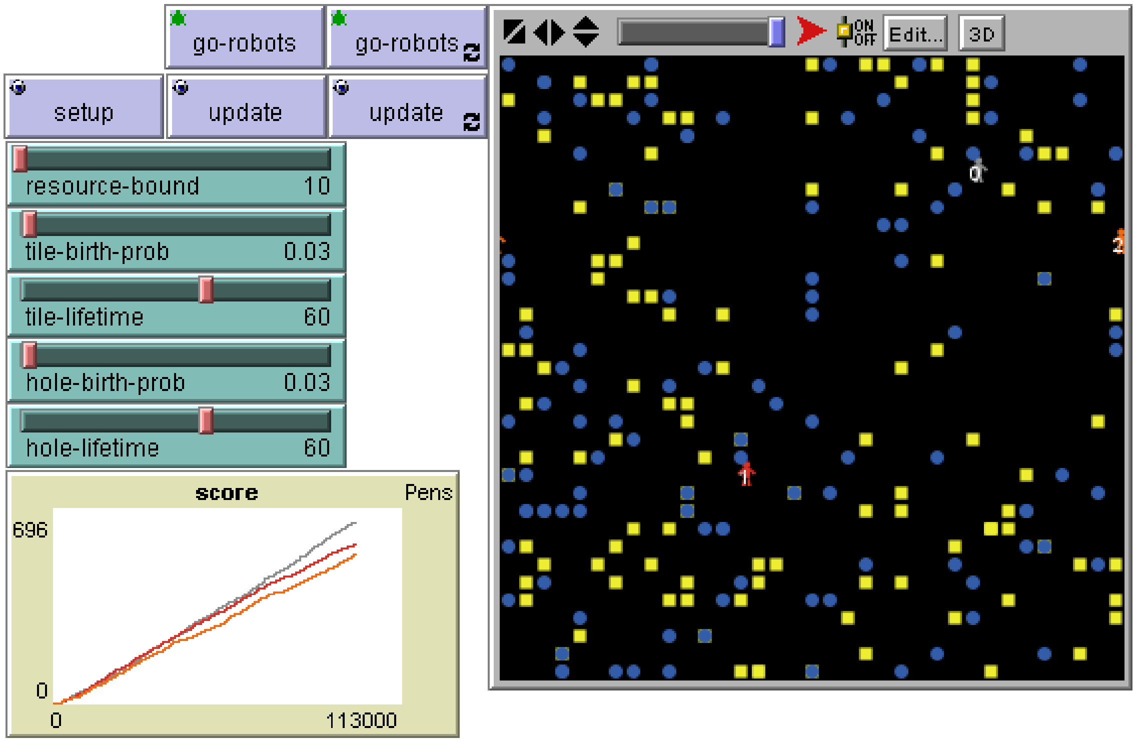


Figure 3.3: Simulation of Tileworld where the objects with person shape are agents, the yellow squares are tiles, and the blue circles are holes.

- Agents can move along four directions: north, south, east or west, and can move to a direction only when they face it.
- An agent can only see holes and tiles along one direction, however it can change its facing direction. For example, when an agent faces north, it can see all the holes and tiles in front of it, but can not see the holes and tiles at the east, west or south. Thus, the perception is limited and agents must partially rely on the information from communication.
- Agents can communicate with each other about what they see, but they may not tell the truth.
- An agent can save information about the location of holes and tiles that it sees or gets from communication messages as its beliefs.
- Considering the resource bound, we limit the storage space of beliefs for each

agent, such that each agent cannot store all information it sees or it receives from communication messages.

To make a comparison between an EBDI agent and a rational agent, we include three agents in the system: one EBDI agent and two rational BDI agents. These three agents have the same basic strategy in how to choose tiles to cover holes, which is described as follows:

1. First ask other agents for the information about the tile closest to its location;
2. Look around to get information about tiles and holes;
3. Deal with request from other agents;
4. Figure out the closest tile using beliefs;
5. Move toward the closest tile;
6. After reaching the tile, ask other agents for the information about the hole closest to its current location;
7. Look around to get information about tiles and holes;
8. Deal with request from other agents;
9. Find out the closest hole, and move the tile to the hole.
10. Repeat steps 1–9.

The differences between the three agents are as follows:

- EBDI agent: has a specific state to store emotion status, which is initially set to neutral. According to the basic strategy, if it decides to choose a belief told to it by some other agent i, then it will take actions just like the strategy mentioned. If it finds out later the hole or tile that the agent i mentioned is not there, it will think that i is lying and will hate i; conversely, if it finds out the information is correct, it will feel thankful to i. Later on, when this agent receives another message from agent i, it will correspondingly decrease or increase the priority the information from the agent i. Also, if this agent hates i to some degree it will lie to i about what it knows.

- Truth telling BDI agent: Always tells the truth when asked about a tile or hole.
- Selfish lying BDI agent: Always lies to other about tiles and holes so as to increase its chances of getting all the tiles into the holes.

In the simulated NetLogo program, the EBDI agent is labeled as "0", the truth telling BDI agent is labeled as "1", and the selfish lying BDI agent is labeled as "2", as in Figure 3.3. Here let's focus on the EBDI agent, and see how to apply the EBDI architecture to this agent in detail. We first consider the four main components.

Assume positive and negative emotions only are used for this agent, and we use positive and negative numbers to present the intensity of the emotions. Then the emotion status can be represented as the set

$$E = \{(Agent_i, e) | Agent_i \in Ag, e \in Z\}$$

where Ag denotes the set of agents in the system, e records emotion status, Z denotes the set of integers. In this specific example, $Ag = \{ag_0, ag_1, ag_2\}$ denotes the three agents corresponding to the label. $(Agent_i, e)$ means that current agent has a emotion e on $Agent_i$. For example, in an agent's emotion status set, there is an element $\{ag_1, -3\}$, which means the agent has negative emotion on ag_1 and the intensity is 3. The initial emotion status set $E_0 = \emptyset$, which means that in the beginning the emotion status for the agent is neutral. This emotion set is an example only, one can use different methods to present the emotional status.

The belief set stores the useful information about the Tileworld, which can be represented as

$$B = \{(Agent_i, obj, location) | Agent_i \in Ag,$$

$$obj \in \{tile, hole\}, location = (x, y)\}$$

where $x \in$ [*west-bound*, *east-bound*], $y \in$ [*north-bound*, *south-bound*]. The constants *west-bound*, *east-bound*, *north-bound*, *south-bound* are integers, and they satisfy *west-bound*<*east-bound*, *north-bound*< *south-bound*, which sets the boundary of the Tileworld environment where agents can move. $Agent_i$ is the agent who sends the message. If the belief comes from perception, then $Agent_i$ will be the agent itself. If the belief is from the communication message sent by agent j, then $Agent_i = ag_j$. For example, belief (ag_j, *tile*, $(3, 5)$) means that the agent obtained a belief from agent j that there is a tile at location $(3, 5)$. The initial belief set $B_0 = \emptyset$.

The desire set stores the agent's current desires (goals). For example, *find-tile* is the desire to find a tile, *find-hole* represents the desire to find a hole, and *carry-tile-to-hole(l)* represents the desire to carry a tile, which we assume the agent is carrying, to a hole at location l. The EBDI agent also has plans associated with each one of these desires. For example, a *find-tile* desire can be satisfied by either searching the space or asking other agents if they have seen any tiles. Since the agents start out with no tiles, initially they have $D_0 = \{$*find-tile*$\}$.

Intention is the agent's currently executing plan. Initially $I_0 = \emptyset$.

We now consider the main functions for the EBDI agent:

There are three belief revision functions:

brf-see gets the belief candidates from perception. For example, if the agent is located at $(3, 5)$ and it faces east, then the agent can see all the tiles and holes locate at $(x, 5)$, where $x \in (3,$ *east-bound*]. Assume there is a hole at $(6, 5)$, then the agent i obtains a belief candidate as $(ag_i, hole, (6, 5))$.

brf-msg obtains the belief candidates from communication messages. Assume agent i asked j about the closest hole to location $(3, 5)$, where i is located, and j returns a message to i that the tile at $(4, 4)$ is the closest one to $(3, 5)$ based on its beliefs. Then the agent i gets a belief candidate as $(ag_j, hole, (4,4))$.

brf-in considers current emotion status and intention as a guide to revising the belief set. For example, if $B = \{(ag_0, hole, (8,5))\}$, $B_p = \{(ag_0, hole, (6,5))\}$, and $B_m = \{(ag_1, hole, (4,4))\}$, if both the emotion status set and intention set are empty, the belief set will be $B = \{(ag_1,\ hole,\ (4,4)),\ (ag_0, hole, (6,5)),\ (ag_0, hole, (8,5))\}$, which orders the beliefs rationally according to the distance to the agent's current location $(3,5)$, such that the front one has the highest priority. If the current emotion status set has a member $(ag_1, -2)$ then this lowers the priority of belief $(ag_1,\ hole,$ $(4,4))$ and results in $B = \{(ag_0, hole, (6,5)),\ (ag_0, hole, (8,5)),\ (ag_1,\ hole,\ (4,4))\}$. If the intention set is not empty and the intention is to reach a tile at $(8,4)$, then the resultant belief set will be $B = \{(ag_0, hole, (8,5)), (ag_0, hole, (6,5)), (ag_1, hole, (4,4))\}$ because the agent's future location will be around $(8,4)$, and the hole at $(8,5)$ will be the closest one by then.

There are two emotion update functions:

euf1 considers the primary emotions. For example, if $E = \emptyset$, and $I = \emptyset$, there are $B_p = \{(ag_0, hole, (6,5))\}$, and $B_m = \{(ag_1, hole, (6,5))\}$, though the agent ag_0 gets duplicate information about the hole at $(6,5)$, it finds out ag_1 tells the truth, and thinks ag_1 is reliable, and then feels happy with ag_1. Thus, an emotion status will be generated as $(ag_1, 1)$. If there is already a emotion status $(ag_1, 1)$ in set E, then the emotion status in E will be updated to $(ag_1, 2)$.

euf2 considers secondary emotions. It works like *euf1* but it uses the current beliefs and intentions. For this simple case, we just ignore it in the simulation.

The *options* function generates new desires based on the agent's current beliefs and intentions. In this example, it mostly serves to generate a new *find-tile* desire after the agent drops its current tile.

The filter function makes a decision on the intention. For example, if the current intention of ag_0 is to find a tile at $(6,5)$, and this information is originally from agent

ag_2, such that $I = \{\textit{find-tile}(ag_2, (6,5))\}$. Assume ag_0 is currently at $(6,1)$, and there is emotion status set $E = \{(ag_2, -2), (ag_1, 2)\}$, and there is $\textit{find-tile}(ag_1, (5,1))$ in D, then ag_0 will change the intention to $I = \{\textit{find-tile}(ag_1, (5,1))\}$, since it trusts ag_1 more.

The plan function generates a sequence of actions based on the intentions. The possible action set in this example can be $Ac = \{turn(direction), move\ (direction)\}$, where $direction \in \{west, east, north, south\}$.

For example, if the ag_0 is currently located at $(6,1)$, faces east, and the current intention is to reach a tile at $(6,5)$, then it may generate a sequence of actions: $\pi = (turn(south), move(south), move(south), move(south))$

For the plan execution function, basically, the agent just follows the sequence of the actions π. Note that every time the agent turns to some direction it can see some new tiles and holes which can trigger the agent's reconsideration.

The above example shows how might build an EBDI agent for the Tileworld. Based on above descriptions about the main components and main functions, the main execution cycle can just follow the interpreter as in Figure 3.2.

We compare the performance of the three agents and test it for dozens of time, all the results are similar, one of which is shown in Figure 3.4, where we set the birth probability be 0.03 and the lifetime be 60 cycles for both tiles and holes. It shows that at the beginning, the differences between the performance are hard to tell, since the information is very limited and every agent does the task by chance; but the emotional agent gets better and better and always ends up with the highest score. Thus, the emotional agent here has better performance than rational agents because it is more adaptive in this dynamic environment.

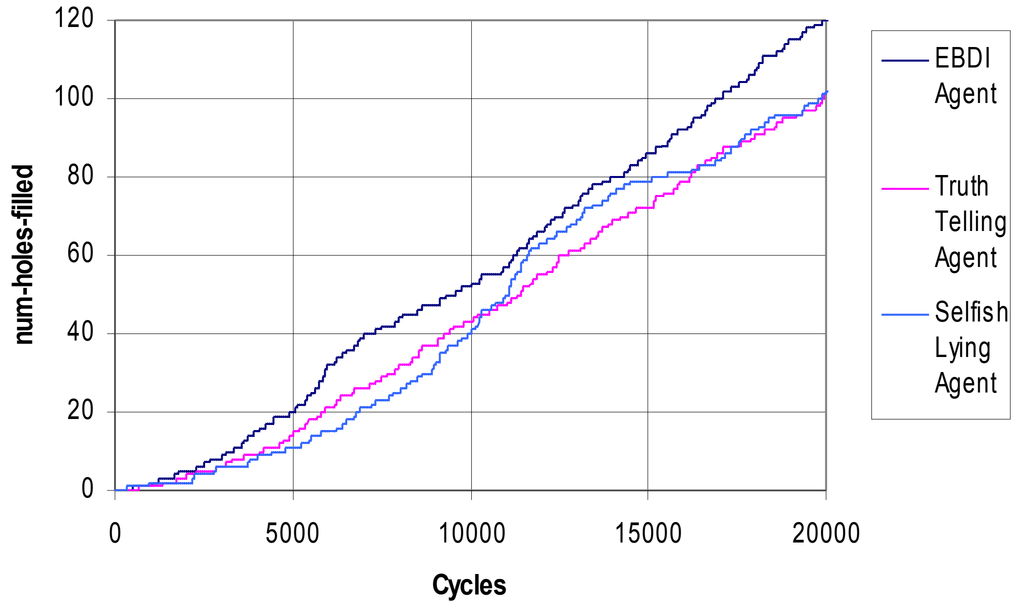


Figure 3.4: Testing Result

3.4 Communicating Emotions among Agents

Communication is an important issue for agents' application. For emotional agents, a necessary concern is how to communicate emotions. Meanwhile, modern agent communication languages (ACLs), as a critical element of multiagent systems and a key to the successful application of agents in commerce and industry, such as the FIPA ACL, already have wide and successful applications. Instead of rebuilding a completely new communication system to work with the emotional agents, can we reuse the existing communication system? Then how can we do that?

This section gives two possible approaches.

3.4.1 Designing Ontologies for Emotional Agents

Current modern ACLs have been designed as general purpose languages to simplify the design of multiagent systems. So the easiest way is to reuse current communication system and let people who want to communicate emotions define their own ontologies and then use only the ones they need. This way is simple, and agents only need to understand a few ontologies. The problem is that there is no universal accepted emotions, so the agents who send the message need to make sure the receiver understand the ontologies it use in advance. However, this may be convenient for small group users and may result in more polychrome and diversity emotions.

3.4.2 Extending Communicative Acts of Modern ACLs

By analyzing the structure of the modern ACLs, we can see that the early work on communicative act based ACLs, such as KQML and FIPA, separated the communication problem into three layers – a message transport layer providing the mechanics of a communication, a domain-independent layer of communication semantics, and a domain-dependent content layer. The ACL speech acts were intended to describe the domain-independent middle layer.

Meanwhile, recent research shows that these ACLs do not support adequately all relevant types of interactions. Serrano and Ossowski [156] report a need for new *ad hoc* sets of performatives in certain contexts, which the FIPA ACL does not support. Singh [162] points out that agents from different venders or even different research projects cannot communicate with each other. In [76], Kinny shows that FIPA reveals a confusing amalgam of different formal and informal specification techniques whose net result is ambiguous, inconsistent, and certainly underspecified communication. He also proposes a set of requirements and desiderata against which an ACL specification can be judged, and briefly explores some of the shortcomings of the FIPA ACL and its original design basis.

Therefore, a complete set of communicative acts in an ACL would be desirable in order to improve understanding among the agents in a multiagent system. Thus, this complete set should also cover our requirement for emotions.

The idea is to keep the structure of the current ACLs, while broaden the semantic coverage of ACLs by formalizing speech act categories that cover the ~4800 speech acts in Ballmer and Brennenstuhl's book [10]. This book proposes an alternative classification of speech acts, which contains both simple linguistic functions such as expression and appeal, and more complex functions such as interaction and discourse. Models for alternative actions are formed and verbs are classified according to the phases of the model. Since the classification is based on an almost complete domain (~4800 speech acts) and the authors claim they provide a "theoretically justified" classification that is "based explicitly and systematically on linguistic data", we believe that to generate a speech act set for ACLs based on their classification will be a powerful way to represent meaning.

The purpose of this work is not for emotions only, but it can solve our problem here. The detailed work is described in [71, 72], also in Appendix A.

Chapter 4

Plug-in: Emotion Update Mechanism

The OCC model, a computational emotion model, developed by Ortony, Clore and Collins [125], has established itself as the standard model for emotion synthesis. There are a large number of studies that employed the OCC model to generate emotions for their embodied characters. However, it is criticized for falling short on suggestions on what to do with the emotional state, and Bartneck [11] points out that the OCC model should be simplified to match the abilities of the embodied emotional character.

This chapter describes how the OCC model can be incorporated into an EBDI architecture as a plug-in. Since the EBDI architecture involves two emotion update functions, correspondingly, it simplifies the OCC model by dividing it into two parts, one fitting into the primary emotion update function, and the other fitting into the secondary emotion update function, as shown in Figure 4.1. Meanwhile, the original OCC model focuses on the types of emotions, and our research focuses on the process of how to generate those emotions. It provides a sample application approach for the EBDI architecture, and EBDI also makes up for a shortcoming of OCC by applying extra emotion effect functions.

The details of incorporating the OCC model into the EBDI architecture are de-

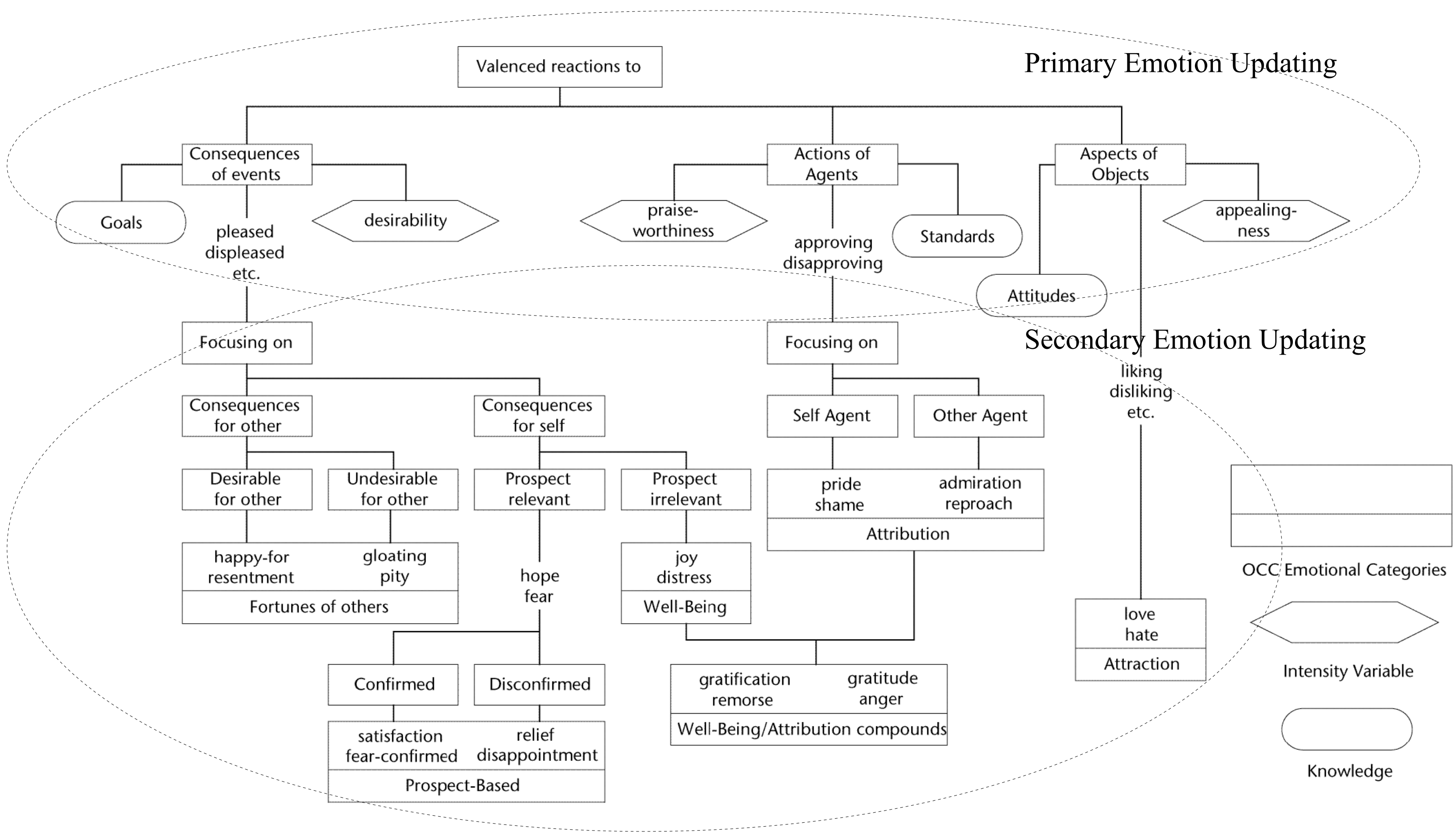


Figure 4.1: Separation of the OCC Model

scribed in the following sections.

4.1 Primary Emotion Update Function

Primary emotions are those that we feel first, as a first response to a situation. Thus, if we are threatened, we may feel fear. When we hear of a death, we may feel sadness. They appear before conscious thought and are thus instinctive or reactive functions of the human brain. When time is limited and we do not have enough time to think about something clearly, primary emotions become extremely useful in decision making. In agents, we can use primary emotions to speed up decision making similarly.

The OCC model assumes that there are three major aspects of the world, or changes in the world, upon which one can focus. The three aspects are events, agents, and objects. One focuses on events when one is interested in their consequences; one focuses on agents, because of their actions; and one focuses on objects, when one is interested in certain aspects or imputed properties of them. So, emotions are valenced reactions, and any valenced reaction is always a reaction to one of these perspectives on the world. Based on this assumption, we have discussions around these three aspects: consequences of event, actions of agents and aspects of Objects

Reactions to Events

The primary emotions of this aspect are states of feeling that arise from attending to some events, which are appraised as being desirable or undesirable. Desirable here describes the feeling of pleasure about a desirable event, and undesirable describes the feeling of displeasure about an undesirable event.

Reactions to Agents

Taking into account the question of how we believe salient events to have come about, we need also to focus on an agent who is instrumental in the event, instead of the event itself. When the agent is judged to have done something praiseworthy, the person experiencing the emotion is inclined to approve of the agent's action, and when the agent is held to have done something blameworthy, the experiencer is inclined to disapprove of the agent's action.

Reactions to Objects

The primary emotions here are momentary reactions of liking and disliking. Specifically, one reaction corresponds to liking or attraction and is occasioned by reacting positively toward some appealing object, and the other corresponding to aversion or dislike, is occasioned by reacting negatively toward some unappealing object.

Thus, according to above description, we can define the reactions to events as a function:

$$\mathit{ReactEvent} : I \times b_{event} \rightarrow v_d$$

where I is the current intention set; b_{event} is a belief candidate related to event or with the type be event; and v_d is the intensity variable, here described by desirability.

The reactions to agents can be defined as:

$$\mathit{ReactAgent} : I \times b_{agent} \rightarrow v_p$$

where I is the current intention set; b_{agent} is a belief candidate about agents or with the type be agent; and v_p is the intensity variable, here described by praiseworthiness.

The reactions to object can be defined as:

$$ReactObject : I \times b_{object} \rightarrow v_a$$

where I is the current intention set; b_{object} is a belief candidate about some object or with the type be object; and v_a is the intensity variable, here described by appealingness.

Then, the primary emotion update function can be represented as in Figure 4.2:

```
euf1 (E, I, B)
    ▷ E: the set of current emotion status;
    ▷ I: the set of current intentions;
    ▷ B: the set of new belief candidates from environment
    (through perception or communication).
    for each b ∈ B
        do type ← CheckBelief(b);
           ▷ type ∈ {event, agent, object}
           if type = event
              then v_d ← ReactEvent(I, b);
                   E ← Mapping(E, v_d);
                   break;
           if type = agent
              then v_p ← ReactAgent(I, b);
                   E ← Mapping(E, v_p);
                   break;
           if type = object
              then v_a ← ReactObject(I, b);
                   E ← Mapping(E, v_a);
                   break;
```

Figure 4.2: Pseudo-code of Primary Emotion Update Function

Where the function $CheckBelief(b)$ is used to check the type of the belief b, and the returned value is a *type* with value *event*, *agent* or *object*. The function $Mapping(E, v)$ is used to map a specific intensity variable to an emotion status, and update current emotion set E.

4.2 Secondary Emotion Update Function

Secondary emotions appear after primary emotions, and may be caused directly by primary emotions or come from more complex chains of thinking. For example, the fear of a threat turns to anger that fuels the body for a fight reaction. For agents, the secondary emotions come from the result of further deliberation and can replace the primary emotions. They are used to refine the decision making, if time permits.

Reactions to Events

According to the events relative to one's own goals, the secondary emotions of this aspect are grouped as three kinds: well-being emotions, fortunes of others, and prospect-based emotions. According to whether the person who experiences the emotions is reacting to the consequences of the focused event with respect only to himself, or also with respect to some other person, the secondary emotion update process continues the primary emotions and is separated into two branches: Consequences for Self and Consequences for Other.

The Consequences for Self branch leads to two groups of reactions: (1) the prospect of an event is crucial, and (2) the prospect is irrelevant. The emotions for which the consideration of the prospect is irrelevant reflect upon one's well-being, and the resulted emotions are something like happiness, joy, unhappiness, sadness, and distress. The emotions follow the group Prospects Relevant, which result from reacting to the prospect of positive and negative events, such as hope and fear, respectively. Upon whether the prospect of a positive or negative event is believed to have been confirmed or disconfirmed, four additional emotions can be generated: satisfaction, disappointment, relief, and fear-confirmed.

The other branch, Consequences for Other, generates emotions that are referred to as the Fortunes-of-others group. There are four distinct emotion types, representing the reactions to events that a person can have when the events are desirable or undesirable relative to the goals and interests of another person. Desirable for Other leads to *happy-for* and *resentment*; while Undesirable for Other leads to *gloating* and *pity*.

Reactions to Agents

Depending on whether the approval or disapproval focuses on the agent itself or other agents, the secondary emotions here are separated into two groups. When the formal agent is itself, the emotion types of *pride* and *shame* can arise. If the formal agent is some other agent, the emotion types of *admiration* and *reproach* can arise. In Figure 4.1, there are another group of emotions labeled "Well-being/Attribution compounds", which include *gratification*, *gratitude*, *remorse*, and *anger*. This group arises from simultaneously focusing on both the action of an agent and the resulting event and its consequences.

Reactions to Objects

The secondary emotions of this kind are ignored, since the reactions to objects are momentary reactions and the emotions of this kind appear to be more immediate, more spontaneous, and less affected by accessible cognitive processes than almost all other emotions. In other words, they are already included in the primary emotion update function.

Thus, the secondary emotion update function can be represented as in Figure 4.3: Where $CheckConsequences(b)$ is to check the belief b and find out if the consequences are for other or self. $CheckDesirable(I, b)$ is to check the belief b and find out if the consequences for other are desirable or not. $FortuneOther(b, v)$ generates emotions based on belief b and the desirability, if it is desirable, then generates happy-for or *resentment*; otherwise generates *gloating* or *pity*. $CheckProspect(I, b)$ is used to check if the belief b is prospect relevant or irrelevant. $Prospect(b, v)$ generates emotions

```
euf2 (E, I, B)
    ▷ E: the set of current emotion status;
    ▷ I: the set of current intentions;
    ▷ B: the set of current beliefs.
for each b ∈ B
    do type ← CheckBelief(b);
       ▷ type ∈ {event, agent, object}
       if type = event
          then if CheckConsequences(b) = other
                  then v ← CheckDesirable(I, b)
                       E ← FortuneOther(b, v)
                  else v ← CheckProspect(I, b)
                       E ← Prospect(b, v)
                       if v = relevant
                          then v_c ← CheckConfirm(I, b)
                               E ← ProspectBased(b, v_c)
                          else E ← Compounds(b, B)
       if type = agent
          then v ← CheckAgent(I, b);
               E ← Attribution(b, v);
               E ← Compounds(b, B);
```

Figure 4.3: Pseudo-code of Secondary Emotion Update Function

based on belief b and the relevance, if it is prospect relevant, then generates *hope* or *fear*; otherwise generates *joy* or *distress*. $CheckConfirm(I, b)$ is to check the prospect relevant belief confirmed or disconfirmed, upon which $ProspectBased(b, v)$ generates emotions *satisfaction* or *fear-confirmed* for confirmed one, and *relief* or *disappointment* for disconfirmed one. $Compounds(b, B)$ is to generate emotions compounded by well-being and attribution. $CheckAgent(I, b)$ is to check if the formal agent is itself or other. $Attribution(b, v)$ is to generate emotions based on self agent or other agent, described by v.

During practical application, the 22 emotions defined in the OCC model can be simplified according to specific representation method for emotion status and the ability to take action. The reasoning process of the EBDI agent can be described briefly as follows: Once the agent perceives some changes or gets some news from the environment, the primary emotion update function will be triggered, and the emotion status is updated; Upon an updated emotion status, the agent updates beliefs, generates desires, and intentions, and leads to action; meanwhile, the secondary emotion update function continually updates the emotion status, if the time permits, the agent will reconsider the beliefs, desires, and intention, and take action; otherwise, the reasoning process will be interrupt, and the action will be based on the current intention.

4.3 Summary

This chapter describes how the OCC model, which is employed frequently to generate emotions for embodied characters, can be incorporated into an EBDI architecture. In a more general sense, the OCC model supplies a needed emotion update mechanism for the EBDI architecture, and provides a guide for applying the EBDI architecture. EBDI also makes up the shortcoming of the OCC model by applying extra emotion effect functions. It simplifies the OCC model by dividing it into two parts, one fitting into the primary emotion update function, and the other fitting into the secondary emotion update function. When time is limited, we run the primary emotion update function, and ignore the other part of the OCC model and lead the agent to action. By applying the secondary emotion update function, we also keep the full characters of the OCC model. Thus, rather than being used for emotional negotiation only, this plug-in can be applied to many other problem domains as well.

Chapter 5

Emotional Negotiation

For human beings, negotiations often evoke a variety of emotions. Emotions can cause intense and sometime irrational behavior, and can cause conflicts to escalate and negotiations to break down [3]. The research shows that emotions play positive and negative roles in negotiation. On the positive side, emotions make us care for our own interests and about people. Empathy can improve understanding and facilitate communication; both hiding emotions and making vigorous displays of emotion can be effective negotiating tactics. Legitimately expressed anger may communicate the party's sincerity and commitment. On the other hand, fear and anger usually play negative roles in negotiation. Another researcher Li and Roloff [93] suggests that while positive emotion leads to cooperation and greater joint gain, negative emotion leads to competition and greater individual gain because of the different cognitive processing styles associated with each. Thus emotion does influence negotiators' cognitive processing, and then affect negotiation outcomes for human beings. However, we are not aware of any research that has tried to incorporate emotional models into automated negotiation yet.

Perfectly rational agents are only affected in their negotiation by features of the problem that directly impact their utility of the resulting deal. However, humans are not perfectly rational and often let their emotions, even those that are unrelated to the negotiation problem, affect their negotiation strategy.

There are many difficulties in incorporating emotional models into automated negotiation. Some main problems are listed as follows:

- How to present the emotion status in negotiation? Or how can we measure the emotions?

- What is the relationship of the emotions with the change of negotiation process?

- And how to correctly reflect this relationship?

- How to convert the affects of emotions into negotiation actions?

This chapter describes a mechanism that shows how emotions can be incorporated into an agent to make its behavior more similar to that of a human experiencing a particular emotional state. It starts by presenting the assumptions upon which the agents will be negotiating. And then introduces the Emotional Worth-Oriented Domain (EWOD), as in Section 5.2. It is a general model based on Worth-Oriented domain.

5.1 Assumptions

Human beings have a value system: things can be valued by given a price, and people compare the values of things by comparing the prices. For example, when you want to buy a house, an appraisal company may give you a price for the house according to where it is located, the structure age, square feet, conditions, other comparable houses' values, etc. Most people who want to buy the house may not really care about how it is appraised but they do care about the value of the house. That is, everyone has their own valuation of every item.

Similarly, we can assume that agents have value function which maps items to the value, which we represent with a real number, the agent ascribes to that item.

Thus, we can simplify the solutions of problems in multiagent system by considering them as some functions on the value system. Still using above example, the appraisal company can be considered as an agent, and this agent's task is to find out the value of the house to the average agent. Every items it considered, such as where it is located, the structure age, etc. are represented as some features, and finally there is some evaluation function to combine all kind of features, and output a value–the price of the house to the agent.

Here, we assume that our agents have utility functions that capture their preferences over possible deals. As such, agents can value the same item differently, which can lead to negotiation. For example, agent A offers agent B a watch with price \$10; B may think it's too expensive and that \$4 would be more reasonable, so then negotiates with A for \$4; A thinks \$4 is not acceptable and asks for \$8, and so on. Non-emotional agents are typically assumed to have fixed utility functions. However, a human's utility valuation can change due to their emotional state, and an agent's should as well.

Meanwhile, emotional behaviors are different from rational ones, but they are not in complete conflict. By adding emotions between the behavior and the unrelated reason, many things become easy to explain. For example, suppose an agent A gets a gift from a friend B today that makes him very happy. When people are in a happy mood they are more willing to help others. C asks A for help. Usually A rejects C, but today A gives C the help that C requests. There is no direct relation between the fact that A gets a gift from B and that A helps C, but by adding emotions we can explain it. Usually the effect of emotions is rational, in other words, there is often a reason why people are happy or sad. On the other hand, from emotion to behavior, there are also some rules to follow.

Emotions do have some effect on people's behavior. However, these effects are usually ignored in automated negotiation protocols. Correspondingly, some features that do not seem to be directly related are also ignored. By adding emotions we can better model the outcome of real human negotiations.

5.2 Emotional Worth-Oriented Domain (EWOD)

Based on above assumptions, let's consider the general models of negotiations. Rosenschein and Zlotkin [142] had a great contribution in this area to introduce a distinction between different types of negotiation domain: task-oriented domains and worth-oriented domains. In the task-oriented domains, the tasks are explicitly defined in the encounter: each agent is given a set of tasks to accomplish, associated with which there is a cost. An agent attempts to minimize the overall cost of accomplishing these tasks. Worth-oriented domain is a more general domain: the goals of an agent are specified by defining a worth function for the possible states of the environment, and the goal of the agent is thus implicitly to bring about the state of the environment with the greatest value. As mentioned in [183], unlike task-oriented domains, agents negotiating over worth-oriented domains are not negotiating over a single issue: they are negotiating over both the state that they wish to bring about and over the means by which they will reach this state. The task-oriented domain is a special case of worth-oriented domain.

Without losing generality we focus on the model of worth-oriented domains (WOD), and take into account the emotions, modify the model to be a new one—Emotional Worth-oriented Domain.

The formal description of this model is a tuple:

$$< E, Ag, J, c, r_e >$$

where

- E is the set of possible environment states;

- $Ag = \{1, ..., n\}$ is the set of possible agents;

- J is the set of possible joint plans, which are called joint plans because executing one plan can require several different agents. A joint plan can be represented as $j : e_1 \rightarrow e_2$, which means that the plan j can be executed in state e_1, and when executed in this state, will lead to state e_2. If the plans are not joint, but can be done by one agent, then it falls to task-oriented domain, and J will be the set of task assignments.

- $c : J \times Ag \rightarrow \Re$ is a cost function, which assigns to every plan $j \in J$ and every agent $i \in Ag$ a real number which represents the cost $c(j, i)$ to i of executing the plan j.

- r_e could be a function with time, or a constant, which represents the emotion degree of an agent i. The range is from 0 to 1. For example, for a completely rational agent, $r_e = 0$; for a completely emotional agent, $r_e = 1$.

An encounter in this model is a tuple:

$$< e, W, W_e >$$

where

- $e \in E$ is the initial state of the environment;

- $W : E \times Ag \rightarrow \Re$ is a worth function, which assigns to each environment state $e \in E$ and each agent $i \in Ag$ a real number $W(e, i)$ which represents the value, or worth, to agent i of state e.

- $W_e : S_e \times E \times Ag \rightarrow \Re$ is a emotional worth function, which gives worth affection of current emotional status , represented by an emotional state function S_e, to each environment state $e \in E$ and each agent $i \in Ag$. It is a real number too.

Reaching agreement involves the agents negotiating over the collection of joint plans. Agents try to reach agreement on the plan that brings about the environment state with the greatest worth. The optimal plan j^i_{opt} will then satisfy the following equation:

$$j^i_{opt} = \arg \max_{j:e_0 \rightarrow e \in J} r_e \cdot W_e(S_e, e, i) + W(e, i) - c(j, i)$$

This equation involves three parts: emotional affection on the worth value, rational worth value, and cost. Similar to task-oriented domain, this equation still tries to maximize the utility, however this utility involves emotion feature and its affection on utility.

5.3 Summary

This chapter simply describes the concept of EWOD by extending a general model - Worth-Oriented Domain. This extension gives us a mechanism that shows how emotions can be incorporated into an agent. Based on EWOD, we can then continually discuss the detail negotiation strategies, and see how would emotions work on negotiation strategies.

Chapter 6

Mapping Descriptive Emotions to Numerical Emotions

As described in Chapter 5, the decisions of automated negotiation are often based on the utility or some worth function, which is numerical. Then the measurement for the emotions is preferred to be numerical as well. Chapter 4 describes a mechanism for updating emotions in an EBDI agent, which is designed as a plug-in. It generates three pairs of emotions for primary emotion update function, and 22 emotions for secondary emotion update function. These emotions are descriptive. So, how can we map these emotions to numerical emotions such that we can apply them into the model of emotional negotiation?

This chapter solves the problem how to measure emotions in negotiation. Here we first describe a numerical measurement for emotions – PAD (Pleasure - Arousal - Dominance) emotional scale in Section 6.1. Then we show the detailed mapping in Section 6.2.

6.1 Pleasure-Arousal-Dominance Emotional Scale

The PAD (Pleasure-Arousal-Dominance) emotional state model is a general but precise three-dimensional approach to measuring emotions. Mehrabian [104] reviews versions of the PAD scales with different dimensions, and lists sets of studies that report development and refinement of a final set of the scales and consistently yield three nearly orthogonal dimensions: pleasure - displeasure, arousal - nonarousal, and dominance - submissiveness. "Pleasure - displeasure" distinguishes the positive - negative affective quality of emotional states, "arousal - nonarousal" refers to a combination of physical activity and mental alertness, and "dominance - submissiveness" is defined in terms of control versus lack of control. The analysis shows that these three dimensions provide a parsimonious base for the general assessment of emotional states.

Specific terms describing emotions can be visualized as points in a three-dimensional PAD emotion space. Alternatively, when the PAD scale scores are standardized, each emotion term can be described succinctly in terms of its values on the pleasure-displeasure, arousal-nonarousal, and dominance-submissiveness axes. The following sample ratings illustrate definitions of various emotion terms when scores on each PAD scale range from -1 to +1: angry (-.51, .59, .25), bored (-.65, -.62, -.33), curious (.22, .62, -.01), dignified (.55, .22, .61), elated (.50, .42, .23), hungry (-.44, .14, -.21), inhibited (-.54, -.04, -.41), loved (.87, .54, -.18), puzzled (-.41, .48, -.33), sleepy (.20, -.70, -.44), unconcerned (-.13, -.41, .08), violent (-.50, .62, .38).

Thus, according to ratings given for "angry," it is a highly unpleasant, highly aroused, and moderately dominant emotional state. "Sleepy" consists of a moderately pleasant, extremely unaroused, and moderately submissive state, whereas "bored" is composed of highly unpleasant, highly unaroused, and moderately submissive components.

Within the PAD Model, there are eight basic and common varieties of emotion, as defined by all possible combinations of high versus low pleasure (+P and –P), high versus low arousal (+A and –A) and high versus low dominance (+D and –D), as in Table 6.1. Thus, for instance, Anxious (–P+A–D) states include feeling aghast, bewildered, distressed, in pain, insecure, or upset; hostile (–P+A+D) states include feeling angry, catty, defiant, insolent, and nasty; and exuberant (+P+A+D) states include feeling admired, bold, carefree, excited, mighty, and triumphant. By focusing on individuals, extroverted, arousal seeking, exhibitionistic, nurturing, and affiliative persons are exuberant (i.e., pleasant, arousal, dominant). However, they may differ in terms of the weights of Trait Pleasure (P), Trait Arousability (A), and Trait Dominance (D) associated with each. Dependent persons are pleasant, arousal, and submissive. Anxious or neurotic persons are unpleasant, arousal, and submissive, whereas aggressive persons are unpleasant, arousal, and dominant (e.g., [105]). Mehrabian [106] provided equations showing relationships of specific personality measures to the PAD temperament dimensions.

Table 6.1: Basic and Common Varieties of Emotion in PAD

PAD Labels	Varieties of Emotion
(+P+A+D)	Exuberant
(+P+A–D)	Dependent
(+P–A+D)	Relaxed
(+P–A–D)	Docile
(–P+A+D)	Hostile
(–P+A–D)	Anxious
(–P–A+D)	Disdainful
(–P–A–D)	Bored

Some sample emotion terms grouped according to Pleasure, Arousal, and Dominance value are given in [104], where the preceding eight groups of emotion terms are derived from ratings of 240 emotion terms with the preliminary state pleasure, state arousal, and state dominance scales. Table 6.2 summarizes the appendix of [104] and some other samples in documents. The verbally described situations that represent a balanced sample of emotional experiences are listed in Appendix B.

Table 6.2: Sample Emotion Terms Grouped on PAD Value

PAD Type	Sample Emotion Terms
(+P+A+D)	Admired, bold, carefree, dignified, elated, excited, masterful, mighty, triumphant
(+P+A–D)	Amazed, curious, fascinated, grateful, impressed, loved, respectful
(+P–A+D)	At ease, comfortable, relaxed, satisfied, secure, unperturbed
(+P–A–D)	Consoled, cruel-admired, docile, domineering-timid, humiliated-lonely, protected, reverent, sleepy, tranquilized
(–P+A+D)	Angry, catty, cruel, defiant, hostile, insolent, nasty, unmotivated-distressed, violent
(–P+A–D)	Aghast, bewildered, distressed, fear, frustrated, hungry, in pain, insecure, neuroticism, puzzled, anxiety, upset
(–P–A+D)	Amazed-daring, awed-domineering, disdainful, humiliated-sad, indifferent, selfish-uninterested, uncaring, unconcerned
(–P–A–D)	Bored, despairing, fatigued, inhibited, lonely, sad, sluggish, subdued

These scales have wide-ranging applications [109]. They are used to assess consumer reactions to products, services, and shopping environments. Additionally the scales can be used to assess the emotional impact of a workplace, an advertisement, or a medical or psychotropic drug. For example, Mehrabian in paper [111] uses PAD value to analyze product preferences of consumers; in paper [107] he illustrates the power of the PAD Emotion Model by analyzing the Positive-Affect, Negative-Affect scales and showing the greater ease with which the PAD framework can differentiate depression from anxiety; Valdez [173] uses PAD Emotion Model for easy and comprehensive study of a large and complex area of research; Mehrabian [108] applies the PAD Emotion Model to broad fields of worker and marital satisfaction and shows how existing findings in the literature and new findings from the study can be conceptualized and summarized easily. There are also some applications in study of shopping environments and evaluating consumer reactions to services and products [109], etc. Recently, there has been some effort to incorporate PAD in Artificial Intelligence

[110], but the research is in its infancy. Thus, we can use this established method to record emotion status.

6.2 Mapping Methods Description

According to Chapter 4, the plug-in by OCC model generates three groups of primary emotions: pleased / displeased responding to consequences of events; approving / disapproving responding to actions of agents; liking /disliking responding to some aspects of objects. They actually show some positive or negative trends corresponding to different objects - events, agents, or aspects of objects.

For the secondary emotions, it generates 22 more specific emotions: happy-for, resentment; gloating, pity; hope, fear; satisfaction, fear-confirmed; relief, disappointment; joy, distress; pride, shame; admiration, reproach; gratification, remorse; gratitude, anger; love, hate.

Meanwhile, we have 8 varieties of PAD emotions, as in Table 6.1, and some sample emotions, as in Table 6.2.

Then, how can we map above emotions to PAD values? Including above resources, other resources we can use are some online tools, such as WordNet [114]. WordNet is a semantic lexicon for the English language. It groups English words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets. It has been used in automatic text analysis and artificial intelligence applications. Instead of giving specific values of P, A, and D, the first step is to map the 22 emotions to the 8 varieties of PAD emotions. If needs be, we can figure out the difference in the value later.

Following the principle of ontology matching, vocabularies generally are separated into lists of classes, predicates and instances, and then compared class vs. class, predicate vs. predicate, etc. However, sometimes it is desirable to compare whole vocabularies without such classification since some authors may represent similar concepts with different types of terms.

Here we have two lists of terms from 22 emotions in OCC model and sample emotions in PAD, and we want to produces a list of matched pairs. Each pair contains two terms: one from the source, OCC emotion set, and the other from the target, sample emotions in PAD, or PAD emotion set. Each term can be multi-word, such as “happy-for”, etc. The matched pairs are then found through combining following four methods [92]:

1. Whole term matching: This is the first as well as the simplest procedure to be executed. The terms in both ontologies are converted to lowercase and then compared for an exact name string match. For example, we have “fear” in 22 OCC emotions, and we also have “fear” in PAS sample emotion terms, thus we directly map “fear” to (–P+A–D).

2. Word constituent matching: This is the second procedure to be executed. Each term is broken into words wherever there is a capital letter, a hyphen or an underscore. Words such as “a”, “the”, “of”, “in”, “for”, etc. are dropped from multi-word terms. For example, “happy-for” is changed to “happy”. Remaining words for each term are morphologically processed and compared in exact string match to words of each term from the target ontology. Using this procedure, un-obvious matching term pairs such as “anger” and “angry”, “satisfaction” and “satisfied” can be found.

3. Synset matching: This is the third procedure to be executed. It explores the semantic meanings of the word constituents by using the WordNet [114] synsets to help identify synonyms in matching. A synset is a WordNet term for a sense or a meaning by a group of synonyms. This procedure is similar to the method in word constituent matching in decomposing multi-word terms into their word

constituents except that it does not perform direct matching between the words. For each word in each term in each ontology, if it is in WordNet, then it must belong to one of the synsets and have at least one WordNet synset index number. The procedure associates the WordNet synset index numbers of the constituent words with the term. The synset index numbers are close for synonyms. Thus the two terms which have the largest number of common synsets are recorded and presented. The synset index numbers for the 22 OCC emotions are shown in Table 6.3.

4. Type matching: This is the last procedure, and it explores the ontological category of each word constituent for matching. It is base on the hyponyms and hypernyms from WordNet synsets. For example, if A is a feeling of joy, and B is also a feeling of joy, however we did not find matching of A and B by using previous method, then we can use this one to match A and B. Thus, using this procedure, terms that cannot be matched by previous methods, either string comparison or sense comparison, will be matched if they represent classes or properties of the same type.

Then we can get a mapping from 22 OCC emotions to PAD sample emotions. The next step is to represent these OCC emotions with PAD. From PAD sample emotions to one of the eight PAD variations is easy. However considering that "Dominance" alludes to some relationship between the subject and the object, we will do some adjustment according to the situation in OCC model. For example, if the kind of emotion is a response of an agent to itself, the "Dominance" value should be set to neutral or 0 since there is nothing about others. The results are shown in following Section 6.3.

6.3 Mapping Process and Results

By following above process, we have following mapping by applying whole term matching and word constituent matching: fear ⇒ fear (–P+A–D); satisfaction ⇒ satisfied (+P–A+D); distress ⇒ distressed (–P+A–D); admiration ⇒ admired (+P+A +D); anger ⇒ angry (–P+A+D); love ⇒ loved (+P+A–D).

To apply synset matching, we first get the synset index numbers from WordNet for the 22 OCC emotions as in the following Table 6.3.

By applying synset matching and type matching we get following mappings as in Figure 6.1.

Considering the relationship between the original OCC emotions and the mapped emotions, and the detailed situation for each OCC emotions, we get the following PAD representation, as in following Table 6.4. If an emotion may be positive or negative value on some dimension, we set it be "N" – neutral. For example, "happy" has one sense that is mapped to "elated" (+P+A+D), and another is mapped to "satisfied" (+P–A+D), then we set "A" value to be "N" since it can be positive or negative. The "D" value is also adjusted according to the relationship between the subject and the object in the OCC model.

6.4 Summary

This chapter gives a mapping from 22 OCC emotions to PAD sample emotions space and then classified to 8 groups. The mapping has following features:

- It is consistent with the senses in WordNet. During our mapping process, the WordNet is interpreted and used as a lexical ontology. The hypernym/hyponym relationships among the synsets is interpreted as specialization relations between conceptual categories.

Table 6.3: Synset Index Numbers for OCC Emotions

OCC Emotions	Word Class	Synset Index Numbers
happy	adjective	01194588, 01088951, 02649875, 01040585
resentment	noun	07446390
gloating	noun	07429832
	verb	00874329, 02147144
pity	noun	07451394, 07205004, 04774185
	verb	01804852
hope	noun	07409610, 07438591, 05875007, 10032734, 10896562, 04792935
	verb	01809579, 01794298, 00697932
fear	noun	07417148, 07422123, 07418802
	verb	01763725, 01763198, 01763564, 01763430, 01761564
satisfaction	noun	07428849, 13800811, 13121678, 01058249
relief	noun	07391167, 14253998, 13119882, 10488747, 01194210, 15073413, 07256304, 01073784, 00351073, 04027856, 01061233
disappointment	noun	07438140, 00066741
joy	noun	07424946, 05756981
	verb	01796740, 01796355
distress	noun	07394350, 14284168, 14134837, 00083473
	verb	01780950
pride	noun	07406373, 07429110, 04831402, 07886934, 00746908
	verb	01755494
shame	noun	07404456, 14248441, 07205004
	verb	02522971, 02483927, 01775138, 01097127
admiration	noun	07398628, 07407864, 01203324
reproach	noun	06623791, 14250018
	verb	00817162
gratification	noun	13800811, 01058464
remorse	noun	07433612
gratitude	noun	07402230
anger	noun	07414249, 13850329, 00747687
	verb	01768967, 01769975
love	noun	07440729, 05741438, 09704247, 07386227, 13421623, 00834429
	verb	01758160, 01811592, 01758531, 01414190
hate	noun	07443888
	verb	01757132

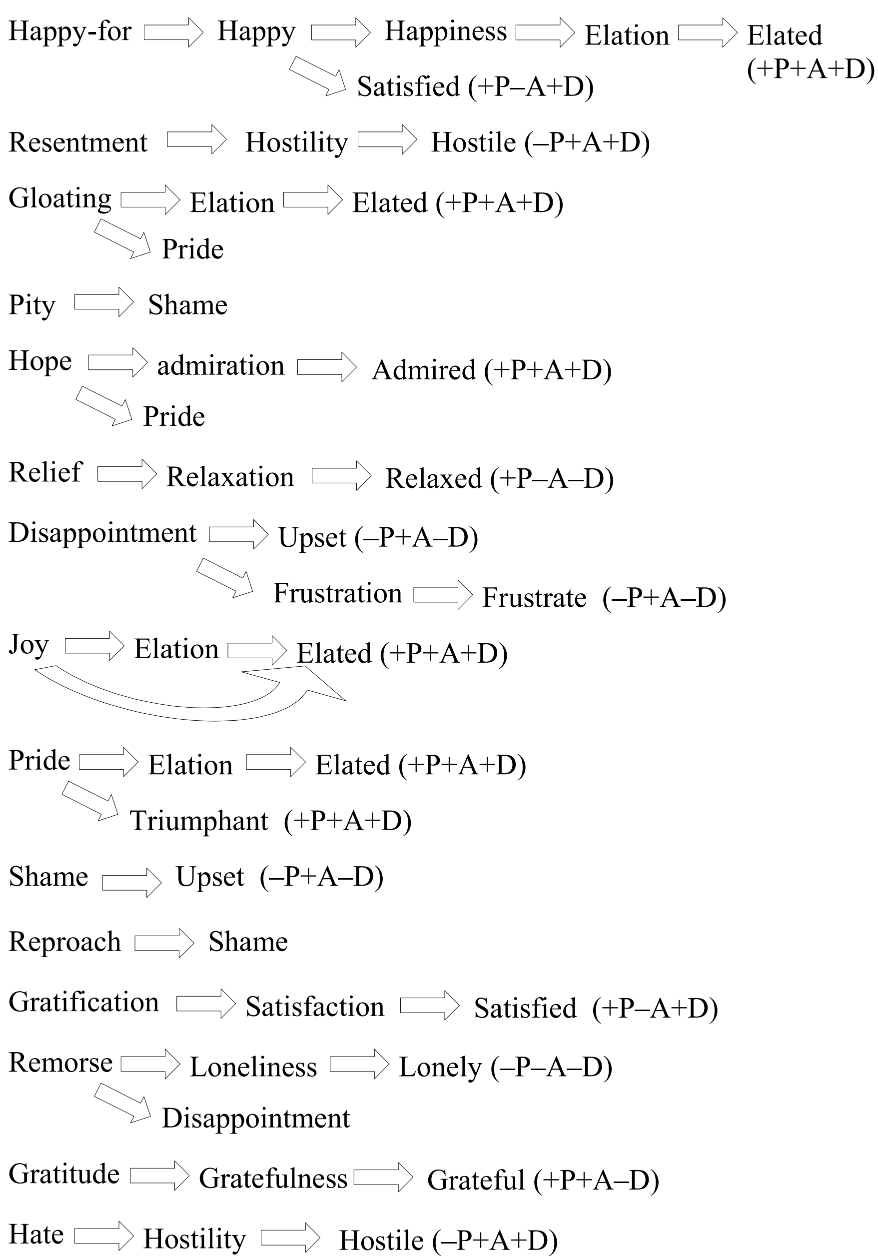


Figure 6.1: Mapping by Synset Matching and Type Matching

Table 6.4: OCC to PAD Mapping Result

OCC Emotions	Pleasure	Arousal	Dominance
happy-for	+	N	+
resentment	−	+	+
gloating	+	+	+
pity	−	+	−
hope	+	+	+
fear	−	+	−
satisfaction	+	−	+
fear-confirmed	−	−	−
relief	+	−	+
disappointment	−	+	−
joy	+	+	N
distress	−	+	N
pride	+	+	N
shame	−	+	N
admiration	+	+	+
reproach	−	+	−
gratification	+	−	+
remorse	−	−	−
gratitude	+	+	−
anger	−	+	+
love	+	+	−
hate	−	+	+

- It is consistent with the situation in the OCC model. Since WordNet is a universal database that has a lot of basic semantic, it does not consider the specific situation in the OCC model. Thus, some adjustment is applied, such that the results fit better.

For better mapping, we would like to apply some statistical method in the future as well: for example, ask 100 or more people to do this mapping subjectively. We can then compare the statistical result.

So far, we have the plug-in based on OCC model to generate and update emotions. For the primary emotions, we just record them as kind of the positive or negative trends, which could simply be applied to speed up the decision process. And for the secondary emotions, the 22 emotions generated from OCC model, we mapped them to the PAD emotion scale. Then the next step is figure out the decision process – how would these 3 dimensions affect the negotiation result. Next chapter will focus on solving this problem.

Chapter 7

Emotional Negotiation Strategy and Evaluation

As introduced above, PAD uses three basic dimensions of emotion: Pleasure – Displeasure (P), Arousal – Non-arousal (A) or general level of physical activity and mental alertness, and Dominance–Submissiveness (D) or feelings of control vs. lack of control over one's activities and surroundings. It is a numerical measurement, however it uses three dimensions: Pleasure, Arousal, and Dominance. So, we need to relate these measurements into automated negotiation, and the problem is how to combine these dimensions to correctly reflect the effects of emotions in negotiation.

Thus, this chapter shows how to combine the three dimensions in PAD emotion scale and use them to implement emotionally enhanced automated negotiating agents. Emotions' effects on negotiation strategy are modeled as in Section 7.1 by mapping a rational negotiation strategy to an emotional one, which is called PAD Emotional Negotiation Model. Section 7.2 gives evaluation about these effects, and the result shows that it reflects human experience and is consistent with negotiation theory.

7.1 PAD Emotional Negotiation Model

To relate above chosen measurements into automated negotiation, first we need to find out the relationships of the three dimensions with human behavior, and map them to the negotiation. We analyze the details of PAD, and the relationships with human behavior and negotiation as follows.

- P: Pleasure–Displeasure. This gives the direction of emotions, positive emotion status / negative emotion status. Generally, for humans, a positive emotional state is more conducive to a person acting in a friendly and sociable manner with others; conversely, a negative emotional state tends to heighten chances that the individual will be unfriendly, inconsiderate, or even rude to others. During negotiation, a more pleasant agent tends to cooperate with others or tends to accept others' offers; on the contrary, a more unpleasant agent tends to reject others' offers. We can reflect this relationship to the value system by assuming that pleasure makes the agent increase the evaluation value and displeasure makes the agent decrease the value.

- A: Arousal–Non-arousal. This gives the degree of effects on the above intentions as given by P. Arousal means to rouse or stimulate to action or to physiological readiness for activity. We can reflect this to the value system of negotiation by assuming that this measure magnifies or minimizes P's affection. For example, if an agent is in pleasure status this emotion makes the agent increase the evaluation value a little; if the agent is also on arousal, it increases even more. But, if the agent is in displeasure, then arousal will make the agent decrease the value more.

- D: Dominance–Submissiveness. This estimates the degree of the ability of being commanding, controlling, or prevailing over all others, or degree to yield oneself to the authority or will of another. This description is close to the idea of power in Network Exchange Theory (NET) [181]. The agent in a dominant state or with more power tends to persist in its own proposal and benefit more in negotiation. However, the D value in PAD is decided by emotional status, which is subjective; the power in NET is objectively decided by the network structure. We can relate this measurement to the value system of negotiation by assuming that since a dominant agent tends to persist in its own proposal it will tend to decrease the evaluation value. On the other hand, if the agent is submissive, it will tend to yield and accept the other agent's proposal.

By analyzing and combining all the above relationships together, we define the following emotional state function:

$$S_e(r_p, r_a, r_d) = r_p \cdot (1 + r_a) - r_d$$

where $r_p, r_a, r_d \in (-1, 1)$ are a measurement of the three dimensions of the PAD model. They define an emotional status. For example, anger is defined as $\{-0.51, 0.59, 0.25\}$, which means $r_p = -0.51$, $r_a = 0.59$, $r_d = 0.25$ [110], fear is $\{-0.64, 0.60, -0.43\}$.

The emotional worth W_e is then defined as

$$W_e(S_e, e, i) = S_e(r_p, r_a, r_d) \cdot W(e, i).$$

We can also define the effects of emotion on the rational evaluation to be given by F where

$$F = r_e \cdot S_e(r_p, r_a, r_d).$$

F tells us how much the rational evaluation will increase or decrease due to the emotional state. For example, if $F = 0.1$ and the rational worth function is given by W then the emotional state makes the worth increase by $0.1 \cdot W$.

The agent's decision is thus based on

$$j^i_{opt} = \arg\max_{j:e_0 \to e \in J} r_e \cdot (r_p \cdot (1 + r_a) - r_d) \cdot W(e,i) + W(e,i) - c(j,i)$$

So far, we have presented a detailed equation for the optimal plan. The negotiation protocol we assume has the agent offering a new proposal or accepting the other's proposal at each time step. If the proposal is accepted then negotiation ends. The difference between the offers at successive time steps at time τ is called d_τ. Different agents may use different strategies for proposing their next offer. As such, d_τ could be a constant, as in the monotonic concession protocol [142], or it could change. We let $S(\tau)$ be the agent's strategy function. For example, an agent using the monotonic concession protocol would have $S(\tau) = d$.

Given a rational agent with a strategy function $S(\tau)$, we can convert it to an emotional one by mapping the original strategy function to a new one $S'(\tau)$, as such:

$$\begin{aligned} S'(\tau) &= M(S(\tau)) \\ &= r_e \cdot S_e(r_p, r_a, r_d) \cdot S(\tau) + S(\tau) \end{aligned}$$

where $M(\cdot)$ is our emotional mapping function that maps rational strategies to emotional strategies.

7.2 Evaluation

We consider a situation where two agents need to reach an agreement on a given issue. We assume the agents use a protocol based on the the strategic-negotiation model [80, 79], which is derived from Rubinstein's model of alternating offers [144].

We consider cases with only two agents: A and B. It is assumed that the agents can take actions in the negotiation only at certain times in the set $T = \{0, 1, 2...\}$ that are determined in advance and are known to the agents. In each period $\tau \in T$ of the negotiation, if the negotiation has not terminated earlier, the agent whose turn it is to make an offer at time τ will suggest a possible deal, and the other agent may either

1. accept the most recent offer or proposal,

2. reject it,

3. opt out of the negotiation.

If an offer is accepted by both agents, then the negotiation ends, and the offer is implemented. If at least one of the agents opts out of the negotiation, then the negotiation ends and a conflict outcome results. If no agent has opted out and at least one agent rejects the offer, then the negotiation proceeds to period $\tau + 1$ where a new offer is made.

In theory, both agents can keep rejecting offers so that an agreement may never be reached (in that case we talk about disagreement or a conflict deal). However, an agent's utility depends on the value at which an agreement is reached as well as on the time at which it is reached, hence disagreement is the worst possible outcome for both players. Our model makes the following further assumptions:

- *Agreement is preferred*: agents prefer any deal at least as much as disagreement.

- *Agents seek to maximize utility*: agents prefer to get larger utility values.

- *Agents have a reservation price*: if the utility is below the reservation price, an agent would rather not reach agreement.

Now, let's consider the following specific scenario: two agents A and B want to split \$10. To see the property of the r_p, r_a, r_d clearly, we let the strategy function

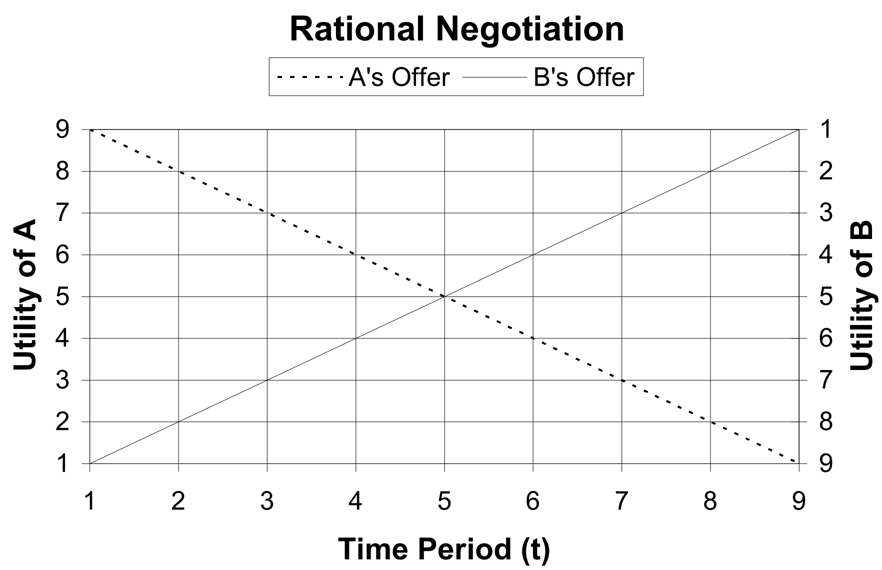


Figure 7.1: Rational Negotiation Process ($r_e = 0$)

of both agents be the simplest one, a constant ϵ, as in the monotonic concession protocol:

$$S(\tau) = \epsilon$$

Assuming $\epsilon = \$1$ for each time round, then rational agents with $r_e = 0$ will have their rational negotiation process described as in Figure 7.1. A's strategy is described as the dotted line. A starts by giving its best offer: "A gets \$9 and B gets \$1". At each time step of the negotiation A will propose a new offer along this dotted linear line until it reaches its reservation price or the negotiation ends. B's strategy is similar and is described by the solid line. All the possible deals are represented as points on the y-axis. For example, A gets \$9 and B gets \$1, or A gets \$8 and B gets \$2, etc. The x-axis shows the time rounds. The cross point shows when and what deal agents A and B will agree on. In this case, they end with A getting \$5 and B getting \$5.

We now show how to add emotions to these agents. We let the agents be completely emotional, that is, $r_e = 1$, and vary their emotional dimensions r_p, r_a, r_d

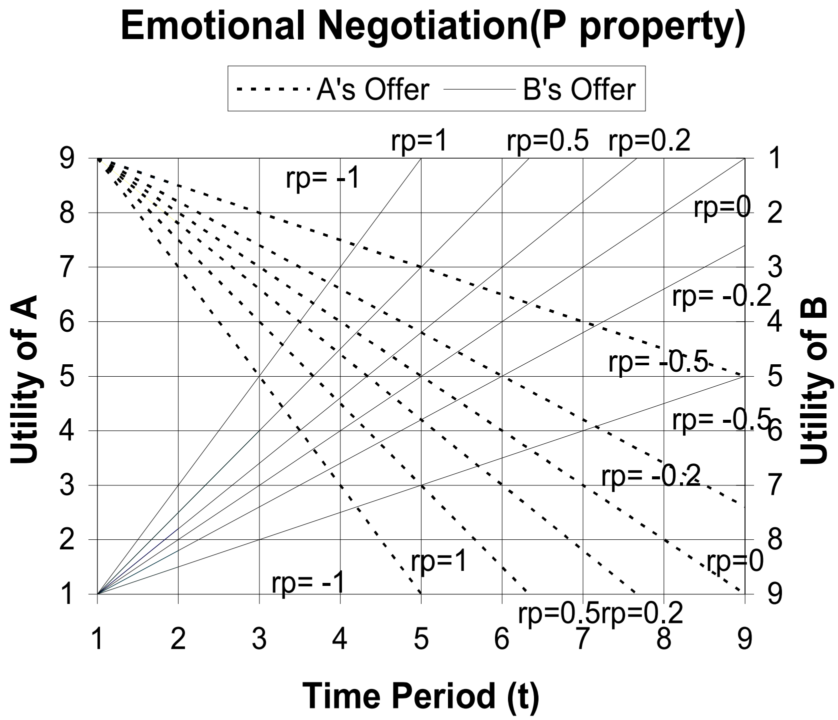
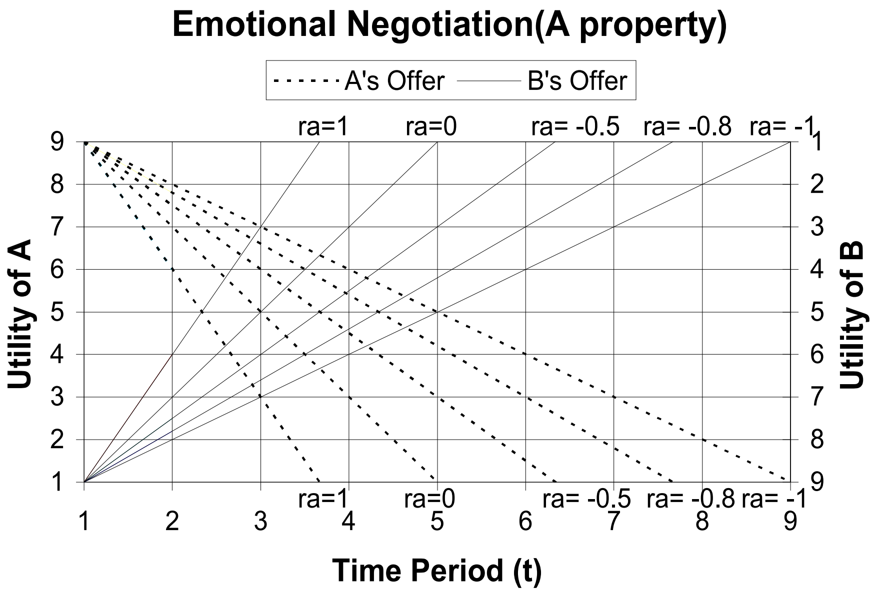


Figure 7.2: Emotional Negotiation Process ($r_e = 1, r_a = 0, r_d = 0$)

separately. Figure 7.2 shows an example where we set $r_e = 1, r_a = 0, r_d = 0$, which lets emotion features A and D be neutral and lets r_p vary from -1 to 1. From this figure, we can see that:

- A more pleasant agent ends up with a deal more quickly, and a more unpleasant agent ends up with a deal more slowly;
- Agents of the same type with the same pleasure status end up with the same benefit. That is, they will reach the deal where A gets \$5 and B gets \$5;
- If two agents of the same type but with different pleasure status engage in negotiation, then they will reach an agreement that is more favorable to the more unpleasant agent.

We can see that the above results fit our intuition as well as our theory. A pleasant person easily accepts any offer, which means he might not benefit as much. However,



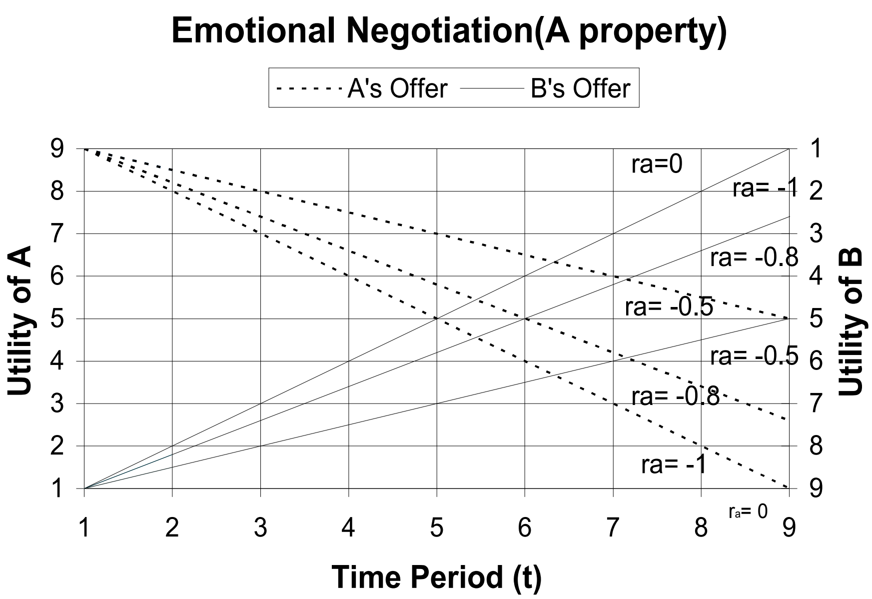


Figure 7.3: Emotional Negotiation Process (Top: $r_e = 1, r_p = 1, r_d = 0$; Bottom: $r_e = 1, r_p = -1, r_d = 0$)

if there are many different negotiations, then we can imagine that a pleasant person will end up with a lot of different deals. Our model is similar. Since a more pleasant agent ends up with a deal more quickly, it has time for other possible trades. Thus, its short-term loss might translate into a long-term gain.

Let's now consider emotion feature A's property by setting $r_e = 1, r_p = 1, r_d = 0$ and letting r_a vary from -1 to 1; then by setting $r_e = 1, r_p = -1, r_d = 0$ and letting r_a vary from -1 to 0. The negotiation process for these two cases is described in Figure 7.3. Notice that in the bottom figure we do not show negotiation lines for $r_a > 0$, because the value will decrease to negative for this extreme case $r_p = -1$, and the negotiation will end immediately, since it is below the reservation price. In other words, the effects of emotion F can't be less then -1, which makes the negotiation stop. By analyzing these cases, we can find the following properties for emotion feature A:

- A more aroused agent with pleasure status will end up with a deal even more quickly, but a more aroused agent with displeasure status will end up with a deal more slowly; a more non-aroused agent with pleasure status will end up with a deal more slowly, but a more non-aroused agent with displeasure status will end up with a deal more quickly. In other words, arousal magnifies the effect of the agent's pleasure/displeasure status, and non-arousal minimizes the effect of the agent's pleasure/displeasure status.

- Two agents of the same type with the same pleasure and arousal status end up with the same benefit. That is, they reach a deal where A gets \$5 and B gets \$5;

- If two agents of the same type with the same pleasure (displeasure) status but different arousal status engage in negotiation, the result is that the one that is more aroused will benefit less (more).

Finally, let's consider emotion feature D's property by setting $r_e = 1, r_p = 0$, and letting r_d vary from -1 to 1. The negotiation process is shown in figure 7.4. As before, we can find the following properties for emotion feature D after analyzing the figure.

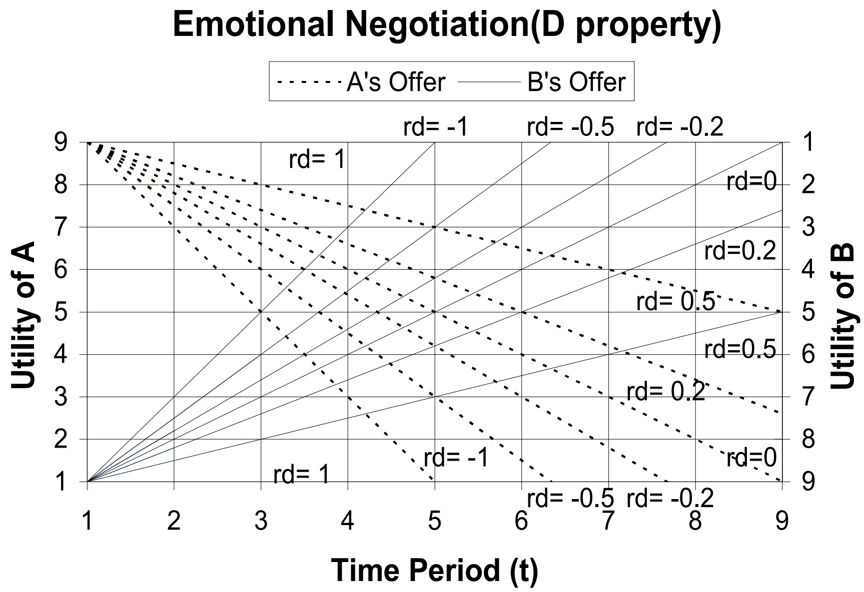


Figure 7.4: Emotional Negotiation Process ($r_e = 1, r_p = 0$)

- A more dominant agent takes longer to reach an agreement, and a more submissive agent reaches an agreement faster;
- Two agents of the same type, with the same dominance / submissiveness status, end up with equal benefit. That is, they reach a deal where A gets \$5 and B gets \$5;
- If two agents of the same type, but with different dominant / submissive status, engage in negotiation, then the dominant agent will benefit more and the submissive agent will benefit less.

Above all, the properties of the emotion features P, A, D in the model reflect human experience, and agents of the same type with the same emotional status end up with a deal with equal benefit for both as in negotiation theory, which is summarized in Table 7.1.

We note that there are some limitations to our analysis. Namely, agents with complex strategies would not be represented by a linear strategy function, but might

Table 7.1: Summary of P-A-D Properties

	P-property	A-property	D-property
Range	$r_p \in$[-1, 1]	$r_a \in$[-1, 1]	$r_d \in$[-1, 1]
Positive	Increase evaluation	Increase effects of P-property	Decrease evaluation
Neutral	None effects	None effects	None effects
Negative	Decrease evaluation	Decrease effects of P-property	Increase evaluation
Effect Range (F)	[-r_e, r_e]	[-1, 2$\cdot r_e$]	[-r_e, r_e]
Negotiation Theory	Match	Match	Match

require complex curves. Also, an agent's emotional state might change during the course of negotiation, which would have the effect of changing the strategy function. However, our analysis should still hold if we consider small enough time intervals. Any curve can be approximated by a line for small enough lengths and a changing function can be approximated by a fixed function for small enough time steps.

7.3 Summary

This chapter proposed an automated negotiation model that incorporates emotions into the agents' strategies. We evaluated the model and showed that it reflects human experience and negotiation theory. Specifically, the P-dimension shows that a pleasant agent ends up with a deal faster but benefits less in a single trade; the A dimension magnifies or minimizes the trends of the P dimension; the D-dimension shows that a dominant agent insists on its own benefit, and it benefits more from the deal but reaches a deal slower. No matter what, agents of the same type and same emotional status will end up with a deal of equal benefit.

In the work above, we show how the emotional status of agents affects their negotiations, which is an important but very basic step. Since there are no previous numerical human experience data we could compare our model with, we simply gave the reasonable range for the model and verified that it does reflect human experience. The popular practical value ranges for each dimension are still to be collected.

Chapter 8

Conclusion and Future Work

This dissertation focuses on the effects of emotions on decision-making. By introducing primary and secondary emotion into BDI architecture, we first present a generic architecture for an emotional agent, EBDI, which can merge various emotion theories with an agent's reasoning process. It implements practical reasoning techniques separately from the specific emotion mechanism. The separation allows us to plug in emotional models as needed or upgrade the agent's reasoning engine independently. Sample agents in Tileworld show that an EBDI agent can have better performance than traditional BDI agents.

Next, we try to apply this architecture in automated negotiation. Then a plug-in for EBDI architecture is designed, which modifies the OCC model, a standard model for emotion synthesis, to generate emotions. And I analyze the possibility to incorporate emotions into negotiation and generate EWOD (Emotional Worth-Oriented Domain), which requires emotions to be numerical. Thus, a mapping from 22 OCC emotions to 3-dimension numerical PAD emotions is given. Finally, how these 3-dimension emotions affect the negotiation strategy is described and an evaluation is given which shows that it can be used to implement agents with various emotional states that mimic human emotions during negotiation.

Thus we can potentially design an agent for negotiation with a high EQ agent according to specific applications and purposes. For example, we can design different emotional agents in negotiation by adjusting the P-A-D value. According to different purpose, we can apply desired P-A-D value. However, to collect these data need a lot of testing and the design should be based on the test result. This is an area for future research.

Since negotiation is already used widely to solve many problems in different domains, and this research is based on a general process of negotiation, the research results can also be widely used in other areas.

Beyond the area of automated negotiation, the EBDI architecture can also be applied to many other areas, such as affective control. These areas are to be extended and applied. Also, the designed plug-in can be used for other application as well.

Bibliography

[1] Li yun (ceremonial usages: Their origin, development, and intention). *The Book of Rites*, 200 B.C.

[2] James Adam. *The republic of Plato, with critical notes, commentary and appendices*. Cambridge University press, 1902.

[3] Robert S. Adler, Benson Rosen, and Elliot M. Silverstein. Emotions in negotiation: How to manage fear and anger. *Negotiation Journal*, 14(2):161–179, April 1998.

[4] Aristotle. Nicomachean ethics. `http://classics.mit.edu/Aristotle/nicomachaen.html`, 350 B.C.

[5] Aristotle. Rhetoric. `http://classics.mit.edu/Aristotle/rhetoric.html`, 350 B.C.

[6] Ronald C. Arkin, Masahiro Fujita, Tsuyoshi Takagi, and Rika Hasegawa. An ethological and emotional basis for human-robot interaction. *Robotics and Autonomous Systems*, 42(3-4):191–201, 2003.

[7] Augustine. On the two souls, against the manicheans. *Nicene and Post-Nicene Fathers, Series One*, 4, 1992.

[8] John L. Austin. *How to do Things with Words*. Oxford University Press, Oxford, UK, 1962.

[9] Rina Azoulay-Schwartz and Sarit Kraus. Negotiation on data allocation in multi-agent environments. *Autonomous Agents and Multi-Agent Systems*, 5(2):123–172, 2002.

[10] Th. Ballmer and W. Brennenstuhl. *Speech Act Classification – A Study in the lexical Analysis of English Speech Activity Verbs*. Springer-Verlag, Berlin, Heidelberg, New York, 1981.

[11] Christoph Bartneck. Integrating the occ model of emotions in embodied characters. In *Proceedings of the Workshop on Virtual Conversational Characters: Applications, Methods, and Research Challenges, Melbourne.*, 2002.

[12] Joseph Bates. The role of emotion in believable agents. *Commun. ACM*, 37(7):122–125, 1994.

[13] Antoine Bechara. The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, 55:30–40, 2004.

[14] M. Ben-Or and N. Linial. Collective coin flipping. *Randomness and Computation*, pages 91–115, 1990.

[15] John Bowlby. *Attachment and Loss*. Basic Books, New York, 1969.

[16] Michael E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.

[17] Cynthia Breazeal. *Designing Sociable Robots*. MIT Press, Cambridge, MA, USA, 2002.

[18] Colin F. Camerer, George Loewenstein, and Matthew Rabin, editors. *Advances in Behavioral Economics*. Princeton University Press, Princeton, NJ, 2003.

[19] Dolores Canamero. Modeling motivations and emotions as a basis for intelligent behavior. In *AGENTS '97: Proceedings of the first international conference on Autonomous agents*, pages 148–155, New York, NY, USA, 1997. ACM Press.

[20] Walter Cannon. *Bodily Changes in Pain, Hunger, Fear and Rage 2nd Ed.* Appleton, New York, 1929.

[21] Man Kit Chang and Carson C. Woo. A speech-act-based negotiation protocol: design, implementation, and test use. *ACM Trans. Inf. Syst.*, 12(4):360–382, 1994.

[22] Louis Charland. Feeling and representing: Computational theory and the modularity of affect. *Synthese*, 105:273–301, 1995.

[23] Cicero. *De Finibus Bonorum et Malorum.* Heinemann, London, 1931.

[24] Cristina Conati. Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, 16(7-8):555–575, 2002.

[25] S.E. Conry, K. Kuwabara, V.R. Lesser, and R.A. Meyer. Multistage negotiation for distributed satisfaction. *IEEE Transactions on Systems, Man, and Cybernetics, Special Issue on Distributed Artificial Intelligence*, 21(6):1462 – 1477, December, 1991.

[26] S.E. Conry, R.A. Meyer, and V.R Lesser. Multistage Negotiation in Distributed Planning. *Readings in Distributed Artificial Intelligence*, pages 367–384, 1988.

[27] Antonio R. Damasio. *Descartes' Error : Emotion, Reason, and the Human Brain.* Avon Books, New York, 1994.

[28] Antonio R. Damasio. *The Feeling of what Happens: Body and Emotion in the Making of Consciousness.* Harcourt Brace and Co., New York, 1999.

[29] Richard J. Davidson, Klaus R. Scherer, and H. Hill Goldsmith, editors. *Handbook of Affective Sciences.* Oxford University Press, New York, 2002.

[30] Randall Davis and Reid G. Smith. Negotiation as a metaphor for distributed problem solving. *Artificial Intelligence*, 20:63–109, 1983.

[31] Ronald de Sousa. *The Rationality of Emotion.* MIT Press, Cambridge, MA, 1987.

[32] Ronald de Sousa. Emotion. Stanford Encyclopedia of Philosophy. `http://plato.stanford.edu/entries/emotion/`, 2003.

[33] Benedictus de Spinoza. *Ethics in The Collected Works of Spinoza vol. 1.* Princeton University Press, Princeton, N.J., 1985.

[34] Benedictus de Spinoza. *Ethics.* Oxford University Press, New York, 2000.

[35] K. Decker and V. Lesser. An Approach to Analyzing the Need for Meta-Level Communication. *International Joint Conference on Artificial Intelligence*, 1, January 1993.

[36] John Deigh. Cognitivism in the theory of emotions. *Ethics*, 104:824–854, 1994.

[37] René Descartes. *Discourse on Method.* Liberal Arts Press, New York, 1956.

[38] D. Druckman. *Negotiations.* Sage Publications, London, 1977.

[39] Edmund Durfee and Victor Lesser. Using Partial Global Plans to Coordinate Distributed Problem Solvers. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 875–883, August 1987.

[40] Edmund H. Durfee. *Coordination of Distributed Problem Solvers.* Kluwer Academic Publishers, Norwell, MA, USA, 1988.

[41] Edmund H. Durfee and Victor R. Lesser. *Negotiating task decomposition and allocation using partial global planning*, pages 229–243. Distributed Artificial Intelligence (Vol. 2). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1989.

[42] M. G. Dyer. Emotions and their computations: Three computer models. *Cognition and Emotion*, 1(3):323–347, 1987.

[43] Paul Ekman and Richard J. Davidson, editors. *The Nature of Emotion: Fundamental Questions*. Oxford University Press, New York, 1994.

[44] Clark Elliott. The affective reasoner: A process model of emotions in a multi-agent system. Technical report, Institute for the Learning Sciences, Northwestern University, 1992.

[45] Jon Elster. *Alchemies of the Mind: Rationality and the Emotions*. Cambridge University Press, Cambridge, 1999.

[46] E. Ephrati and J.S. Rosenschein. The clarke tax as a consensus mechanism among automated agents. In *Proceedings of the National Conference on Artificial Intelligence*, pages 173–178, San Jose, California, July 1991.

[47] Eithan Ephrati and Jeffrey S. Rosenschein. Reaching agreement through partial revelation of preferences. In *ECAI '92: Proceedings of the 10th European conference on Artificial intelligence*, pages 229–233, New York, NY, USA, 1992. John Wiley & Sons, Inc.

[48] Eithan Ephrati and Jeffrey S. Rosenschein. Distributed consensus mechanisms for self-interested heterogeneous agents. In *CoopIS*, pages 71–79, 1993.

[49] Dylan Evans. The search hypothesis of emotions. *British Journal for the Philosophy of Science*, 53:497–509, 2002.

[50] FIPA. *FIPA Communicative Act Library Specification*. Foundation for Intelligent Physical Agent, Geneva, Switzerland, 2002.

[51] R. Fisher and W. Ury. *Getting To Yes: Negotiating Agreement Without Giving In*. Houghton Mifflin, Boston, 1981.

[52] Mark S. Fox. *An organizational view of distributed systems*, pages 140–150. Distributed Artificial Intelligence. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[53] Nico Frijda and Jaap Swagerman. Can computers feel? theory and design of an emotional system. *Cognition and Emotion*, 1(3):235–257, 1987.

[54] Nico H. Frijda, Antony S. R. Manstead, and Sacha Bem, editors. *Emotions and Beliefs: How Feelings Influence Thoughts*. Cambridge University Press, New York, 2000.

[55] Galen. *The Best Doctor is Also a Philosopher*. Galen: Selected Works. Oxford University Press, New York, 1997.

[56] Les Gasser. Social conceptions of knowledge and action: Dai foundations and open systems semantics. *Artif. Intell.*, 47(1-3):107–138, 1991.

[57] Les Gasser. Social knowledge and social action: Heterogeneity in practice. In *IJCAI'93: Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pages 751–757, 1993.

[58] Michael P. Georgeff and François Felix Ingrand. Decision-making in an embedded reasoning system. In *IJCAI'89: Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 972–978, 1989.

[59] Carlos Gershenson. Modelling emotions with multidimensional logic. In *NAFIPS '99: Proceedings of 18th International Conference of the North American Fuzzy Information Processing Society*, pages 42–46, 1999.

[60] Edmund L. Gettier. Is justified true belief knowledge? *Analysis*, 23:121–123, 1963.

[61] Peter Goldie. *The Emotions: A Philosophical Exploratio*. Oxford University Press, Oxford, 2000.

[62] Ted Gracyk. Hume's aesthetics. Stanford Encyclopedia of Philosophy, `http://plato.stanford.edu/entries/hume-aesthetics/`, 2003.

[63] Jonathan Gratch. émile: Marshalling passions in training and education. In *AGENTS '00: Proceedings of the fourth international conference on Autonomous agents*, pages 325–332, New York, NY, USA, 2000. ACM Press.

[64] Lavinia Hall, editor. *Negotiation: Strategies for Mutual Gain*. Sage, Beverly Hills, 1993.

[65] John C. Harsanyi. Approaches to the bargaining problem before and after the theory of games: A critical discussion of zeuthen's, hicks', and nash's theories. *Econometrica*, 24(2), 1956.

[66] Barbara Hayes-Roth. Agents on stage: Advancing the state of the art of ai. In *IJCAI'95: Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 967–971, 1995.

[67] Eva Hudlicka and John Billingsley. Affect-adaptive user interface. In *Proceedings of HCI International (the 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Ergonomics and User*

Interfaces-Volume I, pages 681–685, Mahwah, NJ, USA, 1999. Lawrence Erlbaum Associates, Inc.

[68] Ullrich Hustadt, C. Dixon, R. A. Schmidt, M. Fisher, J.-J. C. Meyer, and W. van der Hoek. Verification within the KARO agent theory. In J. L. Rash, C. A. Rouff, W. Truszkowski, D. Gordon, and M. G. Hinchey, editors, *Proceedings of the First International Workshop on Formal Approaches to Agent-Based Systems (FAABS 2000)*, volume 1871 of *LNAI*, pages 33–47. Springer, 2001.

[69] William James. What is an emotion? *Mind*, 19:188–204, 1884.

[70] Nicholas R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, C. Sierra, and M. Wooldridge. Automated haggling: Building artificial negotiators. In *Pacific Rim International Conference on Artificial Intelligence*, page 1, 2000.

[71] Hong Jiang and Michael N. Huhns. An approach to broaden the semantic coverage of acl speech acts. In Jacky Akoka, Stephen W. Liddle, Il-Yeol Song, Michela Bertolotto, Isabelle Comyn-Wattiau, Samira Si-Said Cherfi, Willem-Jan van den Heuvel, Bernhard Thalheim, Manuel Kolp, Paolo Bresciani, Juan Trujillo, Christian Kop, and Heinrich C. Mayr, editors, *Perspectives in Conceptual Modeling, ER (Workshops)*, volume 3770 of *Lecture Notes in Computer Science*, pages 162–171. Springer, 2005.

[72] Hong Jiang and Michael N. Huhns. Broadening the semantic coverage of agent communicative acts. In *Postproceeding of Agent-Oriented Information Systems III 7th International Bi-Conference AOIS 05 Workshop*, 2006.

[73] R. Johnson. *Negotiation basics.* Sage, Newbury Park, 1993.

[74] C.L Karrass. *The Negotiating Game: How to Get What You Want.* Thomas Crowell Company, New York, 1970.

[75] Peter King. Aquinas on the passions. *Aquinas's Moral Theory*, pages 101–132, 1999.

[76] David Kinny. Reliable agent communication - a pragmatic perspective. In *PRIMA '99: Proceedings of the Second Pacific Rim International Workshop on Multi-Agents*, pages 16–31, London, UK, 1999. Springer-Verlag.

[77] P. Kleinginna and A. Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5:345–379, 1981.

[78] Sarit Kraus. Agents contracting tasks in non-collaborative environments. In *AAAI*, pages 243–248, 1993.

[79] Sarit Kraus. *Automated negotiation and decision making in multiagent environments*, pages 150–172. Mutli-agents systems and applications. Springer-Verlag New York, Inc., New York, NY, USA, 2001.

[80] Sarit Kraus. *Strategic Negotiation in Multiagent Environments*. MIT Press, 2001.

[81] Sarit Kraus and D. Lehmann. Designing and building a negotiating automated agent. *Computational Intelligence*, 104(1-2):132–171, 1995.

[82] Sarit Kraus, K. Sycara, and A. Evenchik. Reaching agreements through argumentation: a logical model and implementation. *Artificial Intelligence*, 11(1):1–69, 1998.

[83] Sarit Kraus and J. Wilkenfeld. Negotiation over time in a multiagent environment: preliminary report. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, pages 56–61, Sydney, Australia, 1991.

[84] Sarit Kraus and Jonathan Wilkenfeld. A strategic negotiations model with applications to an international crisis. *IEEE Transaction on Systems Man and Cybernetics*, 23(1):313—323, 1993.

[85] Sarit Kraus, Jonathan Wilkenfeld, and Gilad Zlotkin. Multiagent negotiation under time constraints. *Artif. Intell.*, 75(2):297–345, 1995.

[86] David M Kreps and Robert Wilson. Sequential equilibria. *Econometrica*, 50(4):863–94, July 1982.

[87] Brigitte Lâasri, Hassan Lâasri, and Victor R. Lesser. Negotiation and its role in cooperative distributed problem-solving. Technical report, University of Massachusetts, Amherst, MA, USA, 1990.

[88] Diogenes Laertes. Diogenes laertes: Lives and opinions of eminent philosophers. The Complete Online Library of Ancient Sources, `http://www.thedyinggod.com/chaldeanmagi/sources/diogenes.html`.

[89] Susan E. Lander and Victor R. Lesser. Customizing distributed search among agents with heterogeneous knowledge. In *Proceeding of First International Conference on Information Knowledge Management*, pages 335 – 344, Baltimore, 1992.

[90] J. E. LeDoux. Emotion, memory and the brain. *Scientific American*, pages 50–57, 1994.

[91] Joseph LeDoux. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Simon and Schuster, New York, 1998.

[92] John Li. Lom: A lexicon-based ontology mapping tool. `http://reliant.teknowledge.com/DAML/I3con.pdf`, 2007.

[93] Shu Li and Michael E. Roloff. *Strategic Emotion in Negotiation: Cognition, Emotion, and Culture*. From Communication to Presence: Cognition, Emotions and Culture towards the Ultimate Communicative Experience. Festschrift in honor of Luigi Anolli. IOS Press, Amsterdam, 2006.

[94] R. Duncan Luce and Howard Raiffa. *Games and Decisions: Introduction and Critical Survey*. Dover Publications, 1989.

[95] William Lyons. *Emotion*. Cambridge University Press, Cambridge, 1980.

[96] Niccoló Machiavelli. Prince and discourses (selections). *Selected Political Writings*, 1994.

[97] Nicolas Malebranche. *The Search after Truth*. Cambridge University Press, Cambridge, 1997.

[98] Thomas W. Malone. Organizing information processing systems: parallels between human organizations and computer systems. *Cognition, computing, and cooperation*, pages 56–83, 1990.

[99] Joel Marks. A theory of emotion. *Philosophical Studies*, 42:227–242, 1982.

[100] Stacy Marsella and Jonathan Gratch. Modeling the influence of emotion on belief for virtual training simulations. In *Proceedings of the 11th Conference on Computer-Generated Forces and Behavior Representation*, May 2002.

[101] Robert Masters. Compassionate wrath: Transpersonal approaches to anger. *The Journal of Transpersonal Psychology*, 32(1):31–51, 2000.

[102] Gerald Matthews and Adrian Wells. *Attention and Emotion: a Clinical Perspective*. Lawrence Erlbaum, Hillsdale, NJ, 1994.

[103] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. *Readings in nonmonotonic reasoning*, pages 26–45, 1987.

[104] Albert Mehrabian. Framework for a comprehensive description and measurement of emotional states. *Genetic, Social, and General Psychology Monographs*, 121(3):339–361, August 1995.

[105] Albert Mehrabian. Relationships among three general approaches to personality description. *Journal of Psychology*, 129:565–581, 1995.

[106] Albert Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14:261–292, 1996.

[107] Albert Mehrabian. Comparison of the pad and panas as models for describing emotions and for differentiating anxiety from depression. *Journal of Psychopathology and Behavioral Assessment*, 19:331–357, 1997.

[108] Albert Mehrabian. Correlations of the pad emotion scales with self-reported satisfaction in marriage and work. *Genetic, Social, and General Psychology Monographs*, 124:311–334, 1998.

[109] Albert Mehrabian. General tests of emotion or affect for evaluating consumer reactions to services and products. `http://www.kaaj.com/psych/scales/emotion.html`, 2005.

[110] Albert Mehrabian. Incorporating emotions and personality in artificial intelligence software. `http://www.kaaj.com/psych/ai.html`, 2005.

[111] Albert Mehrabian and R. De Wetter. Experimental test of an emotion-based approach to fitting brand names to products. *Journal of Applied Psychology*, 72:125–130, 1987.

[112] John-Jules Ch. Meyer. Reasoning about emotional agents. In *Proceedings of ECAI 2004*. IOS Press, 2004.

[113] Francois Michaud and Jonathan Audet. Using motives and artificial emotions for long-term activity of an autonomous robot. In *AGENTS '01: Proceedings of the fifth international conference on Autonomous agents*, pages 188–189, New York, NY, USA, 2001. ACM Press.

[114] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. Introduction to wordnet: An on-line lexical database. *International Journal of Lexicograph*, 3(4):235–244, 1990.

[115] Marvin Minsky. The emotion machine (draft). `http://web.media.mit.edu/~minsky/`.

[116] T. Moehlman, V. Lesser, and B. Buteau. Decentralized negotiation: An approach to the distributed planning problem. *Group Decision and Negotiation*, 1(2):161–192, January 1992.

[117] Luìs Morgado and Graca Gaspar. Emotion based adaptive reasoning for resource bounded agents. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 921–928, New York, NY, USA, 2005. ACM Press.

[118] John F. Nash. The bargaining problem. *Econometrica*, 18:155–162, 1950.

[119] John F. Nash. Two-person cooperative games. *Econometrica*, 21(1):128–140, April 1953.

[120] Jerome Neu. *Emotion, Thought and therapy: a Study of Hume and Spinoza and the Relationship of Philosophical Theories of the Emotions to Psychological Theories of Therapy*. Routledge and Kegan Paul, London, 1977.

[121] Martha Nussbaum. *Love's Knowledge: Essays on Philosophy and Literature*. Oxford University Press, Oxford, 1992.

[122] Justin Oakley. *Morality and the Emotions*. Routledge and Kegan Paul, London, 1992.

[123] Brian Ogren. Aristotle's rhetoric and the cognition of being: Human emotions and the rational-irrational dialectic. *Minerva - An Internet Journal of Philosophy*, 8:1–19, 2004.

[124] Jimena Olveres, Mark Billinghurst, Jesus Savage, and Alistair Holden. Intelligent, expressive avatars. In *Proceedings of the First Workshop on Embodied Conversational Characters (WECC'98)*, October 1998.

[125] Andrew Ortony, Gerald L. Clore, and Allan Collins, editors. *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, UK, 1988.

[126] Martin J. Osborne and Ariel Rubinstein. *Bargaining and Markets*. Academic Press, 1990.

[127] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT Press, 1999.

[128] Rizzo P., Veloso M.V.and Miceli M., and Cesta A. Personality-driven social behaviors in believable agents. In *AAAI 1997 Fall Symposium on "Socially Intelligent Agents", AAAI Press Technical Report FS-97-02*, pages 109–114, Menlo Park, CA, USA, 1997. AAAI Press.

[129] Lin Padgham and Guy Taylor. A system for modelling agents having emotion and personality. In *PRICAI Workshop on Intelligent Agent Systems*, pages 59–71, 1996.

[130] Jaak Panksepp. *Affective neuroscience: the foundations of human and animal emotions*. Oxford University Press, New York, Oxford, 1998.

[131] H. Van Dyke Parunak, Robert Bisson, Sven Brueckner, Robert Matthews, and John Sauter. A model of emotions for situated agents. In *AAMAS'06: Proceedings of International Joint Conference on Autonomous Agents and Multi-Agent Systems (submitted)*, 2006.

[132] David Pereira, Eugénio Oliveira, N. Moreira, and L.Sarmento. Towards an architecture for emotional bdi agents. In *EPIA '05: Proceedings of 12th Portuguese Conference on Artificial Intelligence*, pages 40–47. Springer, 2005.

[133] Emmanuel Peterfreund. *Information, Systems, and Psychoanalysis: An Evolutionary Biological Approach to Psychoanalytic Theory*. Intrernational Universities Press, New York, 1971.

[134] R. Pfeifer. Artificial intelligence models of emotion. *Cognitive Perspectives on Emotion and Motivation*, 44, 1988.

[135] R. W. Picard. *Affective Computing*. MIT Press, Cambridge, MA, 2000.

[136] Juhani Pietarinen. Hobbes, conatus and the prisoner's dilemma. Modern Philosophy. `http://www.bu.edu/wcp/Papers/Mode/ModePiet.htm`, 2000.

[137] Martha Pollack and Marc Ringuette. Introducing the tileworld: experimentally evaluating agent architectures. In Thomas Dietterich and William Swartout, editors, *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 183–189, Menlo Park, CA, 1990. AAAI Press.

[138] Marek Pyka. On emotion and value in david hume and max scheler. Philosophy of Values. `http://www.bu.edu/wcp/Papers/Valu/ValuPyka.htm`, 2000.

[139] Howard Raiffa. *The Art and Science of Negotiation.* Harvard University Press, Cambridge, USA, 1982.

[140] Amelie Oksenberg Rorty. *Explaining Emotions.* University of California Press, Los Angeles, USA, May 1980.

[141] Amelie Oksenberg Rorty. The historicity of psychological attitudes: Love is not love which alters not when it alteration finds. *Mind in Action*, pages 121–134, 1988.

[142] Jeffrey S. Rosenschein and Gilad Zlotkin. *Rules of Encounter.* The MIT Press, Cambridge, MA, 1994.

[143] Alvin E. Roth. *Axiomatic Models of Bargaining.* Springer-Verlag, Berlin, 1979.

[144] Ariel Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–110, 1982.

[145] Ariel Rubinstein. Choice of conjectures in a bargaining game with incomplete information. *Game Theoretic Models of Bargaining*, pages 99–114, 1985.

[146] T. Sandholm and V. Lesser. Issues in automated negotiation and electronic commerce: Extending the contract net framework. *First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 328–335, January 1995.

[147] Jean-Paul Sartre. *The Emotions: Outline of a Theory.* Philosophical Library, New York, 1948.

[148] Stanley Schacter and Jerome Singer. Cognitive, social and physiological determinants of emotional states. *Psychological Review*, 69:379–399, 1962.

[149] Matthias Scheutz. Useful roles of emotions in artificial agents: A case study from artificial life. In *AAAI*, pages 42–48, 2004.

[150] Amy M. Schmitter. 17th and 18th century theories of emotions. Stanford Encyclopedia of Philosophy. `http://plato.stanford.edu/entries/emotions-17th18th/`, 2006.

[151] Amy M. Schmitter. Ancient, medieval and renaissance theories of the emotions. Stanford Encyclopedia of Philosophy. `http://plato.stanford.edu/entries/emotions-17th18th/LD1Background.html`, 2006.

[152] John R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge U. Press, Cambridge, England, 1970.

[153] John R. Searle. *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge U. Press, Cambridge, 1979.

[154] John R. Searle and D. Vanderveken. *Foundations of Illocutionary Logic*. Cambridge U. Press, London, 1985.

[155] John Sellars. Justus lipsius (1547-1606). The Internet Encyclopedia of Philosophy, `http://www.iep.utm.edu/l/lipsius.htm`, 2006.

[156] Juan Manuel Serrano and Sascha Ossowski. An organizational metamodel for the design of catalogues of communicative actions. In *Proceedings of the 5th Pacific Rim International Workshop on Multi Agents*, pages 92–108, London, UK, 2002. Springer-Verlag.

[157] Roger Shank and Kenneth Mark Colby, editors. *Computer Models of Thought and Language*. W.H. Freeman, San Francisco, 1973.

[158] Erin Shaw, W. Lewis Johnson, and Rajaram Ganeshan. Pedagogical agents on the web. In *AGENTS '99: Proceedings of the third annual conference on Autonomous Agents*, pages 283–290, New York, NY, USA, 1999. ACM Press.

[159] C. Sierra, P. Faratin, and N. Jennings. A service-oriented negotiation model between autonomous agents. In *Proceeding of 8th European Workshop on Modeling Autonomous Agents in a Multi-Agent World (MAAMAW-97)*, pages 17–35, Ronneby, Sweden, 1997.

[160] Herbert A. Simon. Motivational and emotional controls of cognition. *Psychological Review*, 74(1):29–39, 1967.

[161] David Simpson. Lucretius (c. 99 - c. 55 bce). The Internet Encyclopedia of Philosophy, `http://www.iep.utm.edu/l/lucretiu.htm`, 2006.

[162] Munindar P. Singh. Agent communication languages: Rethinking the principles. *Computer*, 31(12):40–47, 1998.

[163] Push Singh. Em-one: An architecture for reflective commonsense thinking. PhD Thesis. `http://web.media.mit.edu/~push/push-thesis.pdf`, 2005.

[164] Aaron Sloman. Why robots will have emotions. In *Proceeding of IJCAI*, University of Sussex, UK, 1981.

[165] Robert Solomon. Emotions and choice. *Explaining Emotions*, pages 251–281, 1980.

[166] K. P. Sycara. Persuasive argumentation in negotiation. *Theory and Decision*, 28:203 – 242, 1990.

[167] Katia Sycara. Resolving goal conflicts via negotiation. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, 1988.

[168] Katia Sycara. Argumentation: Planning other agents' plans. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, 1989.

[169] Y. Takeuchi, Y. Katagiri, and T. Takahashi. Learning enhancement in web contents through interagent interaction. In M. Hirose, editor, *INTERACT 2001: Proceedings of the Eight Conference on Human-Computer Interaction*, pages 480–487, 2001.

[170] Jean Tirole. *The Theory of Industrial Organization*. The MIT Press, Cambridge, MA, 1988.

[171] M. Toda. The design of the fungus eater: A model of human behavior in an unsophisticated environment. *Behavioral Science*, 7:164 – 183, 1962.

[172] Masanao Toda, editor. *Man, Robot and Society*. Martinus Nijhoff Publishing, Boston, MA, 1982.

[173] P. Valdez and Albert Mehrabian. Effects of color on emotions. *Journal of Experimental Psychology: General*, 123:394–409, 1994.

[174] Juan D. Velásquez. When robots weep: emotional memories and decision-making. In *AAAI '98/IAAI '98: Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence*, pages 70–75, Menlo Park, CA, USA, 1998. American Association for Artificial Intelligence.

[175] José M. Vidal. *Fundamentals of Multiagent Systems*. Unpublished, 2006.

[176] D. M. Wegner and T. Wheatley. Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54(7):480–492, 1999.

[177] Gerhard Weiss, editor. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press, Cambridge, Massachusetts, 1999.

[178] Michael Wheeler. Negotiating complex deals and disputes. Harvard Business School, http://www.hbs.edu/mba/academics/coursecatalog/2240.html, 2006.

[179] Uri Wilensky. NetLogo: Center for connected learning and computer-based modeling, Northwestern University. Evanston, IL, 1999. http://ccl.northwestern.edu/netlogo/.

[180] Jonathan Wilkenfeld, Sarit Kraus, Kim M. Holley, and Michael A. Harris. Genie: a decision support system for crisis negotiations. *Decis. Support Syst.*, 14(4):369–391, 1995.

[181] David Willer. *Network Exchange Theory*. Praeger Publishers, an imprint of Greenwood Publishing Group, Inc., 88 Post Road West, Westport, CT, 1999.

[182] Richard Wollheim. *On the Emotions*. Yale University Press, New Haven, 1999.

[183] Michael J. Wooldridge. *An introduction to multiagent systems*. John Wiley & Sons, LTD, Baffins Lane, Chichester, West Sussex, England, 2001.

[184] Ian Wright, Aaron Sloman, and Luc Beaudoin. Towards a design-based analysis of emotional episodes. *Philosophy, Psychiatry, and Psychology*, 3:101–126, 1996.

[185] Dajun Zeng and Katia Sycara. Bayesian learning in negotiation. *International Journal of Human Computer Systems*, 48:125–141, 1998.

[186] Frederik Zeuthen. *Problems of Monopoly and Economic Walfare*. G. Routledge & Sons, London, 1930.

Appendix A

Broadening the Semantic Coverage of Agent Communicative Acts

Abstract: Communicative acts-based ACLs specify domain-independent information about communication and relegate domain-dependent information to an unspecified content language. This is reasonable, but the ACLs cover only a small fraction of the domain-independent information possible. As a key element of modern ACLs, the set of communicative acts needs to be as complete as possible to allow agents to communicate the widest range of information with agreed-upon semantics. This paper describes a new approach to broaden the semantic coverage of agent communicative acts. It provides agents with the ability to express more of the semantics of human languages and yields a more powerful ACL. We first describe the main meaning categories and semantics for an ACL, which we derive from prior work on speech-act classifications. Next, we prove the resultant semantic coverage. Finally, we present some example applications, which demonstrate that the approach can combine the benefits of the FIPA ACL with Ballmer and Brennenstuhl's speech act classification, resulting in a more expressive and efficient ACL.

As a critical element of multi-agent systems and a key to the successful application of agents in commerce and industry, modern agent communication languages (ACLs), such as the FIPA ACL, provide a standardized set of performatives denoting types of communicative actions. Such ACLs have been designed as general purpose languages to simplify the design of multi-agent systems. However, recent research shows that these ACLs do not support adequately all relevant types of interactions. Serrano and Ossowski [156] report a need for new *ad hoc* sets of performatives in certain contexts, which the FIPA ACL does not support. Singh [162] points out that agents from different venders or even different research projects cannot communicate with each other. In [76], Kinny shows that the FIPA ACL has a confusing amalgam of different formal and informal specification techniques whose net result is ambiguous, inconsistent and underspecified communication. He proposes a set of requirements and desiderata against which an ACL specification can be judged, and briefly explores some of the shortcomings of the FIPA ACL and its original design basis.

Early work on communicative act-based ACLs, such as KQML and FIPA, separated the communication problem into three layers—a message transport layer providing the mechanics of a communication, a domain-independent layer of communication semantics, and a domain-dependent content layer. The ACL speech acts were intended to describe the domain-independent middle layer. The problem is that the 22 communicative acts in the current FIPA ACL cover only a small fraction of the domain-independent concepts that an agent might want to express. For example, one agent can `inform` another of a domain concept using the FIPA ACL, but cannot `promise` another something. If an agent wants to make a promise, its only recourse is to express it in the content language, for which there typically is no standardized support.

Therefore, a larger set of communicative acts would be desirable in an ACL to improve understanding among agents. Recognizing that the ~4800 speech acts in [10] would be desirable but impractical to use individually, we describe a feasible approach to broaden the semantic coverage of ACLs by formalizing speech act categories that subsumes the ~4800, enabling the meanings of all the speech acts to be conveyed. Different from [21], we focus on the standard messages used for communication instead of designing a conversation protocol.

Specifically, Section A.1 describes prior work on a comprehensive classification of speech acts by Austin, Searle, and Ballmer and Brennenstuhl. The main meaning categories and their semantics are given in Section A.2, where we use FIPA's formal semantic language to represent the semantics of our communicative act categories. This permits our approach to combine the benefits of the FIPA ACL with a broader set of communicative acts. Finally, Section A.3 proves the semantic coverage by comparing it with the FIPA ACL, and several example applications are described in Section A.4.

A.1 Research Background

Current ACLs derive their language primitives from the linguistic theory of speech acts, originally developed by Austin [8]. The most important part of his work was to point out that human natural language can be viewed as *actions* and people can perform things by speaking. Austin also classified illocutionary acts as verfictives, exercitives, commissives, behabitives and expositives [8]. The classification has been criticized for overlapping categories, too much heterogeneity in categories, ambiguous definitions of classes, and misfits between the classification of verbs and the definition of categories [10, 154].

Austin's work was extended by Searle [154, 152, 153], who posited that an illocutionary speech act forms the minimum meaningful unit of language. He classified speech acts into five categories: assertives, directives, commissives, declaratives, and expressives. Searle's speech act theory focusses on the speaker. The success of a speech act depends on the speaker's ability to perform a speech act that should be understandable and successful.

Ballmer and Brennenstuhl [10] criticize six aspects of Searle's classification: clarity, definition of declaratives as a speech act type, principles used in the classification, selection of illocutionary verbs, vague definition of the illocutionary point, and vagueness of the line between illocutionary force and propositional content. They propose an alternative classification, which contains both simple linguistic functions such as expression and appeal, and more complex functions such as interaction and discourse. Models for alternative actions are formed and verbs are classified according to the phases of the model.

Ballmer and Brennenstuhl's classification has motivated us to rethink the speech acts used in ACLs. Since the classification is based on an almost complete domain (~4800 speech acts) and the authors claim they provide a "theoretically justified" classification "based explicitly and systematically on linguistic data", we believe that to generate a speech act set for ACLs based on their classification will be a powerful way to represent meaning. However, this classification is not perfect: the classification for English is obtained by translating the verbs of a German classification, the names of the categories are not systematically chosen, and there is no formal semantic representation for the categories. However, most of these problems can be fixed by rebuilding the categories. Thus, we endeavour herein to derive a reasonable set of categories for agent communication from their theory and to give a formal semantics using more typical English names.

A.2 Semantic Description

This section describes the semantic categories for a relatively complete set of speech activity verbs, derived from the classification in [10]. The categories reflect an ontological and a conceptual structuring of linguistic behaviour. The main categories and their relationships are represented in Figure A.1. The top node, Speech Acts, represents the entire set of speech acts in human language and the four major groups—Emotion Model, Enaction Model, Interaction Model and Dialogic Model—represent four basic functions of linguistic behaviour.

The *Emotion Model* is the most speaker-oriented and focusses on representing kinds of emotional states of a human or agent.

The *Enaction Model* is a function directed toward a hearer, by which a speaker tries to control the understanding of the hearer.

The *Interaction Model* is a function involving speaker and hearer in mutual verbal actions. This group includes three sub-categories to represent different degrees of the mutual competition: (1) in the Struggle Model, the speaker tries to get control over the hearer, or the speaker is more competitive in controlling mutual verbal actions; (2) in contrast, the hearer is more competitive in the Valuation Model; and (3) in the Institutional Model, the hearer and speaker are equally competitive.

The *Dialogic* Model covers a kind of reciprocal cooperation where there is a better-behaved and more rigidly organized verbal interaction. Its three sub-categories focus on different types of the content and the organization: (1) the Discourse Model focusses on the organization and types of discourse; (2) the Text Model focusses on the textual assimilation and processing of reality, i.e. the specific knowledge involved; (3) and the Theme Model focuses on the process of thematic structuring and its results, in other words, the structure or organization of some knowledge system.

In the above ontology, the four basic models can be divided into unilateral and multilateral models. The Emotion Model and Enaction Model are unilateral, because

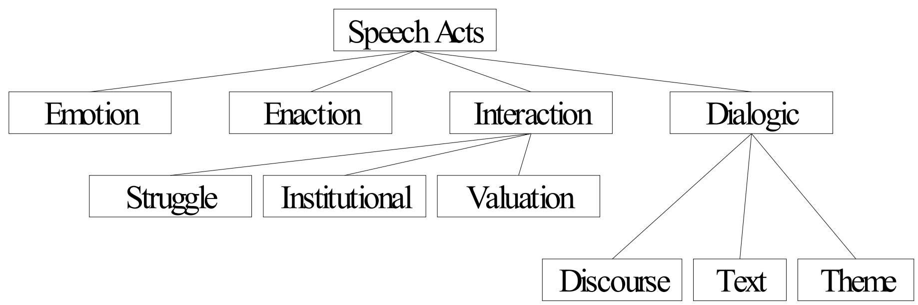


Figure A.1: Ontology of the Main Speech Act Categories

they focus on a single speech action. The Interaction Model and Dialogic Model are multilateral, because they consider the response from a hearer. The Emotion Model and Interaction Model are more original and less constrained, and the Enaction Model and Dialogic Model are more institutionalized and controlled. Practically, these four basic models may be combined.

We next define several formal semantic model notations and then describe the detailed semantics for the meaning categories.

Formal Semantic Model Notations

The semantic model used in representing the categories in this paper follows the formal semantic language described for the FIPA ACL [50]. Components of the formalism are

- $p, p_1, ...$ are closed formulas denoting propositions;
- ϕ, ψ are formula schemes, which stand for any closed proposition;
- i, j are schematic variables denoting agents.

The mental model of an agent is based on four primitive attitudes: belief (what the agent knows or can know); desire (what the agent desires); intention (an agent's persistent goal that could lead to some actions); and uncertainty. They are respectively formalized by operators B, D, I, and U:

- $B_i p$ agent i (implicitly) believes (that) p;
- $D_i p$ agent i desires that p currently holds;
- $I_i p$ agent i intends a persistent goal p;
- $U_i p$ agent i is uncertain about p, but thinks that p is more likely than $\neg p$;

We use the abbreviations:

- $Bif_i \phi \equiv B_i \phi \vee B_i \neg \phi$, which means that agent i believes either ϕ or $\neg \phi$.
- $Uif_i \phi \equiv U_i \phi \vee U_i \neg \phi$, which means that either agent i is uncertain about ϕ (ϕ is more likely) or $\neg \phi$ ($\neg \phi$ is more likely).

To support reasoning about action, we also introduce operators *Feasible*, *Done* and *Agent*:

- $Feasible(a, p)$ means that an action a can take place and, if it does, then p will be true.
- $Done(a, p)$ means that when p is true, then action a takes place.
- $Agent(i, a)$ means agent i is the agent who performs action a.

Generally, the components of a speech act model involved in a planning process should contain both the conditions that have to be satisfied for the act to be planned and the reasons for which the act is selected. The former is termed *FP* (feasibility preconditions) and the latter *RE* (rational effect). We use the same model here, represented as

$$\begin{aligned} &< i, act \quad (j, C) > \\ &FP : \phi_1 \\ &RE : \phi_2 \end{aligned} \tag{A.1}$$

where i is the sender or speaker, j the recipient or hearer, *act* is the name of the speech act, C is the semantic content, and ϕ_1 and ϕ_2 are propositions.

Emotion Model

The Emotion Model focusses on representing the emotional states of a human or agent. We assume there is a finite set of emotions, E, represented as

$$E = \{e_+, e_0, e_-\} \tag{A.2}$$

where e_+ is an emotion in the set of positive emotions, which is characterized by or displaying a kind of certainty, acceptance or affirmation (about the content involved), such as $\{happy, love, ...\}$; e_0 is in the set of neutral emotions, which does not show any tendency, such as $\{hesitate, ...\}$; e_- is in the set of negative emotions, which intends or expresses a kind of negation, refusal or denial, such as $\{angry, sad, afraid, ...\}$.

The Emotion Model is represented as follows:

$$\begin{array}{lll} <i, & em & (j, \phi)> \\ FP: & \neg B_i & (B_j Agent(i, em(\phi))) \wedge D_i(B_j Agent(i, em(\phi))) \\ RE: & B_j & Agent(i, em(\phi)) \end{array} \tag{A.3}$$

where $em \in E$ and the semantic content ϕ can be empty. Here, desire D is used instead of the stronger notion I, since emotions are easy to show for humans. This model represents that agent i sends a message to j that i has emotion em about ϕ or i is in the state of em when ϕ is empty. The FP shows that, when agent i does not believe that agent j knows that i is currently in emotion em about ϕ, and i desires that j knows it, then this message can be sent. The RE shows that the desired result is that agent j believes that i is in emotion em about ϕ.

To simplify usage of this model, we can directly use e_+, e_0, or e_- as communicative

Table A.1: Foundational Meaning Units of Emotional Speech Acts

+	0	-
happy	N/A	sad
love	N/A	hate
excited	nervous	angry
desire	hesitate	fear
N/A	shocked	N/A

acts. In this case, we focus on the effect of the emotion speech act on the content ϕ. That is, for a positive effect, i hopes j knows that i has an intention on ϕ; for a negative one, i hopes j knows that i has a negative intention on ϕ; for a neutral one, i shows its attitude is uncertain about ϕ. Just like human interactions, we do not have to know the precise value of an attitude. Instead, we just need to know that something is viewed favourably, unfavourably or neutrally.

However, detailed emotions are also desirable in some cases. To make this usable, we generate a set of foundational meaning units from 155 emotion speech acts listed in [10]. Table A.1 gives the foundational meaning units of emotions that combine the idea from [1, 43], and they are organized with consideration of positive, neutral and negative values.

In Table A.1, each row represents a kind of meaning unit. In the first row, *sad* has the opposite meaning of *happy*. *Hate* has the opposite meaning of *love* in the second row. *Excited* represents a positive attitude to something with strong feeling, *nervous* represents a strong uncertain feeling about something and *angry* represents a strong negative feeling about something. In the fourth row, *desire* shows a feeling to get something, *hesitate* shows no intentions or some uncertainty and *fear* shows a feeling to avoid something. In the last row, *shocked* shows a neutral feeling about surprise.

Enaction Model

In the Enaction Model, the speaker more or less coercively attempts to get the hearer to do something by expressing an idea, wish, intention, proposal, goal etc. There are many speech acts in this group. To organize them and simplify the usage, we define the set of enactions as:

$$EN = \{en_+, en_-\} \tag{A.4}$$

Unlike the Emotion Model, which describes emotions, the Enaction Model tries to make a hearer do something. Thus, there are no neutral enactions: if agent i does not want j to do anything, i does not have to send any message to j. en_+ is an action in the set of positive enactions, such as $\{intend, desire, askfor$, $encourage$, ...$\}$; en_- is an action in the set of negative enactions, such as $\{$ $warning$, $cancel$, ...$\}$.

The Enaction Model can be defined as:

$$\begin{aligned}
<i, \quad & en_{\pm}(j, \phi) > \\
FP: \quad & \neg B_i\phi \wedge D_i\phi \wedge B_i(B_j\phi \wedge \neg D_j\phi) \qquad for\ en_+ \\
& \neg B_i\neg\phi \wedge D_i\neg\phi \wedge B_i(B_j\neg\phi \wedge \neg D_j\neg\phi) \quad for\ en_- \\
RE: \quad & Done(en_{\pm}(\phi))
\end{aligned} \tag{A.5}$$

where $en_{\pm} \in EN$. This model represents agent i sending a message to j to ask j to do $en_{\pm}$ on ϕ. The FP shows that this message could be sent for en_+ when i does not believe that i can do ϕ and it desires ϕ, while i believes that j can do it, but j does not want to do it. FP is the same for en_-, except ϕ is replaced by $\neg\phi$. The expected result is that $en_{\pm}$ on ϕ is done. Practically, j could just add the action to its action queue for a positive enaction (in this case, $Done(en_+(\phi)) = Done(\phi)$) or delete it from its queue for a negative enaction.

Interaction Model

The Interaction Model is a function involving a speaker and a hearer mutually interacting. We assume an interaction set IN and the communicative act set $Acts$ so that $IN \subseteq Acts$, and for some $in \in IN$ and $act \in Acts$, $\exists rule : in \to act$, such that:

$$\begin{aligned} < i, \quad & in(j, (a, goal)) > \\ FP: \quad & I_i goal \wedge \neg B_i Done(a) \wedge D_i Done(a) \wedge B_i(Agent(j, a) \wedge \neg D_j Done(a)) \\ RE: \quad & Done(a) \wedge (< j, act(i, (a', goal - a) > \vee < j, succeed(i, goal) >) \\ & \vee < j, fail(i, goal) > \end{aligned} \tag{A.6}$$

where a, a' are actions, and $goal$ can be a plan or a sequence of actions. This model represents agent i sending a message to j to ask j to do action a for some $goal$. The FP shows that i intends to achieve the $goal$, so i desires to do a but cannot do it itself, and i believes that j can do it. However, j does not desire to do it. The expected result is j does a first, and then generates another message back to i. This reply message follows the rule $in \to act$. Generally, the message has the form $< j, act(i, (a', goal - a) >$, which means that after j has done a, it generates another action a' and reduces the $goal$. In some cases, for example after j has done a and the $goal$ is already achieved, then j sends back the message $< j, succeed(i, goal) >$, which means that the $goal$ is achieved. Another extreme case is that j finds out that the $goal$ is impossible to be achieved, then it sends back message $< j, fail(i, goal) >$, which means the $goal$ is unachievable.

There are three subcategories of the interaction model representing different degrees of the mutual competition: Struggle Model, Institutional Model and Valuation Model. In the Struggle Model, the speaker tries to get control over the hearer or the speaker is more competitive in controlling mutual verbal actions. In this case, the rule $in \to act$ is decided by the speaker or sender i.

In the Institutional Model, the hearer and speaker are equally competitive. For example, the establishment of a behaviour in an institution equally affects the upholders of and the participants in the institution, especially when entering an institution and thereby adopting its norms, following its norms and rules, violating them and being pursued by the upholders of the institution. Thus, the agents i and j should have some common rule system defined in advance.

In the Valuation Model, the hearer is more competitive, so it decides which communication act to use in its reply. That is, the rule $in \rightarrow act$ is decided by agent j after its evaluation of the previous message. Details of the Valuation Model cover both positive and negative valuations of actions, persons, things and states of affairs.

Dialogic Model

The Dialogic Model covers a kind of reciprocal cooperation, and is a more regular and constrained verbal interaction. For this model, we at first assume a dialogic speech act set DS and the communicative act set $Acts$ so that $DS \subseteq Acts$, and for some $d \in DS$ and $act \in Acts$, $\exists rule : d \rightarrow act$, such that

$$\begin{aligned} < i, \quad & d(j, \phi) > \\ FP: \quad & B_i\phi \wedge D_i B_j \phi \\ RE: \quad & B_j\phi \wedge < j, act(i, \phi') > \end{aligned} \tag{A.7}$$

For agent i to send a message to j about ϕ in this model, agent i believes ϕ, and i desires j to believe it. The expected result is that j believes ϕ and j replies to i with another message about a new ϕ, which is the reasoning result of agent j, and the communicative act used in the message follows the rule $d \rightarrow act$.

Corresponding to the three subcategories that focus on different types of content and organization, we can define three types for ϕ:

- The Discourse Model focusses on the organization and types of discourse, so ϕ points to some kind of type or organization that is predefined. For example, according to the status of a discourse, it could be {*beginning discourse, being in discourse, discourse inconvenience, reconciliation of discourse, ending discourse*}; according to the attitude for some content, it could be {*accept, refuse, cancel* }; according to the number of agents involved in the discourse, it could be {*discourse with several speakers, discourse with one speaker, ...* }; or a kind of irony, joke etc.

- The Text Model focusses on the textual assimilation and processing of the specific knowledge involved, i.e. ϕ describes some knowledge about perceiving reality, producing texts, systematically searching for data etc.

- The Theme Model focusses on the process of thematic structuring and its results, in other words, ϕ points to the structure or organization of some knowledge system.

A.3 Proof of Semantic Coverage

The FIPA ACL has four primitive communicative acts, and its other communicative acts are composed of the primitive acts or are composed from primitive messages by substitution or sequencing [50]. The four primitive acts are:

- The Assertive Inform:

$$\begin{aligned} < i, & \quad inform(j, \phi) > \\ FP: & \quad B_i\phi \wedge \neg B_i(Bif_j\phi \vee Uif_j\phi) \\ RE: & \quad B_j\phi \end{aligned}$$

- The Directive Request:

$$\begin{aligned} <i, \quad & request(j,a)> \\ FP: \quad & FP(a)[i \backslash j] \wedge B_i Agent(j,a) \wedge B_i \neg PG_j Done(a) \\ RE: \quad & Done(a) \end{aligned}$$

  where $FP(a)$ denotes the feasibility preconditions of a; $FP(a)[i \backslash j]$ denotes the part of the FPs of a that are mental attitudes of i; and $PG_i P$ means that i has P as a persistent goal.

- Confirming an Uncertain Proposition (Confirm):

$$\begin{aligned} <i, \quad & confirm(j,\phi)> \\ FP: \quad & B_i \phi \wedge B_i U_j \phi \\ RE: \quad & B_j \phi \end{aligned}$$

- Contradiction Knowledge (Disconfirm):

$$\begin{aligned} <i, \quad & disconfirm(j,\phi)> \\ FP: \quad & B_i \neg\phi \wedge B_i (U_j \phi \vee B_j \phi) \\ RE: \quad & B_j \neg\phi \end{aligned}$$

Furthermore, among the 22 communicative acts of FIPA ACL, the composite ones corresponding to the above four primitive acts are as shown in Figure A.2:

- Inform: accept-proposal, agree, failure, inform-if, inform-ref, not-understood, propagate, propose, proxy, reject-proposal, request-when, request-whenever, subscribe

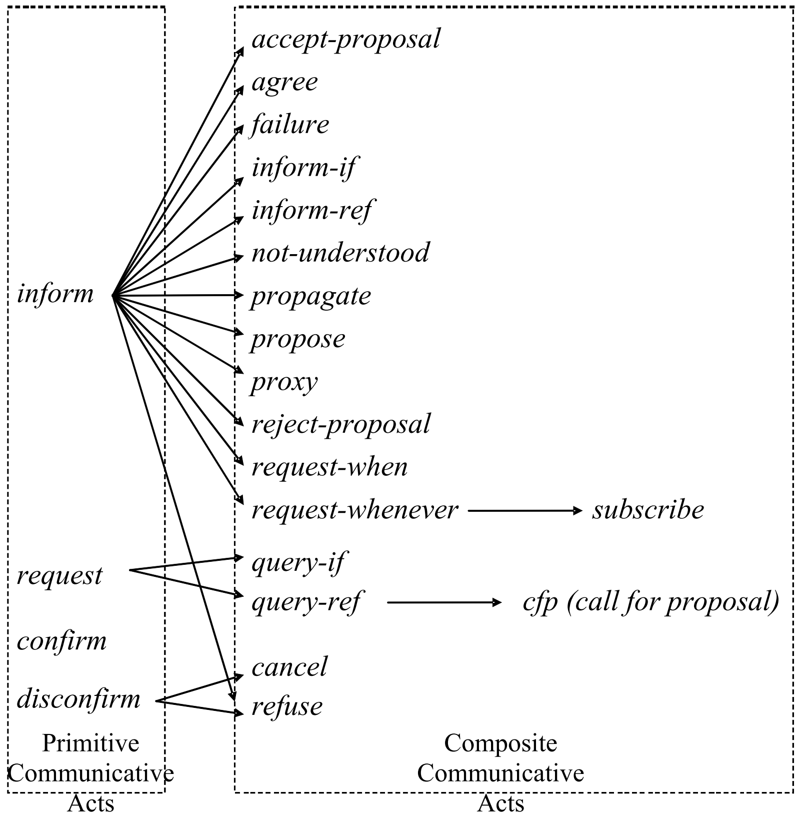


Figure A.2: Relationship of FIPA Primitive and Composite Communicative Acts

- Request: cfp(call for proposal), query-if, query-ref
- Confirm: N/A
- Disconfirm: cancel, refuse

It can be seen that the composite communicative acts relate to the primitive ones unevenly. Most of the communicative acts are derived from *inform*, and even the primitive acts, *confirm* and *disconfirm*, are special cases of *inform*, which can be proved as follows.

Lemma 1. *In the primitive communicative acts of FIPA ACL, confirm* ($< i$, *confirm* $(j, \phi) >$) *is a special case of inform* ($< i$, *inform*$(j, \phi) >$)*.*

Proof. Comparing the definitions of *confirm* and *inform*, we see they have the same message body format and rational effect — RE. The only difference is the feasibility preconditions — *FP*. We can then try to prove that *FP* of *confirm* is a sufficient but not necessary condition of *inform*. That is, when the *FP* of *confirm* is satisfied, the *FP* of *inform* is also satisfied, or the satisfaction of *FP* of *confirm* can trigger an *inform* message; alternatively, the *FP* of *confirm* is not necessary for sending an *inform*.

FP of *inform* is:

$$
\begin{aligned}
& B_i\phi \wedge \neg B_i(Bif_j\phi \vee Uif_j\phi) \\
\equiv \quad & B_i\phi \wedge \neg B_i((B_j\phi \vee B_j\neg\phi) \vee (U_j\phi \vee U_j\neg\phi)) && \text{(A.8)} \\
\equiv \quad & B_i\phi \wedge (\neg B_iB_j\phi \vee \neg B_iB_j\neg\phi \vee \neg B_iU_j\phi \vee \neg B_iU_j\neg\phi) \\
\equiv \quad & B_i\phi \wedge (\neg B_iB_j\phi \vee B_iB_j\phi \vee \neg B_iU_j\phi \vee B_iU_j\phi) && \text{(A.9)} \\
\equiv \quad & (B_i\phi \wedge \neg B_iB_j\phi) \vee (B_i\phi \wedge B_iB_j\phi) \vee (B_i\phi \wedge \neg B_iU_j\phi) \vee && \text{(A.10)} \\
& (B_i\phi \wedge B_iU_j\phi)
\end{aligned}
$$

where Equation (A.8) is derived from the definitions of Bif_ip and Uif_ip. We get Equation (A.9) since agent i not believing j believes ϕ usually means the same as agent i believing j does not believe ϕ.

From Equation (A.11), the last part $B_i\phi \wedge B_iU_j\phi$ is exactly the *FP* of *confirm*. When *FP* of *confirm* is satisfied, that is, when $B_i\phi \wedge B_iU_j\phi$ is true, then Equation (A.11) will also be true. However, *FP* of *confirm* is not a necessary condition, since only if one of $B_i\phi \wedge \neg B_iB_j\phi$, $B_i\phi \wedge B_iB_j\phi$ and $B_i\phi \wedge \neg B_iU_j\phi$ is satisfied, Equation (A.11) will also be satisfied.

Thus, *confirm* is a special case of *inform*.

Lemma 2. *In the primitive communicative acts of the FIPA ACL, disconfirm (*$< i$*, disconfirm (*j*,* ϕ*)* $>$*) is a special case of inform (*$< i, inform(j, \neg\phi) >$*).*

Proof. Comparing the definitions of *disconfirm* and *inform*, we have

$$\begin{aligned} < i, &\quad inform(j, \neg\phi) > \\ FP: &\quad B_i\neg\phi \wedge \neg B_i(Bif_j\neg\phi \vee Uif_j\neg\phi) \\ RE: &\quad B_j\neg\phi \end{aligned} \tag{A.11}$$

Thus, we get the same rational effect format — RE. Let's compare the feasibility preconditions — FP, and similarly Equation (A.11) can be changed to:

$$\begin{aligned} & B_i\neg\phi \wedge \neg B_i(Bif_j\neg\phi \vee Uif_j\neg\phi) \\ \equiv \quad & B_i\neg\phi \wedge \neg B_i((B_j\neg\phi \vee B_j\phi) \vee (U_j\neg\phi \vee U_j\phi)) \\ \equiv \quad & B_i\neg\phi \wedge \neg B_i(B_j\neg\phi \vee B_j\phi \vee U_j\neg\phi \vee U_j\phi) \\ \equiv \quad & B_i\neg\phi \wedge \neg B_i((U_j\phi \vee B_j\phi) \vee (B_j\neg\phi \vee U_j\neg\phi)) \\ \equiv \quad & B_i\neg\phi \wedge (\neg B_i(U_j\phi \vee B_j\phi) \vee \neg B_i(B_j\neg\phi \vee U_j\neg\phi)) \\ \equiv \quad & (B_i\neg\phi \wedge \neg B_i(U_j\phi \vee B_j\phi)) \vee (B_i\neg\phi \wedge \neg B_i(B_j\neg\phi \vee U_j\neg\phi)) \end{aligned} \tag{A.12}$$

From Equation (A.12), the first part $B_i\neg\phi \wedge \neg B_i(U_j\phi \vee B_j\phi)$ is exactly the FP of *disconfirm*. When FP of *disconfirm* is satisfied, that is, when $B_i\neg\phi \wedge \neg B_i(U_j\phi \vee B_j\phi)$ is true, then Equation (A.12) will also be true, which will trigger message $< i, inform(j, \neg\phi) >$. However, FP of *disconfirm* is not a necessary condition, since if $B_i\neg\phi \wedge \neg B_i(B_j\neg\phi \vee U_j\neg\phi)$ is satisfied, Equation (A.12) will also be satisfied.

Thus, we proved that FP of *disconfirm* is a sufficient but not necessary condition to trigger message $< i, inform(j, \neg\phi) >$. In other words, *disconfirm* ($< i, disconfirm(j, \phi) >$) is a special case of *inform* ($< i, inform(j, \neg\phi) >$).

So far, we can conclude that there are actually two foundational communicative acts *inform* and *request*. If we can prove that our approach covers the semantic meaning of these two communicative acts, then our approach covers all the semantic meanings of the FIPA communicative acts, since the others can be derived from these two by adding constraints.

However, we think *inform* is too general. Considering $\neg B_i(Bif_j\phi \vee Uif_j\phi)$ in *FP* of *inform*, it actually lists all the possibility of j's knowledge about ϕ, such that: i does not believe j believes ϕ, or i does not believe j believes not ϕ, or i does not believe j is uncertain about ϕ, or i does not believe j is uncertain about not ϕ. Since at least one of them will be true, $\neg B_i(Bif_j\phi \vee Uif_j\phi)$ will always be true. So *FP* of *inform* can be simplified to $B_i\phi$, which is reasonable because only if agent i has the belief ϕ can it inform j about ϕ. While, we still think it should not ignore the desire to have j to believe ϕ, no matter what i believes or does not believe j's knowledge about ϕ, if i does not have any desire to have j believe ϕ, why does i want to send the message to j?

Based on the above analysis, now we prove that our approach covers the semantic meaning of the two foundational communicative acts *inform* and *request*.

Lemma 3. *The Dialogic model covers the semantic meaning of FIPA's inform.*

Proof. According to our above analysis of *inform*, $\neg B_i(Bif_j\phi \vee Uif_j\phi)$ did not supply any of i's opinion on j's knowledge about ϕ, and i's desire for j to know about ϕ was also be ignored. By adding these considerations, we can then represent *inform* with more precise semantic meaning as:

$$\begin{aligned} < i, \quad & inform(j, \phi) > \\ FP: \quad & B_i\phi \wedge D_iB_j\phi \\ RE: \quad & B_j\phi \end{aligned}$$

Then it has a format similar to the semantic representation of the dialogic model, and the difference is in *RE*. For the dialogic model, we assume a dialogic communicative act set *DS* and the communicative act set *Acts* with $DS \subseteq Acts$, and for some $d \in DS$ and $act \in Acts$, $\exists rule : d \rightarrow act$, If $d \in DS$ is a terminal symbol, that is, there is no rule from d to something else, then in this case, $< j, act(i, \phi') >$ in RE can be ignored, so that we can get the same semantic meaning of *inform*. Thus, we can use the *dialogic model* to represent the semantic meaning of *inform*.

Lemma 4. The interaction model covers the semantic meaning of FIPA's *request*.

Proof. Let's first consider the definition of *request*, which is used to request a receiver to perform some action. Usually it presumes feedback from the receiver. FP of *request* involves three parts:

- $FP(a)[i\backslash j]$: denotes the part of the FPs of action a that are mental attitudes of agent i. We do not know exactly what the mental attitudes will be, although they should satisfy the following conditions for sending out a request: agent i should intend to have action a done—$I_i Done(a)$; and i cannot do a by itself.

- $B_i Agent(j, a)$: i believes that j is the only agent that can perform a.

- $B_i \neg PG_j Done(a)$: this part (in page 36 of [50]) is also presented as $\neg B_i I_j Done$ (a) (in page 25 of [50]), which roughly points out a required condition: i does not believe j intents to do a.

The *goal* in the interaction model denotes a plan or a sequence of actions. To get comparable format of the interaction model, we can let the goal involve only one action, that is, let $goal = Done(a)$. Then the interaction model can be simplified to:

$$\begin{aligned}
< i, \quad & in(j, a) > \\
FP: \quad & I_i Done(a) \wedge \neg B_i Done(a) \wedge D_i Done(a) \wedge \\
& B_i(Agent(j, a) \wedge \neg D_j Done(a)) \\
RE: \quad & Done(a) \wedge < j, succeed(i, Done(a)) >
\end{aligned} \tag{A.13}$$

In the BDI model, intention is generated from desire. If we separate desires into intentional desires (I) and non-intentional desires (NI), then we can represent $D_i p$ to be $I_i p \vee NI_i\, p$, such that Equation (A.13) becomes

$$
\begin{aligned}
& I_i Done(a) \wedge \neg B_i Done(a) \wedge D_i Done(a) \wedge B_i(Agent(j,a) \wedge \neg D_j Done(a)) \\
\equiv \quad & I_i Done(a) \wedge \neg B_i Done(a) \wedge (I_i Done(a) \vee NI_i\, Done(a)) \\
& \wedge B_i(Agent(j,a) \wedge \neg D_j Done(a)) \\
\equiv \quad & I_i Done(a) \wedge \neg B_i Done(a) \wedge B_i(Agent(j,a) \wedge \neg D_j Done(a)) \\
\equiv \quad & I_i Done(a) \wedge \neg B_i Done(a) \wedge B_i Agent(j,a) \wedge B_i \neg D_j Done(a) \\
\equiv \quad & I_i Done(a) \wedge \neg B_i Done(a) \wedge B_i Agent(j,a) \wedge \neg B_i(I_i Done(a) \vee NI_i\, Done(a)) \\
\equiv \quad & (I_i Done(a) \wedge \neg B_i Done(a) \wedge B_i Agent(j,a) \wedge \neg B_i I_i Done(a)) \vee \\
& (I_i Done(a) \wedge \neg B_i Done(a) \wedge B_i Agent(j,a) \wedge \neg B_i\, NI_i\, Done(a)) \qquad \text{(A.14)}
\end{aligned}
$$

The first part of Equation (A.14) has a format similar to the *FP* of *request*:

- $I_i Done(a) \wedge \neg B_i Done(a)$ corresponds to the first part of *FP* for *request*, which presents the detailed required conditions—agent i should intend to have action a done and i cannot do a by itself.
- $B_i Agent(j,a)$ is the same as the second part of *FP* for *request*.
- $\neg B_i I_i Done(a)$ is the same as the third part of *FP* for *request*. We did not use symbol *PG* in our approach, since *PG* is very similar to I, and here this part follows the format on page 25 of [50].

So far, we see that when *FP* of *request* is true, the Equation (A.14) will also be true, and message of *interaction model* will be triggered. However, the *FP* of *request* is not necessary for Equation (A.14) to be satisfied.

Let's continue to consider the *RE* of *request*, which is the same as the first part of *RE* of *interaction model*. However, the second part is also reasonable for *request*, since in most cases *request* implies feedback from the receiver.

Thus, *request* is a special case of *interaction model*, and the *interaction model* covers the semantic meaning of *request* in the FIPA ACL.

In summary, our approach covers the semantic meaning of the two foundational communicative acts, so it also covers all the semantic meanings of the communicative acts in the FIPA ACL. Moreover, our approach also covers additional semantic meanings. For example, our emotion model supplies a way to communicate emotions, which the FIPA ACL does not. We believe it is important to cover emotions in agent communicative acts, since other researchers [27, 176, 13] have already discovered that emotions influence human decision-making; unfortunately, this influence has traditionally been ignored.

A.4 Example Applications

This section provides several examples showing how these defined semantic categories can be used.

Example 1: Bob tells Sue that he loves her. Using the emotion model, the sender is Bob, the receiver is Sue and ϕ = Sue to yield the message on the left below. The expected result will be that Sue has a belief that Bob is in love with her. Since the FIPA ACL does not have a communicative act with a similar meaning, the content must include the expression of emotion, as shown in the message on the right.

(love	(inform
:sender Bob	:sender Bob
:receiver Sue	:receiver Sue
:content (Sue))	:content (Bob loves Sue))

The left message separates domain independent from domain dependent information better and is less ambiguous.

Example 2: Jack commands Bill to turn off the TV. Using the enaction model, the message to be sent is

(command

:sender Jack

:receiver Bill

:content (turn off the TV))

The expected result will be that Bill turns off the TV. The communicative act *command* implies a master-slave relationship between the sender and receiver. The FIPA ACL does not have a similar communicative act, so all the information must be put in the content, as in ***Example 1***, although it is more ambiguous.

Example 3: Bob and Jack work together to open a case with ID 011. Bob gets the key but it is broken. Jack is an expert in fixing keys, so Bob asks Jack to fix the key.

According to the interaction model, the message sent to Jack will be

(interact

:sender Bob

:receiver Jack

:content (fix key) (open case 011))

The goal "open case 011" implies a sequence of actions, which are assumed known to both sender and receiver in advance. Thus Jack tries to fix the key. If Jack fixes the key successfully, he will send a reply to Bob that Bob can pick up the key to open the case now, as shown in the message below on the left. If Jack cannot fix the key, he will then tell Bob that the goal failed, as shown on the right.

(interact	(fail
:sender Jack	:sender Jack
:receiver Bob	:receiver Bob
:content (pick-up key)	:content (open case 011)
(open case 011 - fix key))	

This model is especially useful for multiple agents working together on a project.

Example 4: Bill wants to tell Bob about the structure of the subway system in Boston, which includes the red line, orange line, green line and blue line. According to the dialogic model, the message sent to Bob would be

(structure

:sender Bill

:receiver Bob

:content (Subway in Boston: red line, orange line, green line, blue line)

The expected result will be that Bob records the structure information as one of his beliefs. We can also use FIPA's $inform$ to represent the above message, but the relationship of the subway system and those lines would have to be part of the content.

A.5 Conclusion and Future Work

Comparing our approach to the FIPA ACL reveals that:

Better coverage: our approach covers more of human semantics.

Precise semantics: we adopt the same formalism as used by FIPA for our four basic categories and subcategories.

Easy usage: An ACL must be easy to use, and the FIPA ACL has many successful uses. Instead of replacing it, we substitute our communicative acts and keep its message structure. We organize the communicative acts as an ontology with different abstract levels, so that a user or agent can more easily navigate through them to choose the desired ones.

Better understood: Easy usage requires that the ACL be well understood. However, the original categories given by Ballmer and Brennenstuhl's classification are poor, because the classification is obtained by translating German verbs and the names of the categories are not chosen systematically. We modified their classification by using typical English names, which should be more understandable.

Efficiency: Efficiency is desirable for an ACL. As can be seen in the above examples, our approach separates domain-independent from domain-dependent information better, which can shorten the message sent while improving the semantics.

In summary, our approach combines the benefits of the FIPA ACL and Ballmer and Brennenstuhl's speech act classification. It is more expressive in representing a broader range of domain-independent communication semantics, while remaining consistent with current approaches to ACLs. However, a better communicative act set with reasonable size still needs work. Instead of just considering the categories, some frequently used speech acts also need to be found for the communicative act set.

Appendix B

Sample Emotional Experiences

The following verbally described situations represent a balanced sample of emotional experiences [104].

Pleasant, arousing, and inducing dominance (+P+A+D):

You are in a cabin with snow falling outside. A fire crackles in the fireplace. You are reading a thrilling novel by one of your favorite authors.

Today, you were promoted at work. You are about to tell your spouse that now, finally, you can purchase the kind of home you both have always wanted.

Pleasant, arousing, and inducing submissiveness (+P+A–D):

You are a guest at a celebrity dinner. You almost feel as though you have crashed the party, because many of the people in attendance are well-known figures.

You are water-skiing on a quiet lake. This is something you have always wanted to learn. It is your first time, so falling is still a strong possibility.

Pleasant, unarousing, and inducing dominance (+P–A+D):

You are sitting on the edge of a dock on a warm day and your feet are dangling in the cool water. Your baited fishing line is out a ways in the water.

You are vacationing at a luxurious hotel on a tropical island. You have just finished lunch at the quiet outdoor restaurant of the hotel and are leaning back in your chair and enjoying a cool drink.

Pleasant, unarousing, and inducing submissiveness (+P–A–D):

You and your friend have been skiing all day. The two of you are now settled before a cozy fire in your friend's cabin. Outside, snow drifts gently to the ground.

You are in a forest at night. There is a campfire and you have a cup of your favorite hot beverage. The sky is so clear it looks like you could touch the stars.

Unpleasant, arousing, and inducing dominance (–P+A+D):

Another shopper has just cut in front of you in the line where you have been standing for the last half hour. You are telling the shopper that he/she must go to the end of the line.

You are the manager of a very nice apartment building. You have decided that you must raise rents for everyone in the building.

Unpleasant, arousing, and inducing submissiveness (–P+A–D):

You had your annual physical two days age. Your physician has called you to say that he needs to see you at his office first thing tomorrow morning.

It is late at night. You have driven to the market for a much-needed item. Accidentally, you have locked your keys in your car.

Unpleasant, unarousing, and inducing dominance (–P–A+D):

You are doing some light reading when the phone rings. It is someone trying to sell you magazine subscriptions. You tell them you are not interested.

You have just attended an art show that was drab and unappealing. You are stopped on your way out by a writer from one of the local papers who wants to know your opinion of the show.

Unpleasant, unarousing, and inducing submissiveness (–P–A–D):

You are attending a classical music concert that your spouse insisted you go to. The concert is almost over and all the selections so far have been uninteresting. You tried to go to sleep, but the noise kept you awake.

You have had a long and exhausting day at work. You now must wait for about

30-40 minutes for your ride home.

Note. The abbreviations +P and –P represent pleasant and unpleasant, +A and –A represent arousing and unarousing (new version are arousal and non-arousal), and +D and –D represent dominance- and submissiveness-inducing situations, respectively. The two situational descriptions listed under each of the eight categories (e.g., +P+A+D) were selected on the basis of mean emotional responses of participants to the situations obtained in previous studies.

Appendix C

Publications during Graduate Study

- Hong Jiang, José M. Vidal, and Michael N. Huhns. EBDI: An Architecture for Emotional Agents. In *Proceedings of AAMAS07: Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 2007.

- Hong Jiang and José M. Vidal. The Message Management Asynchronous Backtracking Algorithm. *Journal of Experimental and Theoretical Artificial Intelligence*, 2006. (accepted)

- Hong Jiang and José M. Vidal. From rational to emotional agents. In *Proceedings of the AAAI Workshop on Cognitive Modeling and Agent-based Social Simulation*, 2006.

- Hong Jiang, José M. Vidal, and Michael N. Huhns, Incorporating Emotions into Automated Negotiation. In *Proceedings of Agent Construction and Emotions (ACE 2006): Modeling the Cognitive Antecedents and Consequences of Emotion*, Vienna, Austria, April 2006.

- Hong Jiang and Michael N. Huhns. Broadening the Semantic Coverage of Agent Communicative Acts. *Agent-Oriented Information Systems III, Lecture Notes* *in Computer Science*, 2006. Vol.3529, p.32-47

- Hong Jiang and Michael N. Huhns. An Approach to Broaden the Semantic Coverage of ACL Speech Acts. In *Proceedings of the 24th International Conference on Conceptual Modeling*, Klagenfurt, Austria, Springer Verlag, *Lecture Notes in Computer Science*, October 2005. Vol. 3770, p.162-171

- Hong Jiang and José M. Vidal. Reducing Redundant Messages in the Asynchronous Backtracking Algorithm. In *Proceedings of the Sixth International Workshop on Distributed Constraint Reasoning*, 2005.