Emotion Modelling with Human Belief Revision in Computer Games

BY

Yan Ma

A Thesis Submitted to

Saint Mary's University, Halifax, Nova Scotia

in Partial Fulfillment of the Requirements for

the Degree of Master of Science in Applied Science

March, 2007, Halifax, Nova Scotia

Approved: Dr. Joseph MacInnes (Supervisor)

Approved: Dr. Sageev Oore (Examiner)

Dr.Camilla Holmvall (Examiner)

Dr.Thomas Trappenberg (Examiner)

Date: Dec 19, 2006

Your file Votre référence
ISBN: 978-0-494-29011-8
Our file Notre référence
ISBN: 978-0-494-29011-8

Emotion Modelling with Human Belief Revision in Computer Games

BY Yan Ma

Abstract

Emotion modelling is receiving more and more attention from various fields, e.g. cognitive science, psychology, computer science and neuroscience. Most of these fields share the common research consensus that emotion can be beneficial to human's mental activities. This thesis is also grounded on the same consensus and makes further validations based on the following two hypotheses: One is emotional agents in games should to behave more like human beings than emotionless agents; the other is that agents having full emotional architecture should obtain better playing performance than agents with only partial architecture. Based on theoretical support, the author further hypothesizes that peoples' long term belief can be one of the sources to release complex emotions.

The experiment result suggests the emotional agents did perform significantly better than emotionless ones, but it was unable to significantly reflect the advantages from fully structured emotional agents over the ones of the partial architecture.

March 3, 2007

TABLE OF CONTENTS

List of Figures

Acknowledgements

I wish to express my warm thanks to my supervisor Dr. Joe MacInnes. Without his patience and pertinent advice at all times this thesis could not be finished with less detour and effort.

I would also like to thank Dr. Sageev Oore and Dr. Camilla Holmvall. They provided precious and valuable revision advices regarding the thesis.

I would also like to show my great appreciation to Stuart Crosby. He offered many creative and inspiring ideas regarding my thesis and the experiment.

Furthermore, I owe great amounts of gratitude to my wife Xin Jiang, who not only kept our home together during the research and writing of this thesis, but also assisted me in recruiting subjects for the experiment, and organizing research materials.

Finally, I am thankful to Saint Mary's University for the financial support they provided which allowed me such a wonderful research opportunity.

Introduction

Emotion modelling has received increasingly more attention from various research disciplines over the past decades. Many researchers have taken investigations and studies on emotions, either in their theoretical aspect or application aspect, and have yielded a lot of contributions within different fields. For example, in the field of neuroscience **Damasio (1994)** and **LeDoux (1989, 1996)** revealed a significant understanding of emotions to peoples' thinking process by analyzing the human brain section. Within cognitive science, **Minsky (1986), Sloman (2001)** and **Anderson et al. (2004)** proposed systematic mind architectures by taking account of emotions. In the field of computer science, **Picard (1997), Velaquez (1997), and Gadanho and Hallam (2001)** set up computational emotion models and in psychology **Frijda (1986), Isen (1993) and Loewenstein and Lerner (2003)** analyzed various emotional influences.

In this chapter, we will concentrate on the following two main questions or issues: Why do we need to do research on emotion? What is the current state of emotion theory research?

Why do we need to do research on emotion?

The most practical answer to the above question is to see if emotion theory can be beneficial to us in regard to solving some specific problems or improving our quality of life. There are three ways that emotion research can be considered and applied.

Firstly, emotion research may enhance human believability for game agents in

virtual reality research and possibly in the entertainment industry. In game industry nowadays, game characters equipped with emotion architectures seem more attractive to human players than regular agents, as the former ones think and react like human beings do. For example, they could perform more rational and coherent behaviours or produce non-monotonic strategies while the regular agents do not. Consequently, such emotional agents are not able to create the engaging game environment that human players prefer. For instance, Champandard (**Champandard 2003**) developed a few emotional robots in the PC game "Quake2" that were able to operate emotionally in reaction to the stimulus of their environments. This was displayed through their loss of shooting accuracy when they are frightened or by their loss of perception when they are afraid. They are also able to dance when they are happy and victorious after winning a battle. Other examples of these types of behaviour can be found in (**Velaquez 1998, Bozinovski 1999, Henninger et al. 2003, Marinier and Laird 2004**).

Some advanced topics surrounding emotional believability can be found in the virtual reality arena or the Human and Computer Interaction (HCI) field. Emotion theory, especially emotion modelling, is of great importance to the above two areas as it not only enables agents to behave like humans, but it is capable of making effective interactions with the human users (**Gratch and Marsella 2004a, Gratch and Marsella 2004b, Tangury et al. 2003, Silva et al. 2001**). One great achievement made in HCI regarding the adoption of emotion modelling is the Kismet robot which was developed by the Kismet laboratory at the Massachusetts Institute of Technology,

USA. This robot is able to express rich facial expressions as human people do. Additionally, it is able to perceive peoples' emotional states from their facial expressions or voices and make appropriate responses both verbally and/or physically.

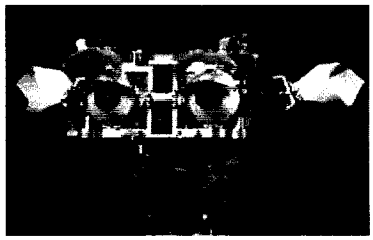

Figure 1.1: Kismet Robot, Image courtesy of P. Menzel

More complex topics in this area do not only include how to model emotions for agents, but also how to enable agents to use emotion theory to correctly identify human users' emotions through their inputs, facial expressions, speech, pitches or tones in order to provide better service (**Liao et al. 2005, Busso et al. 2004, Picard 1997, Fernandez 2003, Neumann and Narayanan 2004**). Such an application is widely used in the electrical tutoring system as well. Emotion theory can serve as a system to assess users' emotional states and insure appropriate interventions by applying different levels of knowledge to problem solving based on the subjects' emotional states thereby greatly improving the users' learning performance, (**Conati and Zhou 2002, D'Mello et al. 2005**).

Emotions also enable computer agents to build flexible and adaptive mechanisms capable of making quick responses to fast-changing environments as their experience increases. Such a characteristic is especially useful when dealing with problems in resource-limited environments such as time-constraint or energy-limited worlds, (**Wright 1996, Sloman 2001, Scheutz 2002**), as the results are quick yet often

sufficient solutions although such kind of solutions after the intervention of emotions are not always optimal. Still, many researchers support and claim that they have developed such emotional mechanisms (**McCauley and Franklin 1998, Sloman 2001, Botelho and Coelho 1998, Cañamero 2003**). Moreover, in the later part of this chapter I will elaborate on how Damasio applied his famous Somatic Markers Hypothesis, (**Damasio 1994**) to explain such a quick reflexive mechanism as emotion. In conjunction with this, McCauley and Franklin, (**McCauley and Franklin 1998),** once stated that "Emotions give us the ability to make an almost immediate assessment of situations. They allow us to determine whether a given state of the world is beneficial or detrimental without dependence on some external evaluation".

In another paper written by Botelho and Coelho, **(Botelho and Coelho 1998)**, the authors held a similar view as McCauley and Franklin and subscribed to the more affective appraisal instead of the traditional cognitive appraisal which helps agents evaluate the situations they face in a fast way. In their implementation, a long chain of cognitive rules can be condensed according to emotion-signals exhibited during the appraisal process. As a result of an affective appraisal based on emotions, the agents' thinking processes, which are normally completed by the cognitive appraisal requiring much longer time frames, is accelerated.

A similar description was made by Sloman, (**Sloman 2001, Sloman 2004**), however, Sloman thought such a quick response mechanism was only a kind of byproduct of the intelligence and should not to be thought of as intelligence itself. On the other hand, **Belavkin, (2001, 2003),** brought solid arguments that rejected

Sloman's assertion, in which he conducted a series of experiments to validate an important fact that the learning process could be accelerated if appropriate emotional stimuli were imposed on the testing objects, rats. That is to say that if we admit that learning is an intrinsic part of the intelligence, then emotion no doubt, impels the learning process to be more efficient and faster.

The third advantage of doing emotion research is to identify certain "negative emotions" that can be overcome in the future. This viewpoint is contributed by **(Scheutz 2004)**, who in his paper suggests that some kinds of emotion such as "guilt" or "infatuation" "can be construed as the loss of control of certain reflective processes that balance and monitor deliberative processes." Actually, Scheutz's suggestion had already been applied into Gratch and Marsella's research (**Gratch and Marsella 2003**), whereby they used their own emotion appraisal theory to detect such negative emotions, (e.g. over-confidence or fear), of soldiers in military planning and training, and how they coped with these emotions after they were identified.

In summation, emotion is of great importance as is its diversity in possible applications to fields such as entertainment, electrical tutoring, Artificial Intelligence (AI), and cognitive science. Based on the above statement, it is necessary for us to learn more about emotion itself, in order to build and tailor emotion modelling to fit our needs for specific purposes.

What is Emotion Theory about?

Emotion theory, unlike other research areas such as mathematics or classic physics

that already have an acknowledged theory framework and foundation, holds different views, opinions and theories from multiple research areas such as psychology, neuroscience, biology, philosophy and cognitive science etc., and does not have a consensus regarding many aspects within the subject field, including the definition of emotion, the role of emotion or some certain phenomena relating to it.

Although arguments have lasted for decades, the presence of Damasio's emotion theory, (**Damasio 1994**), seems to have gained the most recognition and support from the majority of emotion researchers. This is because his theory, from the position of neuroscience, has solid supporting experimental evidence. As a result, the rest of this section will first offer a short discussion on the definition of emotion, and then present an overview of Damasio's emotion theory which explains what emotion actually is.

A Rough Definition of Emotion and Related Discussion

Emotion is a familiar but strange term to everyone. It is familiar as it is experienced every day by us, however, it also seems strange that we can not see, smell, hear or touch it. Instead, we can only conjecture about it through some perceived clues. For instance, some perceptible physiological changes like tearing may indicate excitement or sadness, sweating may indicate anxiety or nervousness. We also emote by means of facial expressions, speech, behaviours and so on.

Since emotion theory has its wide applications as the previous section introduced, our need to research it requires a clear definition of emotion. A commonly seen

definition can be found from a website or a dictionary:

"An emotion, in psychology and common use, is an aspect of a human being's mental state, normally based in or tied to the person's internal (physical) and external (social) sensory feeling. (**Wikipedia 2007**)".

"(Emotion is) the affective aspect of consciousness; (or) a state of feeling; (or) a conscious mental reaction (as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body (**Merriam-Webster 2007**)".

The above definitions seem sufficient to cover the term "emotion" mentioned in our every day life. However, from the perspective of the research, especially for the sake of modelling, the definition seems in short of operation. For example, which causes the emotions, external or internal stimuli? What can the influence of emotion be? How can they affect our behaviour and thinking? Is emotion always the antithesis of the rational thinking? That is, does emotion only impose the negative influence on our deliberation process? Actually, many issues existed on the above aspects. For example in the first question above, some researchers insisted emotion can only be triggered from the external stimuli (**William James** in 1884 from (**LeDoux 1996**)), while others in converse claimed emotion could only be elicited from the internal of human minds (**Glasser 1999**). Some also argued that emotion can be triggered from both of the above sources (**Ekman 2004**). Even though, we can still choose an operational definition from varieties as the following shows:

"Utilizing an uncomplicated view, emotions can be viewed and studied as simple patterned behavioral and physiological responses to specific stimuli (**Gratch and Marsella 2003**)". Of course, like most other emotion researchers, Gratch and Marsella, did not base their full understanding of emotion on such a shallow level. Still, such a simple and plain explanation gives us an intuitive impression of emotion in that emotion refers to a study that focuses on building some "mapping relationships" between specific stimulus and the behaviours they give rise to.

Based on Damasio's findings, (**Damasio 1994**), the above statement seems a little superficial from the perspective of neuroscience. Simple definitions such as this one and others, (the first definition of emotion proposed by William James in 1884), (**LeDoux 1996**), primarily deal with the manifestations of emotion and the possible mappings between stimuli and their responses. They overlook how emotion evolves and manifests within a person's mind. Damasio's emotion theory dealt with this issue and he regarded this internal process as an indispensable component of intelligence. The rest of this section will examine briefly the overview of Damasio's emotion theory and will explain the emotion mechanisms to readers in a convincing way as a means to create a foundation for building my emotional model.

Review on Damasio's Emotion Theory

Damasio's emotion theory was built upon several important findings in the human brain, the most important being that emotions can be triggered in different cerebral pathways. The above proposition was originally inspired by Papez's brain

circuit theory of 1937; a theory about how the internal mechanisms of a creature's brain copes with stimuli, (**Ventura 2000**). In brief, Papez's findings uncovered the fact that external stimuli could be dealt with by the brain, (human or other creature), in two different pathways. The first pathway called "stream of feeling" forwards stimuli directly to the motion system which creates a body response. The other pathway called "stream of thought" redirects stimuli to several components of the brain to induce responses such as reasoning and deliberation before action. Two main realizations are noticeable from these findings. First, feelings are generated in both pathways and second, the processing time through pathway one is considerably shorter than through pathway two.

Papez's brain circuit theory was consolidated by LeDoux's research nearly sixty years later, (**LeDoux 1996**). LeDoux conducted a series of experiments to prove Papez's brain circuit theory by examining the emotion of fear in rats. First, he proved Papez's circuit theory by using the way of "parallel transmission, (see page 15 of **Ventura 2000**)", in that both ways of brain thinking went through amygdala, a brain area identified in LeDoux's time as generating emotions. Furthermore, LeDoux tested the rats' average time needed to direct acoustic stimuli to the amygdala in their brains. The data suggested that through pathway one, the time was twelve milliseconds but the time was almost doubled through pathway two. The second finding proved that pathway one, labelled by LeDoux as the "Thalamatic pathway", required no or less thinking and was much faster than pathway two, labelled by LeDoux as the "Cortical pathway". This said, the former pathway seemed to be less

rational than the latter.

Based on the above facts, it makes sense to see that Damasio further divided emotions into two categories as those two types of emotions follow the different pathways in human brains, i.e. primary emotions through Thalamatic pathway, (or "basic emotions"), and secondary emotions through Cortical pathway. The former referred to some transient psychological states such as joy, fear, anxiety and anger while the latter contrarily referred to durative psychology states such as guilt, infatuation and so on. This kind of taxonomy is widely accepted and used in many emotion researchers' papers because it is supported by experiments, and reveals the essence of the emotional process which happens to people and more broadly, creatures.

Damasio gave a reasonable illustration as to why he separated emotions. Such taxonomy was based on the experimental data collected from his brain lesion studies. These studies refer to the practical way in neuroscience in which the function of a certain brain area is examined by comparing the symptom induced by patients who have had damage to the brain area in question, with others who have had no such damage. He first investigated some old case studies in which patients who suffered from damage in prefrontal lobes, were not able to make as coherent a decision as a healthy person. For example, they were not able to allocate their time or energy toward a goal properly, they overly concentrated on sub-goals, they often lost their grasp on the overall situation, and so on. One fact which needs to be addressed is that the brain impaired patients had the equivalent IQ, (Intelligent Quality), level to

normal people.

Damasio proposed in his own explanation to the above case studies that all emotions could be aroused by two distinct brain areas: amygdala and prefrontal cortices, (part of prefrontal lobes). The amygdala is responsible for producing physiological reactions towards external stimuli such as facial expression changes like laughing, crying and so on, or motions such as jumping, stomping and any other bodily responses like sweating. Damasio thought the arousal of primary emotions was ascribed to the amygdala. Prefrontal cortices, working on the top of amygdala, were in charge of activating secondary emotions based on the memory episodes of past events, which in turn triggered the amygdala to generate some emotional decision that was aligned with the secondary emotion. This explanation actually indicated that the abnormal behaviours made by the above test patients were a result of damage in the prefrontal cortices that resulted in the lack of the ability to generate secondary emotions to perform long term planning or deliberation. Moreover, it is easy to observe that Damasio's explanation very closely parallels Papez's brain circuit theory; while the former explicitly ascribed two kinds of brain thinking processes to two kinds of emotions.

Since the brain working mechanism related to emotions is given, Damasio further illustrated how one's emotions greatly impacted one's decision making process. In order to further bring this illustration to light, we need to first know and understand Damasio's Somatic Markers Hypothesis. Then, evidence to support his hypothesis will be realized with only a brief explanation.

Damasio abstracted his famous Somatic Markers Hypothesis based on the above categorization on emotions. The hypothesis is quoted below:

"In short, somatic markers are a special instance of feelings generated from secondary emotions. Those emotions and feelings have been connected, by learning, to predicted future outcomes of certain scenarios. When a negative somatic marker is juxtaposed to a particular future outcome the combination functions as an alarm bell. When a positive somatic marker is juxtaposed instead, it becomes a beacon of incentive, (page 174 from (**Damasio 1994**))."

From the above statement, we can learn the author presumed such "markers" were aroused from peoples' secondary emotions. They help people to make choices by factoring in the remembered outcome marked by the markers. Negative or positive secondary emotions may lead people to make different choices in that negative emotions predicate a pessimistic future while positive ones an optimistic future.

Damasio further enumerated some features about somatic markers which were necessary for emotion modelling and so I have imported them with the following.

The first feature is that somatic markers only function as filters prior to the subsequent deliberation process, but they do not attend it in actuality. In other words, the markers only "highlight" those options sensitive to them and they will suggest that the subsequent deliberation process involve those emphasized options if those options are coherent to them; or, they will opt out of such options if those options are contradictive to them. No doubt such operations generally speed up people's decision

making process as they at least prune option nodes emotionally before deliberation.

The second feature is that somatic markers help people facilitate long term planning. Damasio emphasized the case that some option may produce immediate consequences which can lead to a positive outcome in the future. Working with understanding, positive somatic markers will help people to "endure sacrifices now in order to attain benefits later."

To demonstrate the somatic markers hypothesis, the following evidence is picked up from Damasio's book, (chapter 9 "Testing the Somatic-Marker Hypothesis" in **Damasio 1994**), with some explanation.

The most convincing experiment that illustrates the somatic markers hypothesis is Iowa Card Gambling Games, originally designed by Damasio's post doctoral fellow Antoine Bechara. In this experiment, a group of patients with prefrontal cortiex damage and a group of normal subjects were both invited to play a card game. The rule was simple: in front of each player, (either a patient or normal subject), there was placed four decks of cards labelled A, B, C and D. Each player was assigned equal amount of fake money, say $2000 at the game's start. Each player would have the opportunity to turn over a card from one of the four decks. Picking cards from A or B would earn $100 each time, but accidentally resulted in a fine that could be as high as $1250 if the card turned over directed the player to do so. On the other hand, picking cards from C or D would only earn $50 each time but only an occasional fine would have to be made and it was usually less than $100 on average. The goal of this game was to earn as much money from the decks as possible. Some restriction

applied to the card game in order to eliminate interference from working memory so that the participants were not allowed to keep notes of their monetary states nor were they informed of their state during the experimentation process.

The experiment results were quite remarkable. The normal subjects drew cards from four decks equally at the outset of the game and converged their choices on deck C and D gradually until the game ended since most of them thought attempts in A and B seemed more risky than in C and D. Patients with prefrontal cortices damage behaved similarly to normal subjects at the initial stage, but drew cards from A and B more frequently as the experiment progressed. One more additional fact was that sometimes patients avoided choosing A or B if the last card drawn from those two decks received a penalty payment. Yet, the patients would soon begin choosing from deck A or B soon after which was different from the normal subjects.

To exclude other possibilities such as patients being sensitive to reward over punishment, Damasio and his students conducted the card game experiment one more time by inverting the order of reward or punishment. The conclusion was that they saw similar results as in the first experiment in that the patients still persisted in choosing cards from A and B which have higher gain but also the possibility of higher loss.

From this conclusion it seems that due to the inability to use somatic markers to mark past events, the patients failed to form biased opinions on those four decks as normal subjects did. Instead, the patient's selection only depended on their momentary feeling induced by their basic emotions.

The above experiment, in addition to other evidence enumerated in Damasio's book, (see chapter 9 in (**Damasio 1994**), shows a potent arguments that emotion essentially provides rationality to the human decision making process although emotion can also result in irrationality, (page 193 in (**Damasio 1994**)).

In summary, through the above overview on Damasio's emotion theory we can clearly apprehend the essence of emotion in that emotion is an indispensable factor to the human thinking process, and acts as a presupposition of efficiency and rationality to people's deliberation and long term planning process.

Summary of the Introduction

In this chapter, two basic questions are addressed. One is why do we need to do research on emotion and the usefulness of emotion theory? The other is what on earth emotion theory is talking about? Regarding the first question, emotion can play several different roles and have several different applications in a variety of settings such as entertainment, electrical tutoring, military training, AI research and Health Sciences. As to question two, the emotion theory proposed by Damasio, having won the most recognition with its solid experimental support, has been given an understandable elucidation. Damasio's emotion theory outlined a clear understanding of emotion from the neuroscience perspective, In other words, emotions should be categorized into two kinds of emotions: basic emotions characterized by transient and quick responses and less deliberation, and secondary emotions with the features of slow responses and deliberation involving past events or

experiences. Finally, from Damasio's Somatic Markers Hypothesis with experimental support, we clearly know the latter response exerts a great impact on the human decision making process, and people may lose rationality towards making long term planning without it. This hypothesis may also be a solid argument to refute the traditional view that emotion should be strictly excluded from intelligence.

Literature Review

In the previous chapter an introduction to emotion theory was discussed with a strong focus on Damasio's findings in neuroscience. His emotion theory disclosed the important fact that emotions intrinsically brought positive contribution to humans' thinking processes, especially to decision making. Despite the specifics of Damasio's findings it is still necessary to more broadly consider the finding of others within the entire emotion research field. Furthermore, to fully apprehend a complete emotional process in one's mind and make further simulation possible, we at least need to have a basic knowledge of the process. For example, we need to know how emotions are elicited, how they are formed in people's minds, and how emotions influence people's decision making or action selection process.

The rest of this chapter is organized as follows: a brief overview will first be given on the development of the emotion research field over past years and then a technical overview on the basic knowledge of the emotional process will be explored.

Past Works on Emotion Research

Issue on the Usefulness of Emotions to Intelligence

As was mentioned earlier, emotion is not a new research topic and can be traced back to 1884 when William James first proposed the definition for emotion, (**LeDoux 1996**). Yet, before its wide applications were identified in science, emotion was traditionally thought to be a hindrance to any rational thinking, (**Young 1943, Hebb**

1949, Toda 1993, McCarthy 1995). For example, McCarthy, (**McCarthy 1995**), suggested that "robot[s] should not be equipped with human-like emotions", as he sustained robots equipped with human-like emotions were far from intelligent robots. Some of McCarthy's contemporaries disagreed with this view, (e.g. **Sloman and Croucher 1981, Minsky 1986**). Statements such as Minsky's that cited "The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without emotions", were considered by most emotion-usefulness supporters.

There was no doubt that Damasio's findings were significantly advantageous for the arguments of emotion-usefulness advocators. After all, those findings originated from the study on brain lesion patients and could be regarded as solid evidence that emotions affected people's thinking process positively, specifically in respect to decision making processes. They showed that without emotion support, people may not be able to make rational selections from among multiple options, or they lacked the ability to contemplate long term plans based on their past experiences. Not only Damasio, but other psychologists and neuroscientists came to similar conclusions that some positive emotions, such as happiness or joy, were able to help people deal with problems effectively, (**Isen 2004**), even if the plans were focused around survival, (**Cacioppo et al. 2004**). Recently, Lerner with her colleague, (**Lerner and Keltner, 2006**), further discovered from their designed experiments that even some negative emotion such as anger could also produce some positive outcomes as it was able to capture more attention and control to cope with situations. However, Sloman, once a

vanguard in emotion research along with Scheutz, still criticized the recent emotion research works, (**Sloman 2001, Sloman 2004, Scheutz 2002, Scheutz 2004**). They mostly thought emotion research misinterpreted Damasio's findings and that emotion should be considered a byproduct of human intelligence and not the requirement for intelligence. The presence of intelligence need not accompany emotions. Sloman deemed the damage on humans' brains lead to the loss of both intelligence and emotional elicitation and not the rise of intelligence. However, they did not design any experiment to support their arguments, nor did they set up any emotional experimentation to counter Damasio's findings.

Although there has been controversy throughout the development of emotion theory, it has seemed that more and more scholars indeed contributed many positive results from emotion theories for AI. For example, emotions can alter people's attention so that they turn their focus to a more relevant and current task (**Frijda 1986**). Emotions were used to perform multiple goal management (**Gadanho 2003**). Emotion could also be used by agents as a useful assessment tool to evaluate their environments or situations so that agents can make appropriate responses by using cognitive appraisal (**Elliot 1992, Gratch and Marcella 2004a, Gratch and Marcella 2004b),** affective appraisal (**Botelho and Coelho 1998**) or directly from Damasio's Somatic Markers Hypothesis (**Ventura 2000**).

Past Works on Emotion Modelling

In the first chapter I introduced some of the broadly known applications of emotion

theory that could be applied to various domains such as the entertainment industry, electric educational tutoring system design or health sciences. I also explained what emotion theory is from the perspective of neuroscience. Now, people may wonder how to connect these two concepts together in order to make the applications come into play. This is the main dilemma for computer scientists. Consequently, it is necessary for us to have a basic understanding of past emotion modelling works so we can tailor different emotion theories into practical agent models for use.

Considering emotions in computer science, particularly in AI, is not a novel idea since as early as in the 1960's, (e.g. **Simon 1967**), scholars proposed some agent architectures which integrate emotions as one component. The real boost of emotion research in AI began with the wide acknowledgement of Damasio's emotion theory; along with other positive findings about the role of emotion in relation to intelligence within the fields of neuroscience and psychology, (e.g. **Isen 1993, LeDoux 1996, Estrada et al. 1997, Isen 2004**). As a matter of fact, many scholars in this field were mostly motivated by Damasio's findings. Some of them totally adopted Damasio's emotion theory to develop their own framework,(**Ventura 2000**), others further built upon it, (**Sloman 1998, Sloman 2001, Wright 1997**), while others used Damasio's emotion theory to develop their own research works, (e.g. **Picard 1997, Cañamero 1997, McCauley and Franklin 1998, Hudlicka 2004**).

In order to summarize past works of emotion modelling, we can categorize them according to the research motivations behind them.

Building emotion architectures is important for exploring and understanding the

human mind from the perspectives of philosophy and or cognition science (**Sloman 1998, Sloman 2001, Wright 1997, Ventura 2000, Hudlicka 2004, Velaquez 1997, Velaquez 1998, McCauley and Franklin 1998**). Within this category, the Sloman's CogAff (the abbreviation of "Cognition and Affect") project was the most dominant. It proposed a clearly-dividing emotional agent structure which defined expandable and flexible interactions between layers, and also it offered well-adapted and reasonable extensions to Damasio's theory. In contrast, Sloman's theory was controversial and arguable since there was no experimental data or solid evidence to support his works.

Velaquez's works, (**Velaquez 1997, Velaquez 1998**), were also heavily referenced and discussed by many researchers as it was easily adoptable and nicely synthesized various emotion theories from psychology into one computable framework. The difficulty was that his work was somewhat mysterious in that he was not able to offer enough implementation details from the emotion theory he referred to and he lightly glossed over many complicated issues within cognitive science, such as issues of memory. For example, he described his mechanical dog design as able to make use of secondary emotions to retrieve memory. He claimed that if similar stimuli were received again, the past experience could be retrieved from memory and could influence the selection of current actions. The problem was that he did not give a convincing statement regarding how to carry out the above processes as he did not explain how memory issues on the human brain work such as how memory is stored and retrieved. He also did not offer any explanation as to how to sort out duplicate or

similar memories for capacity issues or how the retrieved memory episode influenced the selection of behaviours either by altering intensity of action values or by influencing current motivation states.

On the contrary, Ventura's emotion model, (**Ventura 2000**), which was greatly influenced by Damasio's emotion theory, presented three feasible emotion models which loyally reproduced Somatic Markers mechanisms within his double layer framework. Also, this work offered a well-defined mechanism to address some cognitive issues such as memory management mentioned above.

Endeavours in seeking any possibility to connect learning process with emotions were also a remarkable motivation for emotion modelling. As mentioned in the first chapter, such attempts were once made by various scholars. For example, the approach of reinforcement learning theory combined with hormone mechanism, (**Gadanho and Hallam 2001**), showed that the constant release of certain hormones related to emotions such as fear could reinforce learning in a certain situation. The affective appraisal theory (**Botelho and Coelho 1998**), a theory close to Damasio's somatic markers hypothesis claimed that attaching emotion signals to every situation experienced lead to learning the correct choice or positive response by means of retrieving some memory episode. This in turn resulted in creating a compilation of sequential historic rules which would generate an emotional signal thereby the compilation process would speed up the rules of retrieval the next time by identifying the matched emotion signals.

Belavkin, (**Belavkin 2003**), offered a new perspective regarding emotions as one

kind of stimulus, like noise, which could affect a creature's learning process. In his view, the emotional process during learning was termed the Simulated Annealing Process in AI. In other words, appropriate emotional stimulus in the initial stage of learning could lead to a diversity of attempts. In turn, cooling down or a reduction of emotional stimuli occurred when the performance was satisfactory, and warming up could again be performed by increasing emotional stimuli if the goal or environment changed.

How to make communication effective between humans and computers is another important research area in emotion modelling. Most research and work within this category involve a strong application background. Some things being considered in this work are "Why do we need to do research on emotion." Most application work stemmed directly from the emotion theories of psychology, cognitive science or neuroscience and even from AI itself. For instance, Picard, (**Picard 1997**), once adopted HMM (Hidden Markov Model) to imitate people's emotional state transition process. He also indirectly borrowed some architectures or learning ideas from the first and second research motivations explained above. In Gratch and Marcella's work, virtual reality or virtual army training, could be considered as one of the applications of cognitive appraisal introduced in the second category of motivation.

A Theoretical Review on Emotion Modelling

The previous section was mainly an exploration of the developments in emotional research over the past years. In this section, the focus will shift to how to model

emotions. That is, various opinions will be presented regarding sub-courses that can form a complete emotional process which normally includes the following procedures in sequential order: the categorization of emotions, the elicitation of emotions, different descriptions of the emotional process, and the emotional influences on action selection or decision making.

Discussion on the Categorization of Emotions

It should be said that not every emotion scholar made explicit categorization of emotions before conducting research due to different research focuses, but they at least implicitly grouped different emotions into two sets: positive effects or negative effects. This approach was called valence based emotion research. According to Lerner and Keltner, (**Lerner and Keltner 2000**), a valence based approach seemed unreasonable all the time because some emotions with the same valence may have different appraisal outcomes for the same given stimulus which could be either object or event. To prove this assertion, the authors conducted an experiment to compare fear and anger, both negative valence emotions, to appraise risk perception. The result of the experiment showed that fearful people tended to make pessimistic judgements while angry people held contrasting opinions, eventually feeling optimistic about the future.

In view of the above fact, it is necessary to discuss the categorization of emotions as this does not only remind scholars of the specificity that exists among different emotions, but also enables them to focus on a specific research area after

categorization. As mentioned in the last chapter, Damasio divided emotions into basic emotions and secondary emotions according to their different functional mechanisms in the human brain. So, what are these emotions that we experience in our daily life? Izard, (**Izard 1991**), listed eleven different kinds of basic emotions according to the Factor Analysis he proposed, while Ekman (**Ekman 1992**) produced a shorter version containing only six emotions considered primary emotions. These emotions are anger, fear, sadness, happiness, disgust and surprise. Ekman's division is most widely accepted by emotion researchers from different disciplines such as psychology and cognitive science. For instance, psychologists always seem to choose one or two primary emotions from their subjects in order to discover the useful properties pertaining to specific emotions, (**Isen 1993, Isen 2004, Cacioppo et al. 2004, Mellers 2004, Lerner and Keltner 2001, Lerner and Keltner 2006**). Isen specialized in the research of positive emotions such as happiness and joy and concluded that positive emotions, as opposed to negative emotions could result in faster and more creative decisions. Lerner et al. discovered that anger, a negative emotion, does not always produce a negative outcome. It can lead to a positive result since it is an attention grabber and strongly manifests in the mind.

Ekman's division of emotions was not a very unique way to classify emotions. A common way of classification is to regard all basic emotions as consisting of a group of dimensions, while each of them has different value distributions. Smith and Ellsworth (**Smith and Ellsworth 1985**) suggested that there are six aspects by which to measure one's emotional state. These are certainty, pleasantness, intentional

activity, control, anticipated effort and responsibility. A simpler version of the above division method can be found in Mehrabian's P.A.D theory, (**Mehrabian 1995**) whereby Mehrabian used only three dimensions to represent the diversity of emotions. In the P.A.D. theory, P stood for Pleasure-displeasure, A for Arousual-nonarousual and D for Dominance-submissiveness. The first dimension marked one emotional valence, (positive effect or negative one), the second dimension reflected the combination of physical and mental processes, and the last one was the property that controlled the intensity of the emotion, (i.e. if the emotion was able to influence people to a great or lesser extent). As a result, all emotional states that people normally experience could be succinctly represented by the above three dimensional vectors with different values scaled from -1 to 1; such as anger (-0.51, 0.59, 0.25), and elation (0.50, 0.42, 0.23). Furthermore, P.A.D actually categorized emotions into eight different groups with different combinations of (+/- P), (+/- A) and (+/- D).

In contrast, emotion research in computer science has not strictly followed either of the two division methods suggested above. Some scholars simplified emotions into valence based groups, (negative or positive), as computer scientists have a different research emphasis on emotion than psychologists or cognitive scientists. For example, in computer science emotion research may be focused on how to integrate emotions into AI, or how to make improvements regarding how AI can be better served by emotion theory. This only requires that scholars only examine the overall effects or characteristics of emotions. For example, as the last chapter showed, the applications of emotion theory for computer scientists need to involve a "fast and

reflexive" connection between the emotion and the designed agent model, (**McCauley and Franklin 1998, Botelho and Coelho 1998, Belavkin 2001**). On the other hand, there are some emotion researchers with computer science backgrounds that still adopt one of the above categorizations for emotion modelling, (**Velásquez 1997, Velásquez 1998, El-Nasr et al. 2000, Gadanho and Hallam 2001, Henninger et al. 2003, Tanguy 2003**).

Emotions and Elicitations

To model emotions in a specific domain, we need not only to decide on which categorization method we should adopt to extract a relevant subset of emotions, but also to consider how to elicit those chosen emotions in the specified environment.

At the very early stage of emotion research, emotions were thought to be the result of specific external stimuli, (**Watson 1929**), For instance, hunger or a threat would cause a feeling of fear, and loss of parent would induce a feeling of grief.

Later on, the above idea was challenged by various appraisal theories. Generally speaking, most emotion theorists thought emotions were elicited by a human's own appraisal of the stimuli they received.

The conventional opinion in appraisal theory is that emotions are mainly triggered by a human's own cognition and interpretation of external events and not the events per se. Such cognition processes that elicit emotions are called "cognition appraisals" and are widely used by many emotion researchers. Ortony et al., (**Ortony et al. 1988, Elliott 1992, Reilly 1996, Gratch and Marsella 2004b**). Frijda

(**Frijda 1986**), held a similar opinion but further elaborated that emotions were not elicited because of the happening of events, but because the events happened. For example, Frijda explained that "positive emotions can be said to result from events that represent a match: actual or signalled concern satisfaction. Negative emotions result from events that represent a mismatch: actual or signalled interference with concern satisfaction (page 278, **Frijda 1986**)."

Botelho and Coelho (1998) once claimed that they created a more advanced appraisal theory called "affective appraisal" than the cognitive appraisal theory (**Elliott 1992**), which was briefly mentioned previously from the perspective of learning (**Botelho and Coelho 1998**). They believed the affective appraisal contained a much more condensed and concise appraisal process than the cognitive appraisal.

Normally, a complete cognition appraisal consists of three steps to eliciting emotions, (**Elliott 1992**). The first step is to interpret a confronted situation. The second step is to compare the interpreted result from step one to the motives (goals or concerns for example). The last step is to elicit emotion(s) which are relevant to the comparison result from the previous step. On the contrary, Affective appraisal, according to **(Botelho and Coelho 1998)**, explicitly connected a given situation to a related emotion. Such direct mappings were accumulated due to the aforementioned compilation process. That said, affective appraisal may not be able to reflect the variety of responses that emotions can generate since the mechanism can only map a certain situation to a fixed emotional state as long as the rule remains invariant.

Cognition appraisal can produce different emotional results since it is capable of handling different interpretations to the same situation according to different internal motives structures. Such interpretations can be made through EECRs, (Emotion Eliciting Condition Relations), (**Elliott 1992**), or through concerns, (**Frijda 1986**), or through emotion structures which include goals, drives and motivations, (**Reilly 1996**).

Ekman, (**Ekman 2004**) put more emphasis on another form of emotion elicitation called automatic appraisals which had two features. One was to elicit emotions in a very fast way and unconscious way, and the other was to struggle for the most important need, such as those regarding "welfare or survival". He also provided us with an exhaustive analysis about the conditions which were able to elicit emotions, and generalized them into nine categories including automatic appraisals. I will mention some of them in the next paragraph.

External stimuli are not the only generators of emotions as internal stimuli can result in emotions as well. Ekman (**Ekman 2004**) generalized two ways that the human mind could elicit emotions. They are the recalling of a past emotional experience, and imagination. Most scholars seemed more interested in the recalling of experience due to its applicability. Intrinsically, recalling past emotional experiences refers to the process of self evaluation for future improvement. This is one more form of appraisal in conjunction with the other appraisal types introduced above. Some emotion scholars also stated that emotions elicited from the human mind, have an "anticipatory effect", (**Loewenstein and Lerner 2003**), such as

"introspection", (**Wright et al. 1996**), and "belief and sentiment", (**Frijda et al. 2000**). The above concepts and various views in about the emotion process will be discusses in the next section.

Review of the Emotional Process in Our Mind

How do the elicited emotions work in our mind before they produce some emotional signals that influence our behaviours or decisions? In this section, several typical emotion theories or models will be briefly discussed as they interpret the emotional process from different perspectives. Those models include Sloman's three layer mind architecture as a typical emotional process, (**Wright et al. 1996, Sloman 1998, Sloman 2001**); Loewenstein and Lerner's emotional decision-making theory (**Loewenstein and Lerner 2003**); Frijda's emotion theory (**Frijda 1986, Frijda et al. 2000, Frijda 2004**); and Elliot's affective reasoning (**Elliot 1992**). Many other emotion theories can also be found but the emotion theories listed above are sufficient enough to explain how emotions work in our brains.

Sloman's Three Layers Mind Structure

Sloman and his colleagues, who set up their "CogAff" project in 1991, were considered the pioneers in emotion modeling, (**Wright et al. 1996, Sloman 1998, Sloman 2001**). Their emotion model was first well implemented in Wright's PhD dissertation (**Wright 1997**) while Sloman later added new ideas to it, (see Wright's PhD dissertation, **Wright 1997,** for details). Sloman et al. viewed emotions inside the

human mind as the result of millions of years of evolution. Consequently, they thought human's minds were too complex to be fully understood unless correct layering architectures were applied.

Their proposed architecture consisted of three layers: the reactive layer, the deliberative layer and the meta-management layer. In their explanation, the first layer was shared by most creatures on the Earth. Its function is to store rules that empowered creatures to cope with the many situations they had met or inherited from their ancestors. Their evidence was based on the examination of some insects' habitual activities (e.g. fight or flight). The deliberative layer as they suggested operated on top of the reactive layer as the contemplating component such as conceiving plans before actions. Such a characteristic was able to save the storage space required by the reactive layer, and is abundant in most primates. For example, chimpanzees knew to move boxes to raise themselves in order to reach bananas hung on the roof of a house. The third layer, the meta-management layer, which is on top of the deliberative layer, has the ability to choose one strategy from multiple options conceived by the deliberation layer; similar to the deliberation layer choosing actions from the reaction layer. The third layer could also be used to explain more complex phenomena found in most of humankind such as some less perceived moods, (called "tertiary emotions"), like jealousy or infatuation. It is characterized as gaining or losing control, or attention affected by the two bottom layers.

Aside from the above "three layers conjecture" about the human mind, Sloman further gave rise to his "Information Processing Theory" as figure 2.1 shows. In

general terms, Sloman regarded an agent's mind as the centre for processing information from stimuli captured by its perception system and also delivering control signals to its action system after processing. This is the usual way in which a regular agent deals with information, (**Russell and Norvig 2003**). Furthermore, Sloman diversified the pathways of information flows between an agent's perception system and its mind system, and between its mind system and its action system. Additionally, Sloman divided both the perception system and the action system into three levels of sophistication to match the three layers in the mind architecture. Yet, he did not label those different sophistication levels as "Reactive", "Deliberative" and "Meta-management" as with the layers in the mind architecture, as those terms may not be accurate when applied to the above two systems. Instead, the different sophistication degrees were labelled as "low", "middle" and "high".

Emotion, according to Sloman's description, served as an alarm system which was able to direct stimuli to a certain layer of an agent's perception system. The perceived result was then sent to one or more appropriate layers in the mind to concurrently deal with it. Finally the produced control signal was sent to a layer in the action system. A simple explanation of this is as follows: The human mind is comprised of three systems: the perception system, mind architecture and the motion system. Each system is composed of three layers whereby the higher level is always more capable of dealing with received information than the lower level. Emotion is the state from which it is determined which layer in the above three systems will deal with received information. In other words, emotion can forward the stimuli to the

first layer of the perception system resulting in a simple perception or interpretation. Then the perceived result is directed to the second layer or the deliberative layer of the mind architecture to resulting in middle level thinking. Finally, the forwarded result form the mind architecture level is sent to the one of the three layers of the motion system resulting in the appropriate action.

Emotion, once receiving a stimulus, can determine which layer(s) of the mind will deal with the information and deliver the processed result to the appropriate action system depending on the urgency of the emotional state. For instance, lower urgency may send stimuli to a higher level of the perception or the mind system for processing if there is enough time for deliberation and selection. Sloman did not give a precise elucidation of how to run the alarm mechanism or how to use the alarm signal to form different information flows throughout the mind's architecture. He believed more cognitive processes needed exploring before consummating his conjecture on a person's mind since he claimed that people's minds were too complex to be fully understood. Regardless of such a deficiency, Sloman's three layers mind architecture provided a practical and self-contained theoretical framework for our further extension.

Although the control mechanism for information flow was unclear, Sloman still proposed some possible information flows. For example, he exemplified one flow that was shaped like the Greek letter Omega " ". To illustrate it, we can imagine the following scenarios: someone perceived a squirrel as a rat-like animal because they had never seen squirrels before but were familiar with rats, (middle degree

perception). In turn, the person was afraid of it because she considered that she was afraid off all rodents, (meta-management deliberation where abstraction exists). So, they decided to leave their environment, (middle degree action as a high level of plan was presented). This example outlines a rough impression of how Omega Information Flow works in people's minds. For the sake of clarification, two more examples are provided: First, if someone sees a rat, (low degree perception as no associative perception exists), and they felt disgusting, (reactive response), then they may begins to step backwards unconsciously (low degree action). This example illustrates a purely reactive information flow. The second example is if someone perceived a squirrel but had never seen one before, (low degree perception as no associative perception), and they felt disgusting because they thought the squirrels looked like rats and they had a disdain for rats, (deliberation as reasoning), then they began to step backwards unconsciously (low degree action), this would be an example of a lower Omega Information flow.

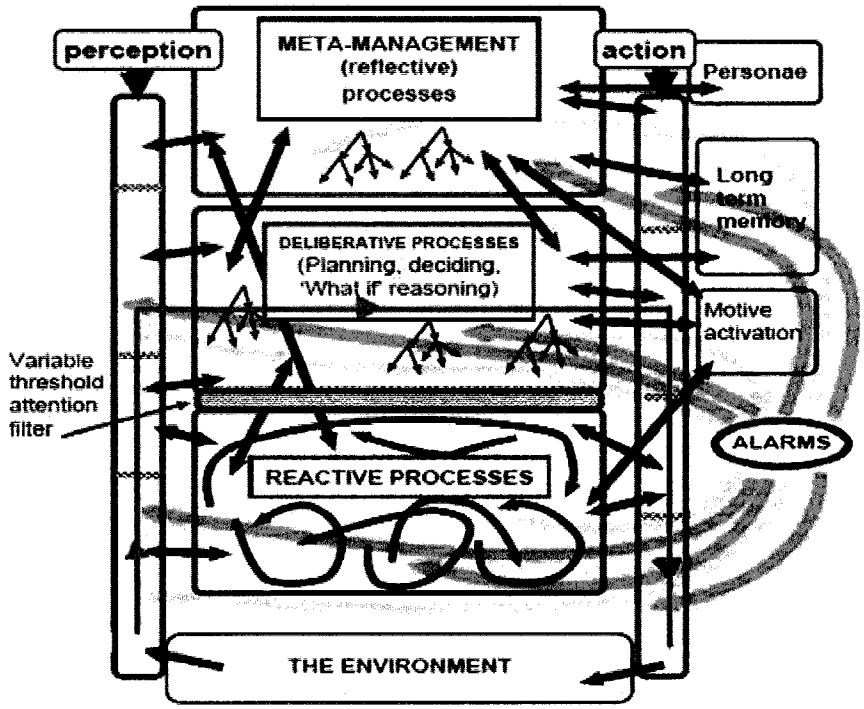

Figure 2.1 Three-Layer Architecture of MIND, by Courtesy of Aaron Sloman. The Bright Red Arrow was added by Yan Ma to reflect the Omega Information Flow proposed by Sloman. A bidirectional arrow denotes an interaction between two layers within a system or two systems. The spreading blue arrows from the alarm system denote the alarm signals can be transmitted to any layers within a system or two systems. The pink arrows pointing to the alarm system denote feedbacks from all other parts of the entire mind architecture to the alarm system. An arrow within one box indicates an isolated inner thinking process within a layer of the mind. The tree-like structure within the deliberative layer or the meta-management layer denotes a type of deliberation process. The bar between two layers within a system refers to filters which can block excessive interactions depending on the urgency level.

Furthermore, Sloman's design based on how people's minds work is basically consistent with Damasio's emotion theory. His reactive layer could be mapped to the amgydala which is in charge of controlling basic emotions to produce reflexive

behaviours; while the deliberative layer was mapped to prefrontal lobes of the brain, which are for secondary emotions that conceive long term plans or make deliberations. What is novel about Sloman's conjecture is that he believed that a higher layer, the meta-management layer, was required to better manage the deliberation process. The idea seemed to leave more free space for exploring the mechanisms regarding how long term processes affected what happened in the lower deliberative layer. In my opinion, (which was influenced by Frijda), (**Frijda et al. 2000**), the meta-management layer is the suitable place for a belief system or other long term concept such as mood or personality. In order to keep agents always conscious of their experiences and history, the layer is expected to mediate agents to form some consistent thinking process similar to humans.

Frijda's Emotion Theory

Frijda's emotion theory (**Frijda 1986, Frijda et al. 2000, Frijda 2004**) and his joint implementation with Swagerman (**Frijda and Swagerman 1987**) provided a new link between emotions and evolution. That is, he regarded the emotion process as able to self-evolve over time. From the perspective of cognition, Frijda refined emotions into concerns and motivations. Those finer components were the right elements to bring evolution to emotions. The latter was able to modify emotions and action readiness to a certain event or object, while the former is relevant to the formation of beliefs.

The disposition of concern, according to (**Frijda 1986**), was defined as "a

disposition to desire occurrence or non-occurrence of a given kind of situation". This referred to the emotional, perceptual and interpretational process. That is, when some event or stimulus occurred, if it was perceived or interpreted as impinging on some concerns one had, it could elicit certain emotions. In (**Frijda et al. 2000**), the authors further pointed out that each emotion had its own concerns towards an object. When the stimulus, which could be either an object or an event, manifested some characteristic relevant to some concerns, the belief for or in that characteristic would be created or strengthened by the emotion holding the corresponding concerns.

Beliefs could be formed or strengthend due to constant stimuli and the related emotions would be stabilized during the process and form "sentiments" (**Frijda et al .2000**). According to Frijda's description, emotion could be elicited by the appraisal result, whether "beneficial or harmful", that was attached to the presence of certain stimuli. Such an appraisal result was a prototype of a belief called the temporal belief. When the same stimuli were eventually realized by someone in a stable and meaningful sense, (i.e. it always resulted in the similar outcome by one's perception and interpretation), the emotional response to these stimuli would be stabilized and appeared similarly. Such dispositional emotional responses were called "sentiments", and the temporal belief would become a long term belief in the meantime.

Moreover, when a situation related to the belief was presented, the belief would manifest its strength by releasing the dispositional emotions, which were called "emotion anticipation" in (**Frijda et al. 2000**). The strength originated from the belief that could potentiate the current emotional state if the stimuli were consistent

with the belief, while being resistant to change even if the stimuli are presented with contradictive information.

In turn, motivation referred to the impulsion to satisfy the elicited emotions concern or goal, and therefore, it would suggest choosing the action or strategy pertinent to the concerned object or event. The relationship between motivation and action was evolutionary and the former could "potentiate relevant action dispositions", "either because a link between the two was wired in, by previous experience, or perhaps by some 'insight' into what a to-be-executed action can achieve. (**Frijda 2004**). In the above quotation, the "action dispositions" actually refer to what is called "action readiness", the tendency to execute a certain action. Similarly, the relation between emotion and its involved motivations is not fixed in that emotion may change its own motivation through the appraisal outcome from stimuli.

From the above statement, we can infer that Frijda's refinement of emotions indeed could bring self-improvement during the entire emotional process. In the formation of belief, it would be useful to enable agents to choose current responses consistent with its past emotional experience; and the dynamics in motivation embodies the adaptability of emotions.

Loewenstein and Lerner's Emotion Theory

Loewenstein and Lerner's emotion theory mainly deals with emotional decision making processes, (**Loewenstein and Lerner 2003**). The framework they proposed has a general meaning to our emotion research as it hypothesized a self-contained

theoretic framework to explain the emotional process running in people's minds.

The authors hypothesized two emotional effects coexisting in people's minds at any moment: the incidental influences and the anticipatory influences. The former is generated because some occasional stimuli have happened, while the latter refers to some effect produced from expected consequences and the emotions attached to the expected emotion. The expected emotion is the emotion which cannot be experienced at the decision making point but experienced before it, and possibly after it, when the expected consequence became true. For example, we may feel excited before we buy a lottery ticket as we've already anticipated the exciting consequence of winning the top prize. Since the incidental influence is often contradictive to the anticipatory influence, "the immediate emotions associated with thinking about the consequences of a decision will differ in intensity and quality from the emotion experienced when the consequence occurs. (**Loewenstein and Lerner 2003**)"

Some interesting findings can be discovered in the fact that the work of Loewenstein and Lerner, and the work of Frijda both mentioned the function of "expected emotions", (which is called "emotion anticipation" in Frijda's work). By applying Frijda's explanation on the power of belief, it is easy to explain why immediate emotions often diverge from the expected emotions since expected emotions can be one kind of dispositional emotion that originated from the belief: The power to persist in a certain opinion which may conflict with the incidental influence currently received.

Models Implementing the Reactive Component Principal

In comparison to the above models, which show in depth the emotional process in the human mind, some emotion models did not explore it. Instead, they were designed as reactive machines which were wired with large amounts of rules that mapped defined situations to certain actions, (**Ortony et al. 1988, Elliot 1992, Reilly 1996**). Part of the implementation works from (**Champandard 2003**) are thought to fall into this category. Those models were actually rich emotion-action representation platforms, on top of which implementations were allowed to build up for various purposes, such as testing psychology theories (**Elliot 1992**), electrical art and recreation (**Reilly 1996, Champandard 2003**).

The reason those models are thought to be reactive is that almost all the mappings between situations and emotions, and emotions and actions are deterministic. Furthermore, the consideration of emotions is irrelevant to adaptation or learning. For instance, (**Elliot 1992**) carried out twenty four emotion types defined in (**Ortony et al. 1988**) and one thousand four hundred emotion induced actions. Each time the observed situation will be assessed according to nine attributes each of which may have two or more different values, and each emotion has three innate attributes connected to actions, and those attributes have finer optional values as well, the span of which could produce diverse combinations each of which connects to some specific responding action. Such lexicon-looking-up style was presented with rich representations between actions and emotions because of complex hard wired relationship, not because of deliberation or other intelligent components suggested by

Sloman.

In conclusion, agents with reactive component only have no or unobvious emotional process in their minds.

Review on Decision Making or Action Selection Process

After the emotion is processed by a human's mind, it can make the final decision on what action or strategy could be chosen next.

Abundant literature regarding this issue can be found. However, in the computational sense, the usual way to make decisions or selections is that the option with the extreme value, either maximum or minimum, will be elected; that is also the well-known strategy in the "winner-takes-all" strategy. For example, in (**Valequez 1997**), the author calculated each primary emotion's intensity according to the following formula:

$$I_{et} = \chi\left(\psi(I_{et-1}) + \sum_{k} L_{ke} + \sum_{l} G_{le} \bullet I_{lt} - \sum_{m} H_{me} \bullet I_{mt} \right) \qquad (2.1)$$

In the above formula, I_{et} is the intensity of emotion e at the time t; χ is the saturation threshold; $\psi(I_{et-1})$ is the last intensity after decay through the function $\psi()$; $\sum_{k} L_{ke}$ is the sum of support received from all elicitors, and $\sum_{l} G_{le} \bullet I_{lt}$ is the sum of support received from the friend emotions, and $\sum_{m} H_{me} \bullet I_{mt}$ is the sum of objection received from the opposite emotions. Each time the emotion with the maximum emotion intensity will be elected to control the current emotion system. Similar approach could be found in (**Gadanho and Hallam 2001**).

Although the computation for this selection is monotonous, the criteria vary. The most commonly occurring variation is the cost-benefit assumption that assumes people always chose the "good enough" option that balances the expected gain and the cost. It is represented by Anderson's ACT-R theory (**Anderson 1991, Anderson et al. 2004**) where ACT-R stands for "Adaptive Control of Thought - Rational". Its mathematical form is as below:

$$U_i = P_i G - C_i \tag{2.2}$$

In the above formula, U_i is the utility of the ith option; P_i is the probability of achieving the goal by choosing option i; G is the expected gain of the current goal; and the C_i is the cost of executing option i. Certainly, the option with the maximum utility according to (2.1) will be elected for execution.

Belavkin (**Belavkin 2003**) contributed an innovative prospective to Anderson's work. Instead of making selections according to the maximum utility strategy, he proposed the idea of choosing the option with minimum ratio of expected effort to benefit under more constrained conditions, Poisson distribution.

$$\tilde{C}(x) = \frac{k(x)\overline{C}(x) + \xi(\overline{C}(x))}{k(x)+1} \tag{2.3}$$

In the above formula, x denotes an optional solution; $\tilde{C}(x)$ denotes the expected cost; $k(x)$ is the number of trial times, and $\overline{C}(x)$ is the past cost which is the ratio of past effort to past trial times plus 1 (see (2.4) below), $\xi(\overline{C}(x))$ is the randomly generated noise that is used to resolve conflicts in the initial stage (see (2.5) below), i.e. when the trial time is small. Here,

$$\overline{C}(x) = \frac{t(x)}{n(x)+1} \tag{2.4}$$

Where $t(x)$ denotes the effort (time) spent previously, and $n(x)$ is the number of successes with the effort.

$$\xi(\overline{C}(x)) = rand \in (0, 2\overline{C}(x)) \tag{2.5}$$

Belavkin (**Belavkin 2003**) further suggested the expected cost derived from (2.3) could be thought of as the optimal moment to give up one solution. In other words, when the effort that has been spent is greater than the expected cost, we may choose to give up the current solution and switch to another. It is done so because it has been proven that the optimal moment to register the first time of success is "when the probability of success equals the probability of failure", if the distribution of the solution obeyed the Poisson distribution (see page 103 of **Belavkin 2003** for details).

Thagard (**Thagard 2002**) once proposed a connectionist network HOTCO2 (Hot Coherence 2) to make emotional decisions. The main idea behind the theory was to choose most coherent hypothesis when emotional coherence was reached in the given environment. "Coherence" in the context of his connectionist network, meant there was no obvious change between the candidate hypothesis nodes and the evidence nodes within the network. To elaborate how HOTCO2 works, we must first understand his early coherence theory ECHO.

ECHO, as explained by Wang (**Wang 1998**), stands for "Explanatory Coherence by Harmony Optimization", and is a connectionist network composed of two groups of units: proposition units and evidence units. Each unit has its activation value to reflect its own potential for influence. For example, the greater the activation value

of a node is, the greater influence it is able to exert on its linked nodes. Such a kind of influence between two linked nodes could be positive or negative and it is determined by their linking weight. This means that the positive linking weight indicates the supporting relationship between two nodes, and negative linking weight indicates the opposing relationship.

ECHO can be used to make belief revision, as it is able to inference the most coherent hypothesis from the presented evidence after updating the activation values and linking weights of all the active nodes. This updating process will be briefly introduced two paragraphs down, and more detailed information about it can be found in the works of Thagard (**1989**) and Wang (**1998**). Wang (**1998**) made improvement on ECHO and derived his own concept which he called UECHO (Uncertainty-aware ECHO) that features a dynamically updating ECHO network by adding the consideration of sequential evidence and the ability to quickly converge to the most coherence proposition by updating the linking weights.

We cam start a quick review of how UECHO works. When setting up a UECHO network, a harmony value should be specified before running the network. It is the criterion to judge if the entire UECHO reaches the harmony. When we run the UECHO, if the final output value is less than the harmony value, we say the entire network is harmonized and the proposition with the maximum value can be elected as the output of the decision.

When one event is perceived, it is sent to the SEU (Special Evidence Unit), which is used to transfer perceived events to their representative evidence units. SEU

performs the sequential evidence updating. It resets the linking weight between evidence units that represent the perceived events and SEU itself to the original value (which is also the maximum value in UECHO, say 1.0, so as to indicate the perceived events present the latest evidence). Also, the SEU weakens its linking weights with other evidence units, some of which can be disabled if the values of the linking weights are below the specified lower bound such as 0.01. The linking weight updating formula is displayed as below:

$$LW(t+1) = LW(t) \bullet (1-d)^{\sqrt{t}} \tag{2.6}$$

Where LW_t is the linking weight at the loop t, d is the decay rate given the value of 0.9 under my implementation.

After updating the linking weights between SEU and Evidence Units (EUs), it starts updating the ones between EUs and Proposition Units (PUs), also called Hypothesis Units, in order to obtain the net input for each node. This process is identical to (2.6). The change in linking weight between a PU PU_i and its linked evidence ev_k is:

$$\Delta w(ev_k, PU_i) = \begin{cases} \alpha \bullet (Act_{\max} - Act_{PU_i}) \bullet Act_{ev_k} & (ev_k \text{ and } PU_i \text{ positively linked}) \\ -\alpha \bullet (Act_{PU_i} - Act_{\min}) \bullet Act_{ev_k} & (ev_k \text{ and } PU_i \text{ negatively linked}) \end{cases}$$

(2.7)

In this equation $Act_{\max}$ and $Act_{\min}$ denote the maximum and minimum value within the network, say 1 and -1 respectively, and α is a constant coefficient with a positive value less than 1, say 0.3. The new linking weight between PU_i and ev_k at the moment of t+1 is updated as:

$$LW_{(ev_k,PU_i)}(t+1)=\begin{cases} LW_{(ev_k,PU_i)}(t)+\Delta w\times(w_{\max}-LW_{(ev_k,PU_i)}(t)) & \Delta w\geq 0 \\ LW_{(ev_k,PU_i)}(t)+\Delta w\times(LW_{(ev_k,PU_i)}(t)-w_{\min}) & \Delta w<0 \end{cases}$$

(2.8)

In this equation $w_{\max}$ and $w_{\min}$ represent the upper and the lower thresholds, 1 and -1 respectively, and they are used to normalize the produced linking weight.

The net input for a node e_i, which can be either a PU or an EU, at the loop t+1 is the sum of all linked nodes activation values multiplying their linking weights:

$$net_{e_i}(t+1)=\sum_i LW_{(e_{ij})}(t+1)\bullet Act_j \tag{2.9}$$

According to (2.9), the activation for e_i can be updated as:

$$Act_{e_i}(t+1)=\begin{cases} Act_{e_i}(t)\bullet(1-\theta)+net_{e_i}(t+1)\bullet[Act_{\max}-Act_{e_i}(t)], (net_{e_i}(t+1)>0) \\ Act_{e_i}(t)\bullet(1-\theta)+net_{e_i}(t+1)\bullet[Act_{e_i}(t)-Act_{\min}], (net_{e_i}(t+1)\leq 0) \end{cases}$$

(2.10)

Here, $Act_{\max}$ and $Act_{\min}$ represent the upper and lower thresholds, 1 and -1 respectively, and they are used to normalize $Act_{e_i}(t+1)$; θ is the decay rate for the old activation $Act_{e_i}(t)$, say 0.05.

Given the net input values for all the nodes within the network, the sum of all the net inputs in the loop t+1 is:

$$H(t+1)=\sum_i net_{e_i}(t+1) \tag{2.11}$$

The absolute value of the difference of the net input at loop t+1 and t is:

$$D(t+1,t)=\left|H(t+1)-H(t)\right| \tag{2.12}$$

If $D(t+1,t)$ is greater than the specified harmony value, the procedure will repeat from (2.7) until the final difference is less than the harmony value. We can name it "Coherence Calculation" for the above UECHO's working mechanism in order to make reference to it later.

Thagard further added the factor "Valence" to form the HOTCO2. Valence referred to the subjective judgements to some propositions. That is to say, a proposition in his new emotion model will be influenced by both the cognition and the affect. As a result (2.9) is changed as follows:

$$net_{e_i}(t+1)=\sum_i LW_{(e_{ij})}(t+1)\bullet Act_j+\sum_i LW_{(e_{ij})}(t+1)\bullet Act_j\bullet V_i(t+1) \tag{2.13}$$

The valence of the node e_i is V_i and is updated the same as in (2.10).

Even though the addition of the factor valence could better simulate people's emotional thinking manner, Thagard himself realized the valence based approach was unable to simulate a variety of emotions (**Thagard 2003**). This is the same conclusion Lerner and Keltner arrived at that we mentioned earlier (**Lerner and Keltner 2000**).

Many other ideas were also used as criteria towards making decisions, such as maximizing the pleasure and minimizing the pain (**Tomkins 1984, McCauley and Franklin 1998, Mellers 2004**), satisfying the motivation in the maximum level (**Frijda 1986, Cañamero 1997, Cañamero 2003, Frijda 2004**), and non-linear probability weighting (see discussion in **Loewenstein G. and Lerner 2003,** p624 in

Handbook of Affective Science) that mainly argues people do not always consider the option with the most occurring probability, but sometimes consider some options which happen less frequently.

Summary of Literature Review

In this chapter an overview on emotion research was provided. Two main concerns were addressed: the first was a brief overview on the development of the emotion research in the past decades and the second was a review on the theories of how to implement a complete emotion model.

Regarding the development of emotion research, the section first discussed whether emotion theories could bring new meanings to intelligence, AI in particular, since arguments around this issue have been happening for decades. We then did a brief review of the achievements made thus far according to different research motivations.

The section detailing the theoretical overview introduced various scholars' interpretations to each key sub-course of a complete emotion process so we can obtain a broad enough overview for further modelling works.

Methodology

In this chapter, agent architecture with emotion support will be proposed first within the framework suggested by Sloman (**Sloman 2001**). The purpose of building such an emotional agent originates from two motivations: one is to test if emotional agents would be preferred by most of the game players over emotionless ones; the other is to test my hypothesis that the integration of beliefs as long term emotions could lead to more coherent behaviours for agents. We will then present the experiment design in order to test the above two hypotheses under my proposed agent architecture.

The design features a new way to interpret how emotions affect human decisions. The decision making process is not only affected by the current formed emotional state, but also determined by the action readiness which has the potential to evolve as the experience increases. This design is originally inspired by Damasio's "Somatic Markers Hypothesis" (**Damasio 1994,** introduced in the first chapter) and Frijda's "motivation" idea (**Frijda 2004,** see the literature review), but realized by using Belavkin's conflict resolution formulas (**Belavkin 2003,** see my literature review). Instead of adopting the valence based approach, like Lerner (**Lerner and Keltner 2000,** see my literature review), three primary emotions with their own characteristics were chosen to embody different emotion effects to the motion system. Another highlight of this research is considering belief as one more critical factor to emotion modelling, and as Frijda has suggested, this enables the agent to make decisions coherent to both the current emotional state and the past formed belief (**Frijda et al. 2000,** see literature review).

Expected Results

The research was based on the game Quake2, a first perspective shooting game under PC, which features furious battle scenarios and requires the subjects to make fast responses. The task of Quake2 within my thesis is simple: subjects only need to find and eliminate their opponent in one game map that synthesizes different landscapes. The fast response requirement and simple task setting are expected to reduce the degree of the human player's deliberation, which also eliminates the need of an elaborate deliberation process for robots. Indeed, such a simple game environment is more obvious for game players to recognize any potential transition between emotions through the robot's actions.

Grounded in the above settings, the research expects that the agent equipped with emotion components is able to make coherent responses to subjects' challenges. That is to say, its behaviours and strategies are expected to be coherent to both of the long-term and short-term stimuli it has received. For instance, if it is experiencing more loss than gain (long term stimuli) or if it is heavily injured in a fight (short term stimuli), it is expected to act conservatively with less consideration on attacking and more on dodging or retreating. In the reverse situation, it is anticipated to behave aggressively. Such simulation may break a new path on how to design agents behaving coherently according to their own experience, which would fit people's thinking manner while receiving little research interest from my investigation. Therefore, such emotional agents are also expected to have a much better performance than the regular emotionless agents.

Starting with Rule Based System

We start with a simple model to build an emotional agent. In the gaming industry, more and more concern is focused on how to make game characters think and behave more like humans. This is because hard wired robots are not able to persist in human players' enthusiasm beyond the point when humans get familiar with all the scenarios, potential rules, and/or the agents' behaviour styles in a game.

Traditionally, adopting a rule based system is one economic and fast way to build agent architectures for games (another option considered is to use the finite state machine), as it is able to provide quick retrieving ability that finds an existing solution for agents to perform and it is also easily extendible to add more knowledge for agents if needed. Normally, a rule based system works as follows:

1. Certain situation is encountered;

2. Start retrieving the relevant rule chunks for seeking matched conditions from the rule base;

3. a) If only one rule is found, perform the actions or solutions corresponding to the condition.

b) If more than one rule satisfies the current condition, only the one placed in the preceding position will be chosen and the actions it includes will be executed.

c) If no matched condition is found, no action will be triggered.

It is easy to tell that such a simple mechanism can only produce monotonous results. For example, we set a rule in Quake2 to deal with the situation that when the agent finds itself in low health condition, it should choose to escape. As a result, a

human may deduce reversely from the robot's escape that it must be easily defeated in its low health condition. Such an "honest" opponent, which always faultlessly exposes its intent to its opponent (the escape behaviour always informs the opponent that it is too weak to fight: "Well, I am now too weak and close to death as there are not many health points left, just come and kill me"), would make its opponent humans bored after a while. To view the problem from another perspective, we may realize that such simple mapping may not cover all the reasons which could cause the escape, that is, the robot may retreat because of its low confidence in fighting ability, or just as a trick to tempt humans to attack and try to kill the opponent. On the other hand, a low health condition does not mean always choosing to escape, maybe a better solution exists such that the agent may go for the health bonus offered, and fight against its opponent after refilling itself with health points (HP), or just fight desperately. Generally, a simple rule based system binds a condition tightly with an action or a solution. It is impossible to map multiple conditions to multiple actions unless all combinations of both two sets are specified (See figure 3.1 below).

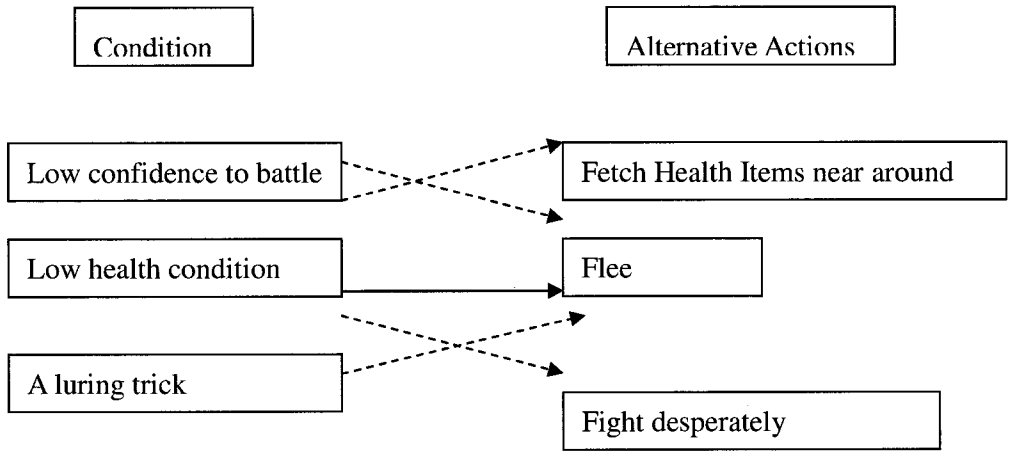

Figure 3.1: A Typical Case in Rule Based System. Solid arrow represents an existing rule, and the dashed arrows represent other possible rules that could be added.

Implementing emotion theory could be a good choice to add diversity to the rule selection of agents. As the first chapter introduced, people's decision making processes were intertwined with emotions, and the latter was important to the former. Consequently, incorporating emotional processes will greatly increase the flexibility during the rule selection process, and furthermore, it would also be possible to enable robots to think and behave like humans.

Still following the instance enumerated previously, by adopting the emotion theory, what could be specified is only a more common condition, like "enemy seen", and then all the optional actions above could be included under the same premise. The agent can then use its current emotional state as criteria to make its subsequent selection. The selection process is correlated to the agent's current emotional state instead of a clearly defined condition: if the agent feels fearful even in a satisfactory health condition, it will still choose to escape; on the contrary, if it becomes infuriated, it probably chooses to fight to death and does not care much about its current health

condition (these two assumptions are in accordance with the average person's manner of thinking). Therefore, running emotion theory to make selection to the above rule sets would proceed as follows:

1. Certain situation is encountered;
2. Obtain the current emotional state.
3. a) Retrieve the relevant chunk of rule sets, and find one to best match current emotional state.

   b) If no matched condition is found, no action will be triggered.

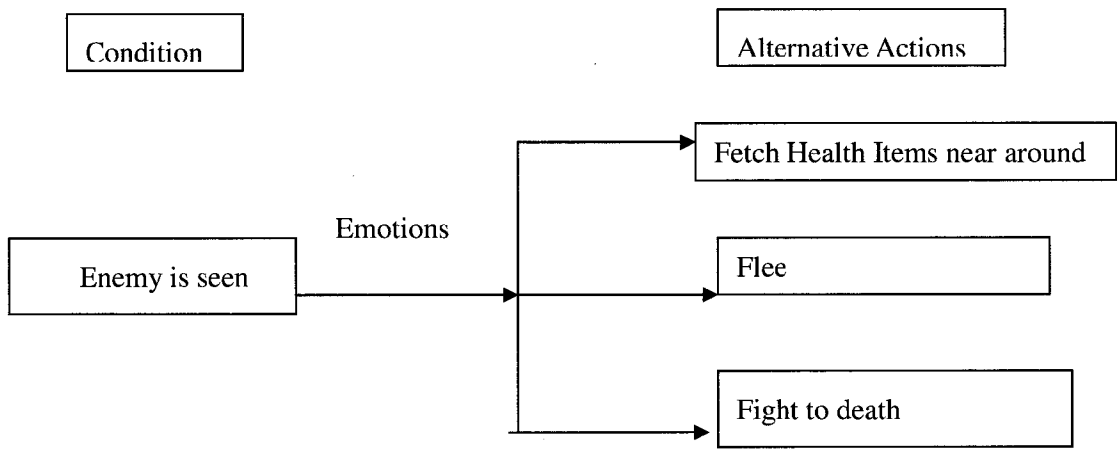

Figure 3.2: Using Emotions to Make Choice on Actions.

Compared to the regular rule based system, the above system with emotion consideration could save the space of storing rules. For example, the regular system needs to specify five rules to represent the scenario of "enemy seen", while the latter system only needs three.

Although the above example could be evidence that emotion theory could improve the selection ability for rule based systems, it is far from constructing an

emotional agent. The above selection process can only be thought of as the final decision signal coming out from the agent's mind, but what the agent's mind will be and how the decision is made are kept unknown. Therefore, in the next section, my implementation details based on Sloman's three layers will be elucidated.

Implementation under Sloman's Three Layers Mind Architecture

As mentioned in the literature review, Sloman's Three Layers Mind architecture could be regarded as an ideal framework for extension. It hypothesized the most necessary structures in an adult's mind, the three layers mind architecture, but left abundant space to allow more specific emotion theories to customize features in domain-dependant environments. In this section, I will illustrate how to build an emotional robot according to Sloman's three layers mind architecture but with several improvements.

Modifications to Game Quake2

Since Quake2 has its own characteristics as a game and I have my own research emphasis, it is not possible to implement Sloman's three layers mind architecture without any modifications. I will list all the differences between my design and Sloman's architecture followed by a brief explanation, and most of the differences will be explained in detail in the subsequent sections.

The following places are tailored from Sloman's theory to fit the game Quake2:

The first is that the Deliberation Layer was designed to only handle deciding. The other deliberation processes such as planning or reasoning will be ignored as the nature of Quake2 determines.

Within my research context Quake2 is a first perspective shooting game that requires fast responses in real time and less deliberation process. Such a process, interpreted by LeDoux's brain pathway theory, may mostly occur in the path involving basic emotions, and few through the path involving the secondary emotions. Such processes are characterises of many trivial and immense changes in the meantime. For example, we may not expect robots to pay much attention to human behaviours' performance sequence during fighting in order to conduct a learning process, as even eight basic behaviours[1] in Quake2 could produce millions of combinations (such as jumping first, aiming to the enemy, then finally firing; or left moving first, then ducking, and finally turning away). It is not necessary to guess where the human player will go to next from its current walking route, as the map is large and human players are not in grids. The above processes may appear either too random to happen, or too trivial to be "marked" by secondary emotions (as Somatic Marker Hypothesis suggested), therefore building HMM or finite state machines to do some estimation job for deliberation may consume much computing resource which is highly restricted during the game playing. Instead, to let an agent evaluate its fighting performance seems more meaningful, and I will explain later how to carry it out using Damasio's Somatic Markers Hypothesis in the deliberation process

[1] Eight basic behaviours in Quake2 include Step forward, Step backward, Move left, Move right, Jump, Duck, Turn around and Fire.

combined with my conjecture.

The second is that only Fear, Happiness and Anger were chosen from the six primary emotions proposed by Ekman (**Ekman 1992**) mentioned in the literature review. The reasons are identical to what Gadanho and Hallam (**2001**) explained that some primary emotions were probably not useful to be implemented in a certain experiment environment. Under the game Quake2, the emotion Disgust could not be used as there was no such situation for the agent to feel disgust unless we add some toxic food or other settings; the emotion Surprise could be felt by the agent when it accidentally met the human player or some other items, but there was no appropriate action in the reactive layer for it to perform for those scenarios; the emotion Sadness could also be felt but the actions for it could be highly overlapping to those for Fear. Therefore, choosing Fear, Happiness and Anger from the 6 emotions were believed enough to empower an agent to express its behaviours in different emotional ways.

On the other side, I do not support the popular point of view in computer science that oversimplifies emotions into two valence based groups according to their effects, positive effect or negative effect. A counterexample is Surprise, which may lead to either positive effects or negative effects, and more evidence can be found in the literature review.

The third is the information flow, based on Sloman's alarm mechanism, and will be simplified to hard wired events and behaviours for different upper layers; those of the Deliberation layer and Meta-Management layer, as he did not derive any specific mechanism to distinguish stimuli or outputs from "Mind" for different layers,

although he suggested setting attention filters in each layer and combining them with his alert system to form some information flow. The implementation process will be illustrated in the next section.

Reactive Layer

Under Sloman's agent architecture, it is easy to tell how the rule based system could serve as the reactive system of his designed three layer architecture. Since it is responsible for receiving emotional signals from the upper layer, and deliver it to trigger appropriate decisions or behaviours, how to implement it properly becomes the main issue. Damasio suggests (page 196-198, **Damasio 1994**) that only enough "factual knowledge" provided could drive the Somatic Markers to make effective judgements or selections. **Loewenstein and Lerner (2003)** further pointed out such knowledge should be complete and accurate. Plus, Damasio suggests the knowledge should also be categorized prior to use, as "...prior categorization allows us to discover rapidly whether a given option or outcome is likely to be advantageous, or how diverse contingencies can modify the degree of advantage". Within my implementation on the game Quake2, such knowledge was presented as symbols in a conventional way to guide robots to perform various tasks such as fight, pursuit, escape, seeking items, and wandering, to name a few. And those symbols were carefully layered depending on their abstractness degree and were also categorized under their concerned themes. For example, "flee" could be thought of as a more abstract concept to "jump"; and "Jump Fire", "Forward Fire", "Dodge Fire" could be

under the same theme "chunk.Fight". These themes were layered into three, each of which contained a few themes, and each theme included none, one, or more symbols which could be thought of as atomic behaviours; if the theme did not contain any symbols, itself would be thought as an atomic behaviour. Although it was a manual job to classify symbols and layered the themes, it explicitly carried out Sloman's proposition about layering behaviours or decisions to reflect different emotional influences as introduced in the last section. The following map illustrates my design:

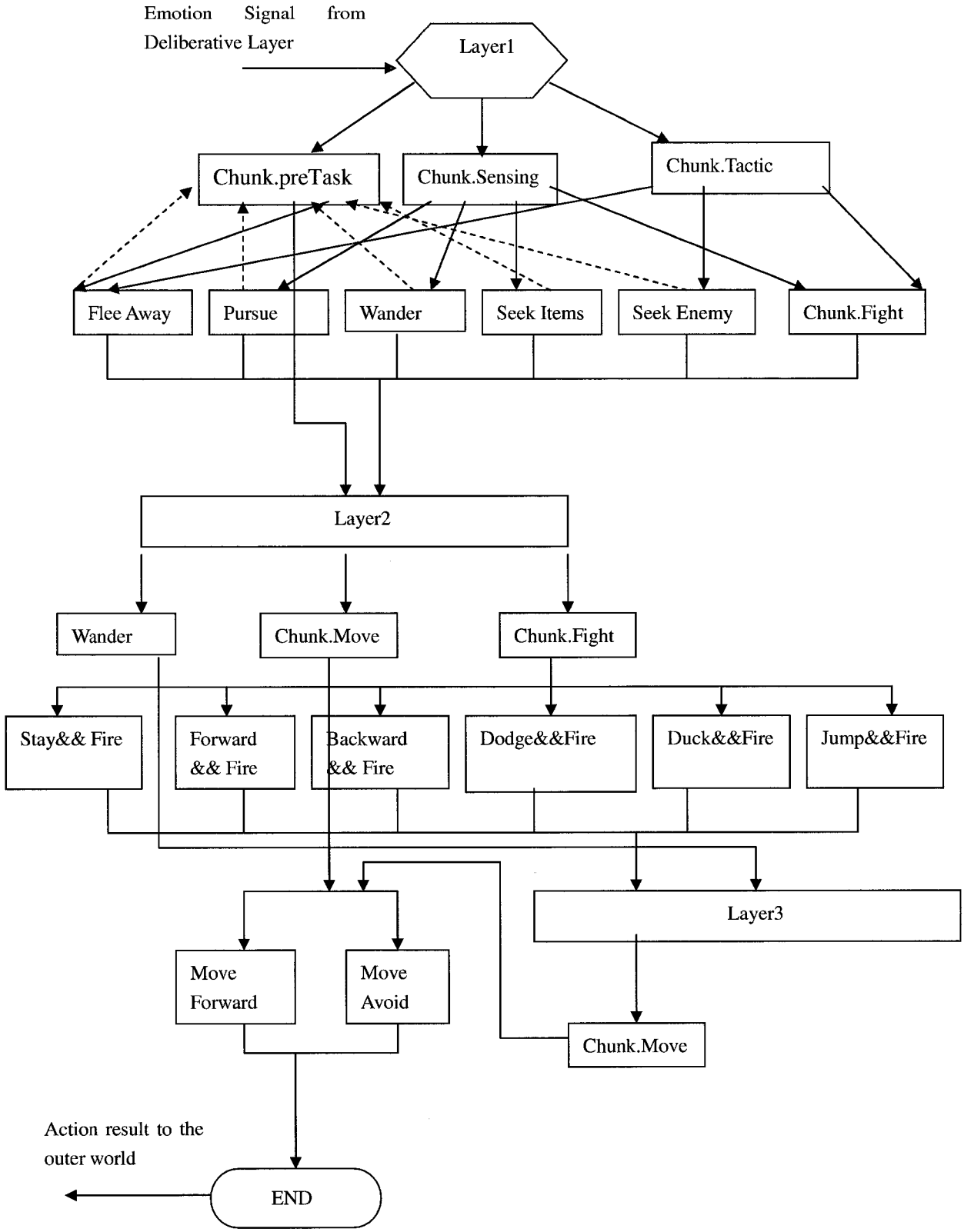

Figure 3.3: Inside the Rule Based System. A solid arrow denotes a possible route

derived from a symbol represented by a square when satisfying some condition; the dashed arrow denotes one continuous task that will be triggered from the source symbol since the next round in the rule based system, and the task will be represented by "Chunk.preTask".

The above figure presented provides us with a rough impression of how the rule based system works: it takes in the emotion signal as the criteria to make a choice or decision, and makes three different choices in three layers. The choice from the upper layer represents some abstract decision (such as escape or pursuit), while the lower layer contains a more concrete choice on how to execute the abstract idea selected from the upper layer coherently. To make the working mechanism clear, it is better to provide an example for illustration: initially, suppose the agent now is very angry and also under a low-health condition due to many previous battles lost, it starts its selection from the layer 1, which loads some general options: "Chunk.preTask", "Chunk.sensing", and "Chunk.tactic". The first option "Chunk.preTask" normally is used to perform some continuous tasks that cannot be done in a single loop, e.g. Escape ("flee") or Pursuit and so on. The second option "Chunk.sensing" suggests to the robot to perceive, and the last option "Chunk.tactic" enables the robot to start deliberation regardless of the current situation. Its anger cannot make it keep executing the current continuous task "seek some health item" to remedy itself, so it simply disrupts it, and it finds itself too upset to sense anything around it, which makes it impossible to choose the "Chunk.sensing". As a result, it chooses "Chunk.tactic" to make a further selection. In "Chunk.tactic", there are an additional three options for choosing: "Seek Enemy", "Flee" and "Chunk.Fight". Again, it

chooses "Seek Enemy" to match its current emotional state as it does not see the enemy yet, or it would choose "Chunk.Fight" otherwise. And it also registers the current strategy "Seek Enemy" for "Chunk.preTask" for the next loop. In the third layer, it chooses "Chunk.move" and makes further selection "Move Forward" to make its hunting process contiune. In the next loop, if it sees the enemy, or if its emotional state changes to less angry (maybe it accidentally obtains a health bonus to refill its health point), it will again disrupt its current continuous task "Seek Enemy" to "Wander" or anything else to best match its current emotional state.

From the above description, some advantages are worth discussing: the rule based system builds up a flexible mechanism for making decisions. As mentioned before, instead of making decisions completely depending on the current situation, the mechanism allows the agent to make choices according to its current emotional state. Furthermore, if one notices the words "very angry" or "less angry" used in the above paragraph, he or she may be aware of how to make those terms perceptible to the robot. As mentioned in the literature review, emotion researchers normally assigned the values to each emotion to denote its intensity, and the one with maximum value will obtain the power to represent the agent's current emotional state and be considered in the following decision making process (**Velásquez 1997, Gadanho and Hallam 2001** mentioned in my literature review). This winner-takes-all strategy seems less rational, for example, calling an emotional state "Fear" as it is composed of 50% fear, 49% anger and 1% happiness. The so-called "Fear" is far cry from the emotional state 90% fear, 9% anger and 1% happiness as the former one should have

much more violent intentions to choose more aggressive task than the latter. In my implementation, the selection process appears to be more "smooth" as it is able to distinguish the "very angry" and "less angry" states. Another main highlight of my adoption is to make decisions to best match the agent's current emotional state, rather than making selections to best match current dominant emotion or to seek maximum pleasure (see my literature review). Although those features are embodied in the reactive layer, the actual decision has already been made from its upper layer, the deliberative layer. I will illustrate how to implement the above advantages in the next main section, the deliberative layer.

Deliberative Layer

According to Sloman's description, (as was mentioned in the literature review), the deliberative layer holds many intelligent components such as reasoning, planning and deciding that sufficiently distinguish humans from most of the other creatures on the Earth. As explained in the "Modifications to Game Quake2", we may temporarily set aside those parts of intelligence other than decision within the context of my thesis. As mentioned earlier, the deliberative layer is capable of managing basic rules in the reactive layer and without its support, we or any other intelligent creatures may require more memory space to remember all possible combinations between conditions and actions. Such characteristics of the deliberative layer were actually demonstrated by the ways of emotional selections in the previous main section "Starting with a Rule Based System". How can we make such an emotional selection

process work rationally like humans do? This is something which is going to be explained in this next sub section.

According to my design, this layer consists of two major systems that lead to human-like decision making process: One is the emotion elicitation system and the other is the action readiness system. A brief illustration of the general working mechanism in the deliberative layer will first be presented, and then it will be followed by the illustrations on those two sub systems separately. Lastly, an issue will be discussed on why the synthesized emotional signals from the two sub systems would generate human like decisions or behaviours for the reactive layer.

Regular Working Mechanism in the Deliberative Layer

To give an intuitive impression on what the deliberative layer is and how it is related to the reactive layer, the structure for those two layers is presented as below:

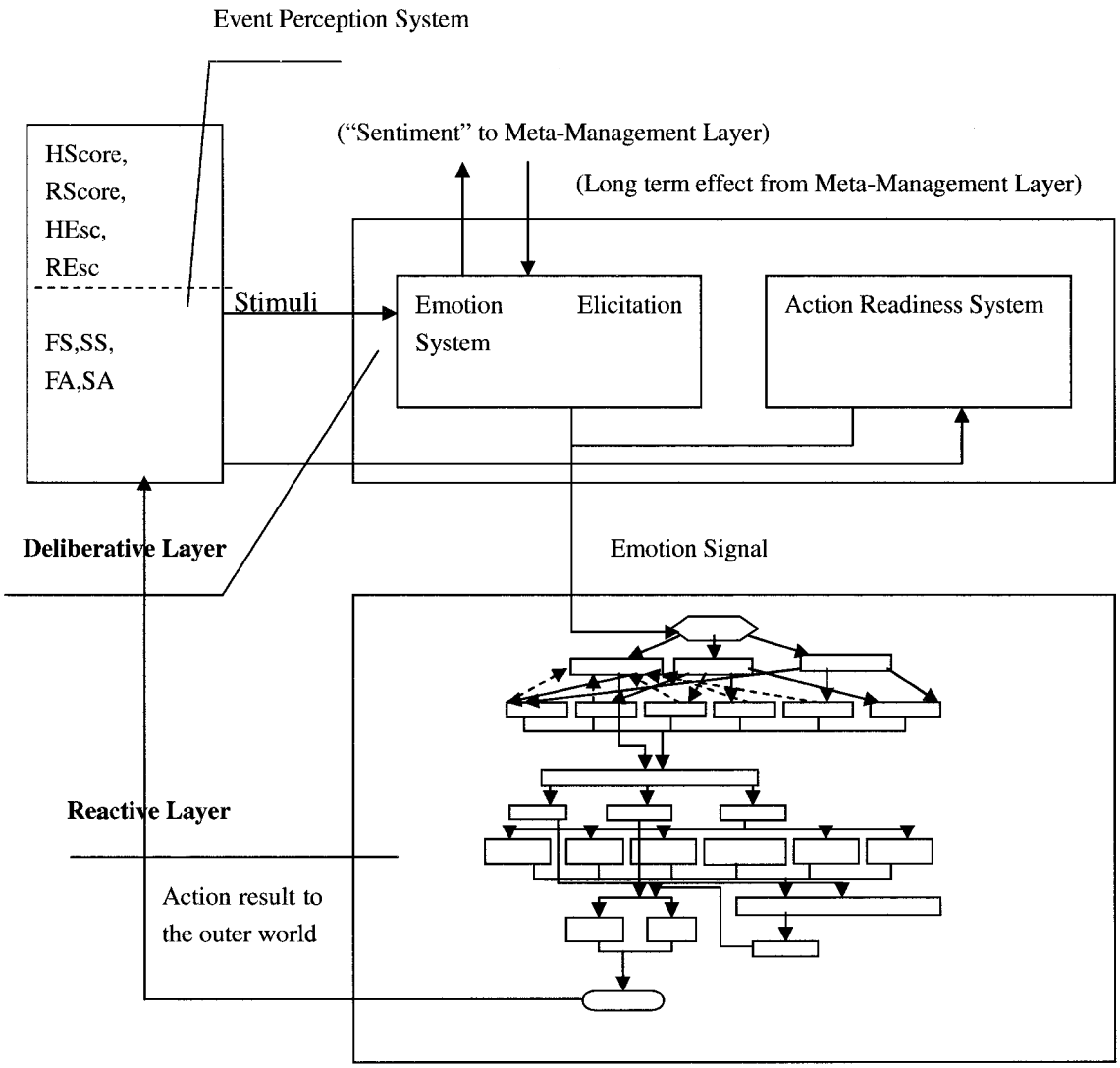

Figure 3.4: Reactive Layer and Deliberative Layer. The figure clearly marks the position of the deliberative layer in the entire agent architecture, its composition, and its interactions between its upper and lower layers. The dashed separating line within the box of event perception system denotes the division of the perceived events.

The above figure clearly marks the interactions between the reactive layer and the deliberative layer.

Before introducing the entire working procedure of the deliberative layer, it is necessary to mention how to categorize events perceived from the outer world that

could be the agent's perceptions. As mentioned in the literature review of Sloman's mind architecture, the perception should be divided into three levels according to the degree of sophistication. Here within the context of Quake2, the sophistication degree of perception was dependent upon the significance of the perceived event. That is to say, the events that result in great emotional changes were perceived as high degree perceptions; and the events with no emotional changes were automatically dealt with by the Quake2 game engine and thus lead to no perceptions within the agent architecture. These events belonging to none of the above two cases were perceived as low level perceptions.

Since the main goal of the game Quake2 is to earn the highest score possible, all the events related to the scores were set in my implementation to bring about the greatest emotional changes. As a result, they were perceived as high degree perceptions by the agent. Those events included the HScore (the Human won Score last time), the RScore (the Robot won Score last time), the HEsc (the Human made Escape last time), and the REsc (the Robot made Escape last time). Since they occurred at a low frequency, they were called "long term interval events". The perceived result was processed in both the deliberative layer and the meta-management layer, which also agreed with the findings mentioned in the literature review.

Some occurrences from the events are not related to the score but still lead to slight emotional changes. They therefore, were perceived as middle degree perceptions. Those events include SS (Successfully shot the opponent), FS (Failed

shooting the opponent), SA (Successfully avoided the opponent's bullet), and FA (Failed to avoid the opponent's bullets). The above events will be evaluated during the fighting scenarios. They are called "short term events" due to their relatively short duration and high frequency of occurrence.

The other events, such as perceiving the walls and landscapes, are not related to any emotions, and consequently do not induce any level of perception.

The rest of this section will give a brief introduction on the working procedure of the deliberative layer.

Events are first perceived from the outer world by the agent and then the perceived result is delivered to the deliberative layer or higher.

When the appraisal result enters the deliberative layer, it will be handled concurrently by both the Emotion Elicitation System and the Action Readiness System. The above two systems in the deliberative layer, along with signals from the Meta-Management or higher layer, will be synthesized to generate emotional signals that guide the reactive layer to perform the appropriate behaviours. If the event is a long term interval event, the deliberative layer will feedback its current emotional state along with long term interval events, to its upper meta-management layer to then be dealt with. Detail process will be illustrated in the next section titled the "Meta-Management Layer".

After executing a decision, the result will be acted out in the exterior world. This may induce another event in the external world, and if this event can be perceived by the agent next time, the procedure will repeat from the start of the working procedure.

Emotion Elicitation System

The Emotion Elicitation System takes in any event perceived outside the mind architecture and produces coherent emotional states. In the context of Quake2, the conditions of eliciting or updating emotions will be first explained, and then the design of the emotion elicitation system will be introduced, a design based on the connectionist network UECHO. Last of all, an improvement on UECHO, one that can better embody diverse emotional responses to the same stimulus, will be presented.

Two Types of Elicitations

In the literature review, four types of elicitations were discussed. To allow for generality, only two types of the four were taken into account for the game Quake2. One is the elicitation by cognition appraisal which follows Frijda's suggestion mentioned in the literature review, but with a slight revision. For instance, an emotion is elicited by a cognition appraisal outcome of external events, and the events are linked to only three different specific emotions as opposed to positive or negative emotions. These emotions are fear, happiness and anger. There are two types of scenarios we will explore in Quake 2. One is the battle scenario whereby an enemy is seen by the agent. The other is the normal scenario whereby no enemy is perceived. Regarding concerns mentioned in Frijda's narration, the three above emotions, become wired with one concern object, in each type of appraisal scenario. In the first scenario, four events "FS", "SS", "FA" and "SA" may be perceived and may elicit the corresponding emotions according to the concern object ΔHP, "delta of

health points". Also, if long term interval events such as HScore, RScore, HEsc or REsc, are perceived, they will elicit emotions in the second scenario. That is, emotions will be elicited in both the deliberative and the meta-management layers, which is in conjunction with Sloman's "Ω" conjecture mentioned previously, only with the different concern object $\Delta Score$, "delta of the score". How to calculate the emotion intensity according to the above two concern objects will soon be explained later in this section.

The other elicitation occurs when a past memory is recalled just as Ekman suggested as seen within the literature review. For example, one may recall a part or a complete scene from a past memory and thus elicit some emotion. Within the context of Quake2, such a process is implemented through the beliefs that reside in the meta-management layer. Each time events are dealt with by the deliberative and or meta-management layers, the deliberative layer will incorporate the anticipatory emotional signals produced by beliefs from the meta-management layer and thus produce the ultimate emotional signal. The next main section, titled "Meta-Management Layer", will elaborate on the process of generating anticipatory emotional signals.

Create Emotion Elicitation System as Connectionist Network

Building an emotion elicitation system by means of a connectionist network is not a new attempt, (**Velásquez 1997, Bozinovski 1999, Gadanho and Hallam 2001**). As explained in the previous section, some external stimuli will be hardwired to elicit or

update emotions. Furthermore, in the next section, the "Meta-Management Layer", it will become clear that long term interval effects from the upper layer can also exert their impact on the emotions. Yet, within the framework of this section, the focus is on how short term events influence emotions. To fit the settings in game Quake2, an emotion elicitation system is designed as followings:

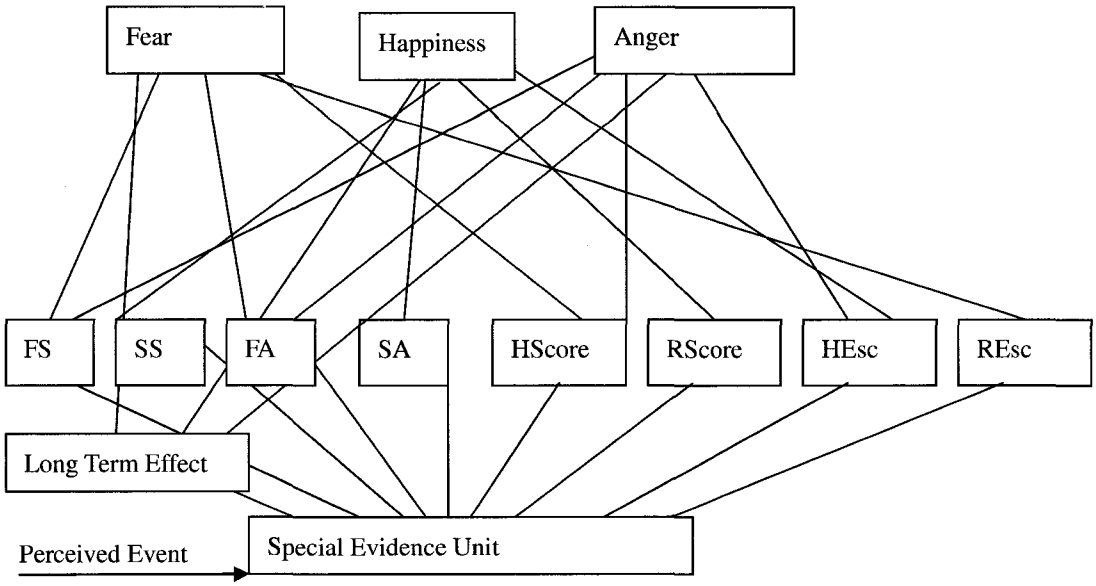

Figure 3.5: The Connectionist Network inside Emotion Elicitation System. The solid lines between each two boxes indicates positive effects between them, while the dash lines represents negative effects between two emotions.

The above figure exhibits how to connect primary emotions to their relevant events by means of the connectionist network inside the emotion elicitation system. The first row displays three primary emotions which are in the position of proposition nodes. The second row is lined with evidence nodes which are the events perceived externally. Those events are the elicitors of the emotions in the first row. For example, all positive events can trigger happiness, and all negative events can trigger fear, as do

most negative events plus HEsc (Human made escape before), while REsc triggers anger. Additionally, the dashed lines between two primary emotions indicate the opposing influences existing between them. This setting is a common conjecture in emotion research (**Minsky 1986, Velásquez 1997**).

In order to calculate emotion intensities, I have primarily adopted Velásquez's formula as was mentioned in the literature review.

$$I_{et} = \chi\left(\psi(I_{et-1}) + \sum_{k} L_{ke} + \sum_{l} G_{le} \bullet I_{lt} - \sum_{m} H_{me} \bullet I_{mt} \right) \qquad (3.1)$$

The explanation of the above formula can be referred to in the literature review.

Since the above formula has a highly overlapping form to the activation updating formulas 2.9 and 2.10 under UECHO (**Wang 1998**), and Velásquez did not increase implementation details in his own emotion models,(**Velásquez 1997**), I have placed (3.1) under the framework of UECHO to obtain more extensibility for my own research. Another important reason to use the above transplant is because the Meta-Management layer requires the use of beliefs which are present in the UECHO mode. I will elaborate on this later.

Create Event Intensity for Emotion Elicitation System

Although UECHO ameliorated upon ECHO in many places, it is not sufficient to be directly used as an Emotion Elicitation System without modification. The main deficiency of UECHO is that it is unable to embody the individuality of each of the primary emotions. This was also the problem that Thagard's HOTCO2 had. We find from the literature review that when two propositions are linked to the same evidence,

UECHO only updated the activation value for a proposition through its old weight. As a result, two propositions in the above case always attained the same weight value because they received the same effect from the same event at all times.

For example, if "FS" is discovered, then the link between "FS" and "Anger" and the one between "FS" and "Fear" will both be updated. Normally, the ECHO or UECHO system will update the proposition "Fear" and "Anger" equally over their linking weights as well as their activation values. The issue is that we cannot always expect that Fear and Anger will attain the same activation updating value since that is unable to reflect the diversity of different emotions to the same stimulus. Instead, it is better to suppose that the emotional state is sometimes composed more of "anger" than "fear", and other times, more of "fear" than "Anger". In order to carry out the above effect, the concept of "event intensity" has been created. This represents the impact exerted by the perceived event on the agent's current emotional state if the event relates to some concerned object that are pertinent to the emotion. As each emotion has its own feature set, they are expected to be influenced by the same event in different degrees.

As mentioned in the literature review, an emotion will be elicited in one of two ways, as mention earlier in this chapter, by the cognitive appraisal. The appraisal process is triggered when some concern is impinged on by the received stimuli. In the battle scenario, HP was wired to be the main object to be concerned with for all three of the emotions. To make such a concerned object computable by the event intensity, the HP was elaborated as ΔHP. That is,

$$\Delta HP = \frac{HP_{agent} - HP_{opp}}{HP_{\max}} \tag{3.2}$$

In (3.2), HP_{agent} and HP_{opp} represent health points of the agent and its opponent respectively, both in [0, 100]. $HP_{\max}$ = 100. ΔHP will consequently be in the range [-1,1].

Similarly, the concerned object, or "Scores between the agent and its human opponent", was wired for the three primary emotions in the non-battle scenario. Again, $\Delta Score$ was used to compute the concern caused by long term interval events:

$$\Delta Score = \begin{cases} \dfrac{Score_{agent} - Score_{opp}}{Score_{\max}}, & \left|Score_{agent} - Score_{opp}\right| <= Score_{\max} \\ 1, & Score_{agent} - Score_{opp} > Score_{\max} \\ -1, & Score_{agent} - Score_{opp} < -Score_{\max} \end{cases} \tag{3.3}$$

In (3.3), $Score_{agent}$ and $Score_{opp}$ represent the score the agent made and then that of the opponent respectively. $Score_{\max}$ is 10. $\Delta Score$ is bounded in [-1,1].

Indeed, (3.2) and (3.3) express people's different concerns within their environments. In battle scenarios, people are mostly concerned with the condition of their health. So if "FA" occurs and people are in good health, they may not care much about the health loss from FA and in this case, the concern value caused by FA will be low, thus leading to a minor change in people's emotional states. In contrast, someone with a FA occurring in low health may begin to panic. This type of consideration is regarded as the "alarm mechanisms" or "Urgency" within the mainstream of emotion modelling works, (**Sloman 2001**, **Frijda 1986**, **Damasio 1994**,

Cañamero 1997). Similarly, long term interval events may induce a similar change in people's concern when in non-battle scenarios.

After assigning different concerns to different groups of events, we use the following formula to calculate the event intensity for a primary emotion e_i in a clear way:

$$I_{evt}(e_i) = sign \bullet \frac{e^{k \bullet V_c} - \min(e^k, e^{-k})}{\left| e^k - e^{-k} \right|} \tag{3.4}$$

In (3.4), $I_{evt}(e_i)$ denotes the intensity of an occurred event and the influence it has on a certain emotion e_i; its value consistently falls within the range of [0, 1] which is determined by the right side of the equation;. *sign* indicates the valence of the event to a primary emotion. That is, some event may have a positive or a negative influence on a primary emotion, FA is negative to happiness for example. k is a scaling parameter within the range of [-1, 1], and the less the absolute value of k is, the more drastic the change I_{evt} can make. V_c is the value of the current concern and can be either ΔHP or $\Delta Score$.

Within the emotion elicitation system, one presumption was made that only one of the three primary emotions was most positively influenced by a certain event in (3.4), while the other two primary emotions were less sensitive to the event. Such an assumption enabled one emotion in a certain event to hold the higher potential to change activities to better satisfy the emotion. Such mechanism allowed competition opportunity for other emotions. For instance, fear was set to be the most sensitive and positive emotion to the event "FA". During a "Chunk Fight" battle, the emotion of

fear increased in an agent that had been shot more than once. The intensity of this emotion will increase much faster than the other two emotions and force the agent to retreat instead of continuing the fight. On the other hand, it is possible for the dominant emotion, say anger, to cause resistance to fear if its emotion intensity is high enough, or if it can obtain enough support from its positively linked event (FS linked to anger for example). Also, if the value of the concern (ΔHP or $\Delta Score$) is too large, which result in that the most felt emotion only receives a much smaller gain (see the formula 3.4 combined with the table 3.1), such a condition also make it possible for the dominant emotion to maintain its current leading position. The following table listed the scaling parameters used in Quake2 for each primary emotion under different events; they were chosen to reasonably match the above description:

	FS	SS	FA	SA
Fear	+, -1.0	−, -3.0	$+,\begin{cases}-0.1(\Delta HP>0)\\0.1\ (\Delta HP\le 0)\end{cases}$	−,-3.0
Happiness	−, -3.0	$+,\begin{cases}-0.1(\Delta HP>0)\\0.1\ (\Delta HP\le 0)\end{cases}$	−, -3.0	$+,\begin{cases}-0.4(\Delta HP>0)\\0.4\ (\Delta HP\le 0)\end{cases}$
Anger	$+,\begin{cases}-0.2(\Delta HP>0)\\0.2\ (\Delta HP\le 0)\end{cases}$	−, -3.0	+, -0.5	−,-3.0

Table 3.1 Scaling Parameter Table for Event Intensity under Short Term Effect

In the above table, the one with two conditional values is the mainly positively

affected emotion.

Similar to the table 3.1, the following table displays the scaling parameter setting for events related to the object"$\Delta Score$".

	RScore	HScore	REsc	HEsc
Fear	—, -0.5	—, -0.5	$+,\begin{cases}-0.1(\Delta Score>0)\\0.1\ (\Delta Score\le 0)\end{cases}$	—,-3.0
Happiness	$+,\begin{cases}-0.1(\Delta Score>0)\\0.1\ (\Delta Score\le 0)\end{cases}$	—, -0.5	—, -1.0	$+,\begin{cases}-0.1(\Delta Score>0)\\0.1\ (\Delta Score\le 0)\end{cases}$
Anger	—, -1.0	$+,\begin{cases}-0.1(\Delta Score>0)\\0.1\ (\Delta Score\le 0)\end{cases}$	+, -0.5	—,-3.0

Table 3.2 Scaling Parameter Table for Event Intensity under Long Term Effect

After obtaining the event intensity, we can update an emotional state by updating the linking weight of each primary emotion in the connectionist network. Then we can re-calculate the activation value of the emotion (the new emotion value) according to their updated weight.

The weight updating formula is changed from

$$\Delta w(ev_k, PU_i)=\begin{cases}\alpha\bullet(Act_{\max}+I_{evt_k}(PU_i)\bullet u-Act_{PU_i})\bullet Act_{ev_k} & (ev_k \text{ and } PU_i \text{ positively linked})\\ -\alpha\bullet(Act_{PU_i}+I_{evt_k}(PU_i)\bullet u-Act_{\min})\bullet Act_{ev_k} & (ev_k \text{ and } PU_i \text{ negatively linked})\end{cases} \quad (3.5)$$

The above formula, $\Delta w(ev_k, PU_i)$ denotes the linking weight update between Aemotion e_i and its supporter, event evt_k. The right side of the formula is mainly based on Wang's weight updating formula (2.7), but with one revision. The component event intensity $I_{evt_k}(e_i) \bullet \mu$ was added in order to solve the deficiency mentioned in the beginning of this section; μ is the maximum effect of event intensity and is given the value of 0.2.

If we replace (2.7) with (3.5) in the last section, and follow the coherence calculation process introduced in the literature review, we can obtain the entire working mechanism inside the Emotion Elicitation System. Such a working mechanism is called the "revised coherence calculation" and will be needed for later reference.

Action Readiness System

If we only use the current emotional state obtained from the emotion elicitation system to affect the agent's action, it seems in vain. That is, we still do not know why one emotion will trigger a certain action, nor do we know what the mapping relationship is between emotion and action. The Action Readiness System is applied as a crucial bridge that connects emotions to actions and results in rational "emotional actions". This is how the mapping relationship can be understood, (the second question above. The adaptability of this system, fits with Damasio's Somatic Markers Hypothesis, and thus answers the first question.

Connect Primary Emotions to Symbols in Rule Based System

As mentioned in the literature review, (**Lerner and Keltner 2000**) once suggested that each emotion (at least among basic emotions) has its own distinct judgement or effect to the same event or object. They used the emotion of anger to exemplify this fact, (**Lerner and Keltner 2001**) and their theory is what inspired me to use Fear, Happiness and Anger as the three primary emotions in my research.

Furthermore, in order to offer a general yet accurate description of these emotions, it is necessary to explore evidence from within research works from the field of Psychology. Consequently, by synthesizing the opinions of a few psychologists within the content of this thesis, (**Lerner and Keltner 2000, Lerner and Tiedens 2006, Mellers 2004, Isen 1993, Isen 2004**), I have come to the conclusion that Fear is emoted as a pessimistic attitude in combat and results in risk aversion strategies or behaviours. On the contrary, the emotion of anger results in an optimistic attitude about the future and produces attacking strategies by dominating over other emotions such as the need to escape. Happiness on the other hand always produces rational and optimistic decisions that are always positive. For example, it may not produce the same sort of results for escape as fear but it may instigate an attack if it is constantly influenced by positive events.

By following Frijda's emotion theory mentioned in the literature review, we can further extract three motivations respectively for the primary emotions. Therefore, Risk aversion can be the motivation for Fear while escape is the end result. Self improvement is the result of happiness, and assault is the end result of anger. In this

context, self improvement implies anything which results in improved conditions for the agent such as score gaining, health point refilling or armour equipping.

We can assume that one emotion does not only respond to its own motivation, but it also responds to the other motivations of the other two emotions. This enables competition between emotions during the decision making process, and is consistent with the connectionist network in the last section. The following table reflects this assumption:

	Risk Aversion	Self Improvement	Assault
Fear	++	+	−
Happiness	−	+	+
Anger	−	−	++

Table 3.3: Motivations to Emotions. "+" refers to the supportive attitude one emotion has to a motivation, "−" to averse attitude.

Since motivation, as mentioned before, acts as the bridge connecting emotion and symbols in a rule based system, it is required that it too can specify the relationship between motivations and symbols. If we consider the effect a certain symbol can produce, we can easily construct the following table to reflect the relationship between motivations and symbols:

	Risk Aversion	Self Improvement	Assault	Resulting Vector
Chunk.Tactic	+	NULL	+	0.25,0,0.25
Chunk.Sensing	NULL	+	NULL	0.25,0.25,-0.25
Chunk.Fight	−	−	+	-1,0.25,1
Flee	+	NULL	−	0.75,-0.5,-0.75
Pursue	−	NULL	+	-0.75,0.5,0.75
Wander	NULL	NULL	NULL	0,0,0
Seek Items	NULL	+	NULL	0.25,0.25,-0.25
Seek Enemy	−	NULL	+	-0.75,0.5,0.75

Table 3.4: Part of Symbols in Rule Based System to Motivations. "NULL" in the table refers to no effect between a symbol and a certain motivation

Notice the last row in the above table. One column among the vectors denotes a finally formed action readiness for a symbol, which is obtained by the product between the matrix of table 3.3 and the matrix formed by the first three rows of the table 3.4. We can express this mathematically as follows:

$$Vec_{m\times k} = [Emotion_m, Motivation_n]^T \times [Motivation_n, Symbol_k] \quad (3.6)$$

The above formula represents the action readiness of emotion "m" to symbol "k" bridged by motivation "n". During the above calculation, the "+" is assigned the

value of 1, "—" 1, "++" 2, "——" -2, "NULL" 0. All the vectors in table 3.4 have been normalized by the base 4.

One thing to be aware of is that not every symbol in the rule based system is assigned an action readiness value; in that not all actions need to be elicited by emotions (**Frijda 2004**). Two examples of this is "Wander" in table 3.4, or "Move Forward" which is not displayed in the table. The symbols under the theme "Chunk.Fight" were not given any action readiness value, either. The reason for this is that I will be demonstrating how action readiness can be adapted constantly under Damasio's Somatic Markers Hypothesis (**Damasio 1994**), in the next chapter.

Demonstration on the Adaptability of Action Readiness by Somatic Markers Hypothesis

In this last section, the basic knowledge of initial action readiness for every symbol within a rule based system is presented. Thereby, the agent can use the settings to produce emotional decisions. One main question that transpires is what if the agent has no prior emotional experience in relation to an action or theme? One possible answer lies in using the Somatic Markers Hypothesis, (**Damasio 1994**, see chapter 1 of my thesis for details) to make adaptive updating on the action readiness along the accumulation of the experience. Since the setting in table 3.3 represents the nature of those three primary emotions, we may keep it fixed; while the one in table 3.4 could be revisable as it represents the attitudes of emotions to symbols. Such attitudes could be constantly updated because the feedback from executing a certain symbol

changes or diversifies over time. As mentioned in the last section, I will begin my examination for such adaptability by looking at the Somatic Markers Hypothesis and the theme "Chunk Fight".

The process of updating the action readiness is actually the one to update the three emotions' attitudes towards one action. Since each emotion has different motivations in choosing actions, it can be assumed that one specific emotion is most inclined or predisposed to the action that most satisfies the motivation. This was also Frijda's viewpoint mentioned in the literature review. This was best seen in the theme "Chunk.Fight". We saw the performance of "Dodge" for risk aversion and "Hit" for assault, while self improvement was sensitive to any positive outcome produced by "Dodge" and "Hit" as they can trigger the emotion happiness.

Given the above motivation settings, we may further apply Belavkin's conflict resolution approach to calculate the action readiness for each option under the theme "Chunk.Fight". That is, the determined cost for each motivation is first calculated, and then its reciprocal form is accepted as the action readiness. The meaning behind the above procedure is interpretable as follows: the smaller the expected cost to satisfy one motivation is, the more solid the link between the motivation and the action will be. This correlates with the right maps in Frijda's description on motivation in the literature review whereby the motivation could "potentiate the action disposition (action readiness)" because of "previous experience".

For "dodge", the effort is the number of times the opponent is hit, while success is the number of accomplished dodges. Similarly, for "attack", the ratio is the

number of missed target shoots relative to the number of hits on the target. To enable the outcomes from "dodge" and "assault" in a comparable amount level, dodge is given one point of value for every fifteen successful dodges made. This is because to dodge is much simpler that to accomplish a successful assault. For happiness, things have been done differently as it is affected by both the positive events and outcomes, and costs of the above two categories. The action readiness value of happiness will be the midpoint or median in between the two other values for motivations. The following formulas show the attitude updating for three motivations respectively:

$$\widetilde{Attitude_{Dodge}}(x) = \frac{w \times k_{Dodge}(x) + 1}{k_{Dodge}(x)\overline{C_{Dodge}}(x) + \xi(\overline{C_{Dodge}}(x))} \quad (3.7a)$$

$$\widetilde{Attitude_{Assault}}(x) = \frac{k_{Assault}(x) + 1}{k_{Assault}(x)\overline{C_{Assault}}(x) + \xi(\overline{C_{Assault}}(x))} \quad (3.7b)$$

$$\widetilde{Attitude_{SI}}(x) = \frac{w \times k_{Dodge}(x) + k_{Assault}(x) + 1}{k_{Dodge}(x)\overline{C_{Dodge}}(x) + k_{Assault}(x)\overline{C_{Assault}}(x) + \xi(\overline{C_{Dodge}}(x) + \overline{C_{Assault}}(x))} \quad (3.7c)$$

In the above formulas, the "w" with the value of 15, is the regulation parameter to enable the action readiness value from "dodge" to be comparable with one from "attack" as mentioned previously. By synthesizing the matrix calculation (3.6) with table 3.3 which defines the attitudes of emotions to three motivations, and makes assumption that all the attitudes are originally set to be positive "+" (as all three attitudes from (3.7a) to (3.7c) produce positive values), we may derive three action readiness values for the three emotions, accordingly:

$$\widetilde{AR_{Fear}}(x) = 2\times\widetilde{Attitude_{Dodge}}(x) + \widetilde{Attitude_{SI}}(x) - \widetilde{Attitude_{Attack}}(x) \quad (3.8a)$$

$$\widetilde{AR_{Happiness}}(x) = -\widetilde{Attitude_{Dodge}}(x) + \widetilde{Attitude_{SI}}(x) + \widetilde{Attitude_{Attack}}(x) \quad (3.8b)$$

$$\widetilde{AR_{Anger}}(x) = -\widetilde{Attitude_{Dodge}}(x) - \widetilde{Attitude_{SI}}(x) + 2\times\widetilde{Attitude_{Attack}}(x) \quad (3.8c)$$

Among the above three formulas, AR represents "action readiness".

Again, by following Belavkin's conclusion mentioned in the literature review the regarding the optimum moment to give up the current trying solution, the above formulas could be used to determine when to give up. During the fighting process, it is certain that the probability of the potential first hit and first dodge from a hit will correspond to the Poisson distribution, therefore, I adopted his theory by using the reciprocal forms of (3.8a) through (3.8c) to guide the agent to redirects its actions when necessary. That is, if the number of hits an agent received from its opponent is beyond the expected times, or if the number of its failed attempts to shoot its opponent is more than the expected times, it may smartly know to switch from the current fighting actions to other options which may have better action readiness. Before switching, it will update the expected cost of the current action to a larger value which allows for greater failure tolerance at the expense of less opportunity to be chosen next time. The action readiness will be updated accordingly. On the contrary, if the attempts of one action always generate positive outcomes, (ie: the agent always hits the opponent and avoids attacks effectively), the expected cost to perform this action will be lower than before. This lower expected cost could most probably result in the agent failing to perform the action successfully. Nontheless, it is evident that after several attempts, the expected cost will inevitably reach a point of

balance in that the action readiness will reflect a stable performance for the agent to choose

Since the above action readiness mechanism has been set up, we can now start to see how an emotion chooses action through commensurate action readiness. As mentioned before, the conventional approach is the use of the "winner-takes-all" mechanism, so that the dominant emotion which has the greatest emotion value over the other emotions will choose the action according to its characteristics. It is argued in the last section that such a mechanism may ignore the other emotions and their effects, although they may be less noticeable at some times over others, but not always. To extend Frijda's idea (**Frijda 2004**) mentioned in the literature review, one certain emotion has a propensity to choose the action that most satisfies its motivation. We may think the agent will choose the action which best fits the current emotional state, instead of a single emotion. Such an extension could be represented by the following formula:

$$\overline{Action}(x) = \arg\max \sum_{m=\{Fear, Happiness, Anger\}} \overline{AR_{emotion_m}}(x) \times \overline{V_{emotion_m}}(x) \qquad (3.9)$$

The above formula (3.9) embodies the action selection process in a mathematical way. That is, an emotional state will choose one action which has the maximum value from the summation of the products between one emotion's intensity and its corresponding action readiness.

(3.9) explicitly points out that the action selection process is not determined by only one dominant emotion. Instead, it is codetermined by the current emotion intensity and also by the related action readiness. (3.9) argues a decision making

process does not only depend on the current emotional state, but also relies on the "impression" formed on each option. After many times of practice, certain opinions about certain objects or events will be gradually formed. For example, after thousands of attempts, the emotion "Fear" will finally discover that "forward&fire" is not suitable for it because it always leads to bad performance for "dodge", the motivation most valued. As a result, when the robot is in a "mainly fearful" state, it chooses other appropriate actions such as "dodge&fire" rather than "forward&fire". Furthermore, applying the above formula to action selection enables us to generate a more "smooth" effect, since it considers multiple emotional affects rather than one. Due to this fact, it can express some "fuzzy" and complex emotional actions. For example, if after a battle values have evolved as follows: "jump&fire" has the action readiness vector (0.80, 0.70, 0.10) and "forward&fire" equals (0.60, 0.55, 0.40); and the agent is mainly happy and only slightly angry at (0.00, 0.80, 0.20). This emotional order is "Fear, Happiness and Anger". By working out the above settings through (3.9) we can obtain 0.58 for "jump&fire" and 0.52 for "forward&fire", and it then becomes evident that the agent will choose "jump&fire". If however, the agent is still happy but becomes more angry, say (0.00, 0.55, 0.45), we can again obtain 0.43 for "jump&fire" and 0.4825 for "forward&fire"; Within this scenario, the agent will now choose"forward&&fire", in that the agent will prefer to attack than to dodge since anger results more in assault whereby the agent will exhibit more attacking intent. Consequently, we can discover that even under the same dominant emotional state such as happiness as cited in the above example, the agent may behave differently.

This illustrates well that (3.9) has the ability to represent various "fuzzy" emotional states in contrast to the emotion signal from Mind architecture, which is not capable of such complexities if two or more emotional states share a dominant emotion.

It is time to turn back to see if the above mechanism reflects features in Damasio's Somatic Markers Hypothesis (**Damasio 1994**). As I mentioned in the first chapter, the main feature of Somatic Markers Hypothesis is that those markers are able to highlight some options for us in the decision making process. Such ability is acquired due to the accumulation of experience from long term events. Still, Damasio only offered a vague description regarding how negative emotions can predict bad outcomes which act as an emotional "alarm", and how positive emotions inspire a sense of optimism which offers hope and the ability to move foreward. The implementation in this research defines Somatic Markers in a more specific way. First, we admit that diversity existed in various emotions in that each emotion has its own distinct feature set. Corresponding to the first point, Somatic Markers need to imbue different emotions with their different attitudes to various objects or events in a specific domain. For example, in the research area of fighting explored in this thesis, such a process has already been implemented by using motivations as a crucial bridging factor to connect emotions to objects or events within a domain. By synthesizing the above two premises, an action readiness updating mechanism is built up which mimics the "highlight" process of Somatic Markers Hypothesis. An illustration of this is one certain emotion will eventually choose a certain action through experience; as an opinion to this action is eventually formed, it is marked

according to the degree it satisfies the motivation. In other words, if the action always satisfies some motivation of the emotion well and successfully, then the emotion will score a high mark to this action in return. As a result, the emotion becomes inclined to choose that action when in the decision making process.

In order to make a clear idea of how to design the action readiness system, it is necessary to offer a generalized summary of the procedure as follows:

1. Build up the knowledge base specialized in the research domain (i.e. build up the "reactive layer"), as in this thesis the fighting knowledge was categorized under different themes in different abstract layers.
2. Seek evidence from psychological theory to form a description set for each emotion that could exert outstanding effect in the research domain. For instance, I applied fear, happiness and anger into fighting.
3. Specify the discriminating motivation for each emotion that could act as a bridge between emotions and various themes or symbols under them (Table 3.3 for instance). Also specify the relationship between motivations and symbols or themes (Table 3.4 for example), and work out action readiness for each symbol through (3.8).
4. Calculate the emotional affect as illustrated through (3.9).
5. If the agent does not have any emotional experience but abundant options, it still gets the chance to be adaptive to the theme consisting of those options and it could gradually update its own action readiness according to the process introduced in this section. This idea fits well with Damasio's

Somatic Markers Hypothesis.

Meta-Management Layer

In this section, the problem will be proposed first in order to induce the necessity of introducing meta-management layer; then the reason to add belief in the layer will be explained; finally, the design for this layer will be presented.

Problem Identified without Meta-Management Layer

Although the two layers we have set up, the reactive layer and the deliberative layer, can produce various decisions in accordance with the agent's current emotional state, the agent may still behave in a less human fashion. One main deficiency exists in that the robot may seem oblivious to all the scenarios it has experienced before and may be determined to continue challenging you no matter how many battles it has lost, (ie: 10(human):0(robot)). On the other hand, it may always try to escape from you no matter how weakly you behave, (ie 0(human):10(robot)).

In this way, the agent has no sense of coherency or continuity with its past experiences. In order to rectify this situation, we can create two rules that will program the agent to make the correct response. First, if the you (the robot), has consecutively beaten your opponent more than five times, the robot should act more aggressively. On the contrary, if you (the robot) have continued to loose out to the opponent more than five times, you act less aggressively. If neither of the two scenarios exists, then you (the robot) should continue using your current strategies.

It must be said however, that this simplicity is far from the complexity of human capability. People may start to behave aggressively after beating their opponent three times, ten times or even after they have lost the first round. In other words, the first example exhibits diverse behaviours people may choose and a set number or value is not representative of these differences. The second example demonstrates that past experiences can exert continuous effects on a person's immediate decision and cannot be simulated by a set of rigid and inflexible rules.

Due to the above facts, it becomes necessary to add in another layer, the meta-management layer (**Wright et al. 1996**). This is placed on top of the agent brain architecture which is capable of producing long term signals in terms of past experiences. As a result, by synthesizing the signals from both the deliberative and meta-management layers, the final output from the agent's mind to its motion system is believed to guide the agent to produce behaviours coherent to both its current emotional state and past experience.

Adding Beliefs into the Meta-Management Layer

The need to involve the meta-management layer in the agent's mind architecture is to gain higher control or influence over the deliberative layer. This was suggested by (**Wright et al. 1996**, **Sloman 1998**) in the literature review. There are two main concepts involved in this layer. First, with it we can make evaluations or comparisons to the strategies or plans created in the deliberative layer, so that we can suggest that people make better choices in the future. Second, we are able to

persistently offer a new viewpoint even if it is in contradiction with the one offered in the deliberative layer. Sloman thought there were some "tertiary emotions in the Meta-Management layer, such as infatuation, jealousy, grief or pride, that enabled people to ignore or reject something inconsistent with them. For example, people in group A who are jealous of those in group B will most likely not register the latter group's achievements.

Tertiary emotions in the meta-management layer (**Wright et al. 1996**) were interpreted as being enduring, highly resistible and perturbing (referring to the ability to interrupt the current ongoing thinking process and take control). More specific to the problem raised previously, I will make use of the former two features of tertiary emotions to form "belief", while the third feature of "perturbance" is actually partly implemented in the lower layer by the competition mechanism of UECHO, (i.e. one continuous task could be interrupted and switched to another due to a change in emotional state The difference is that "perturbance" may exert a sudden impact that forces the decision generated in the deliberative layer to be changed right way, eventhough the competition mechanism among emotions in the deliberative layer appear to be more "soft" and gradual.

The reason to choose "belief" in the meta-management layer is that it possesses similar characteristics to the tertiary emotions, (i.e. the ability to hold some opinion for the long term, and the ability to resist it), Also, it more accurately describes the "coherence" process as the nature of UECHO determines (**Wang 1998**). Second, the formation of emotional beliefs as explained by (**Frijda et al. 2000**) in the literature

review, provided the evidence needed to consider the addition of beliefs into the meta-management layer. It is important because the formed belief is useful in monitoring the agent's decision making process, (i.e. to keep the selection process coherent with the history), therefore it seems logical to add this necessary component.

Designing the Meta-Management Layer

To be compatible and consistent with the deliberative layer, the design for meta-management layer is still under the connectionist network but with different settings and meanings behind it.

The first design issue is to choose an object for belief updating. That is, what might the agent care about all of the time? The best choice is to look at the evaluation of the opponent's overall performance, as this exerts the greatest impact on people's long term thinking. The three beliefs that the agent possesses in relation to the opponent while fighting are: skilful, comparable or inferior and we may use "Potent","Equivalent" or "Weak" to describe these beliefs.

The second issue surrounds how we choose the inputs for updating the agent's belief. This is similar to the design in the deliberative layer, but events consisting of long term intervals are specified. They are "HScore", "RScore", "HEsc", and "REsc". These four events were chosen for inputs because they best represent the specified belief in the "human's overall performance". It is clear that the higher the score the better the win and the less escapes made, the better the performance can be

rated.[1] Other events such as "SS" and "SA" only represent the local performance for a player, (i.e. the performance during one battle), and they are not chosen as the inputs for long term belief updating.

We have learned from the literature review that the formation of a certain belief is also related to the formation of some dispositional emotion. We may hypothesize that a potential causal relationship existed between the elicitation of emotions and the formation of long term beliefs, (i.e. within a fixed event or object, similar emotional stimuli which is constantly elicited will eventually form a permanent belief). By mapping such a relationship into the connectionist network, we can conclude that there exists one more link between emotion and temporal belief. Such a setting that uses "sentiments" as an important influence on beliefs is also considered to be part of the "internal perception" process in the meta-management layer proposed by (**Wright et al. 1996**). For example, stimuli are not only externally perceived events but also present in the mind.

The third issue is how to connect emotional stimulus to beliefs. It is difficult to discern such a connection until we apply the concept of "concern" mentioned in the literature review. At this time, concern refers to some perception by the mind and not from external stimulus. To simplify and clarify, this assumes that a belief type is completely linked to only one certain concern which may be held by one or more emotions with different focuses.

---

[1] Notice: one may argue people choose escape not only because of fear, but also because of other reasons such as lurk and sneak attack later or seeking rescue items for a better fight. It is still evident that someone, with the score of 5(people): 0(agent) and no escape at all, has a better overall performance than someone with the same score but 10 more escapes.

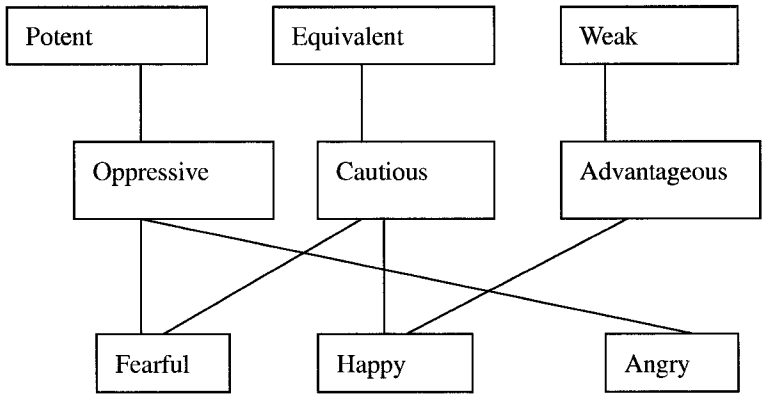

Figure 3.6: Network between Emotions, Concerns and Beliefs.

The above figure shows a rough implementation graph of Frijda's conjecture regarding the relationship between emotions, concerns and beliefs. The top row indicates three beliefs surrounding the opponent's fighting performance. The middle row exhibits the three concerns to which beliefs are attached. The bottom row displays three emotions which may hold one or two concerns from the second row. The figure explains several possible cases in the meta-management layer: When the agent believes its opponent to be strong, the agent feels oppressed causing the agent to act in two ways. One, it either chooses withdrawal because of fear or two, bursts out due to anger. In another case, if the agent believes it has the equivalent strength to the opponent, the pressure is not the same as when the opponent is perceived as strong. When the agent is mostly happy, it is cautious of fighting. On the other hand, if the agent perceives the opponent as weak, it will feel at an advantage and happiness will be increased. As the above design illustrates the definition given before of the three emotions, it is not surprising to see happiness play in two different situations as its nature of rational thinking determines.

From the above illustration, we generated the connectionist network design for the meta-management layer as follows:

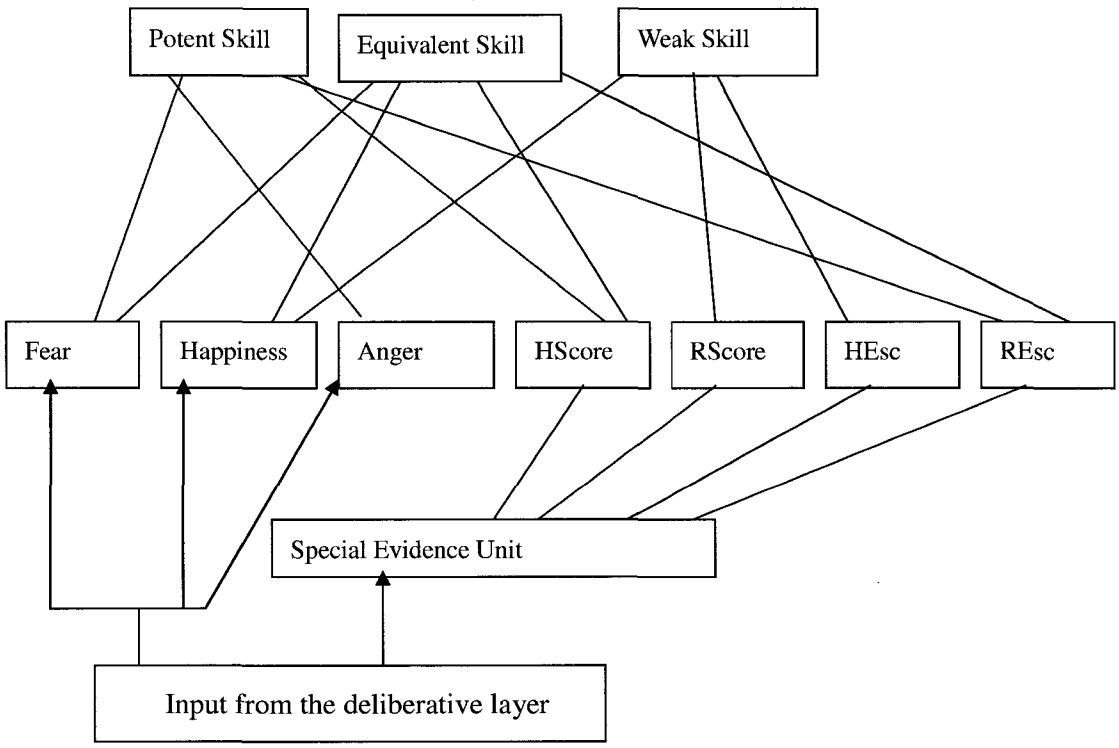

Figure 3.7: The Connectionist Network of the Meta-Management Layer. The solid lines between each of the two boxes indicates a positive effect between them; while the dashed lines represent a negative effect between two emotions.

The above figure displays the structure of the belief network in the meta-management layer according to the description in the previous paragraphs. Notice that the layer of concerns is omitted and that the beliefs are directly linked to emotions. We have clearly seen the relationship between those two groups without the assistance of concerns. Aside from the emotions which help form beliefs with their activation values from the deliberative layer, four long term interval events are able to

foster beliefs, too. The entire belief revision process mainly follows the revised coherence calculation (see page 54). The scaling parameters for those long term interval events used to create event intensity are displayed in the following table:

	RScore	HScore	REsc	HEsc
Potent Skill	−, - 1.8	+, -1.8	$+,\begin{cases}-0.1(\Delta Score>0)\\0.1\ (\Delta Score\leq 0)\end{cases}$	−, - 1.8
Equivalent Skill	−, - 1.8	$+,\begin{cases}-0.1(\Delta Score>0)\\0.1\ (\Delta Score\leq 0)\end{cases}$	+, -1.8	−, - 1.8
Weak Skill	$+,\begin{cases}-0.1(\Delta Score>0)\\0.1\ (\Delta Score\leq 0)\end{cases}$	−, -1.8	−, - 1.8	$+,\begin{cases}-0.1(\Delta Score>0)\\0.1\ (\Delta Score\leq 0)\end{cases}$

Table 3.5 Scaling Parameters Table in the Meta-Management Layer

The fourth issue is how to reflect on how the impact from the meta-management layer to the deliberative layer keeps the agent's thoughts coherent with the past. As already mentioned in the literature review, and the early part of the previous section, such an impact can be exerted by the beliefs. Briefly speaking, one formed belief can elicit its own strength to influence the current emotional state by releasing the dispositional emotion, the sentiment. To map this point, the implementation of beliefs will be involved in the revised coherence calculations that happen in the deliberative

layer. Something that needs to be stressed is that the emotional effects from beliefs are "dispositional", meaning that the impacts under the current belief state should be constant in a certain period. The activation value of beliefs are not allowed to be updated during the revised coherence calculations until some long term event is perceived which signals the start of belief revision in the meta-management layer.

If we turn back to the figure 3.5, we can find the box labelled "long term effect" which was left unexplained. It actually denotes the impact from beliefs. When comparing figure 3.5 with figure 3.7, it is easy to tell that the box "long term effect" in figure 3.5 is composed of three smaller boxes each of which represents a belief candidate in figure 3.7. Since the meaning of the "long term effect" is clearly explained here, it is necessary to raise an example to illustrate how to keep the agent's decision coherent to its beliefs. Suppose the current belief state is (0.1, 0.1, 0.8) in the order of "Potent, Equivalent, or Weak", and the current emotional state is (0.6, 0.3, 0.1) in the order of "Fear, Happiness, or Anger": it is obvious that the current belief will relieve some fearful feelings in the agent after the revised coherence calculation in the deliberative layer is something like (0.41, 0.25, 0.44). This kind of emotional signal guides the agent to make some wiser decisions by synthesizing both the incidental effects and the long term effects. For example, the agent may choose "dodge&fight" instead of "forward&fight" suggested by the current belief state, or "escape" suggested by the current emotional state.

It is time to generalize the entire "coherence" working process designed between the deliberative layer and the meta-management layer. When the agent encounters a

long term interval event, (HScore, RScore and so on), the whole emotional state is calculated in the deliberative layer by integrating the long term interval event. Next, it will send the produced emotional state with the event together to the meta-management layer where current beliefs will be updated according to the received emotional stimuli and event. Once the beliefs are updated, the meta-management layer will return the updated beliefs and the updated emotional state to the deliberative layer as the initial state of the next battle. So, when the agent encounters its enemy the next time, its emotional state in the deliberative layer will always be calculated by integrating the updated beliefs when the same events are perceived. Doing so makes the produced emotional state coherent to both the current situation and the past impression. When one long term interval event happens, the process will start again at the beginning as outlined in the beginning of this paragraph.

A Complete Working Flow in the Agent's Mind Architecture

So far the complete introduction has been finished regarding the three layers agent architecture. We may have a look at the integrated map of it in order to form a full impression on it (See figure 11 below):

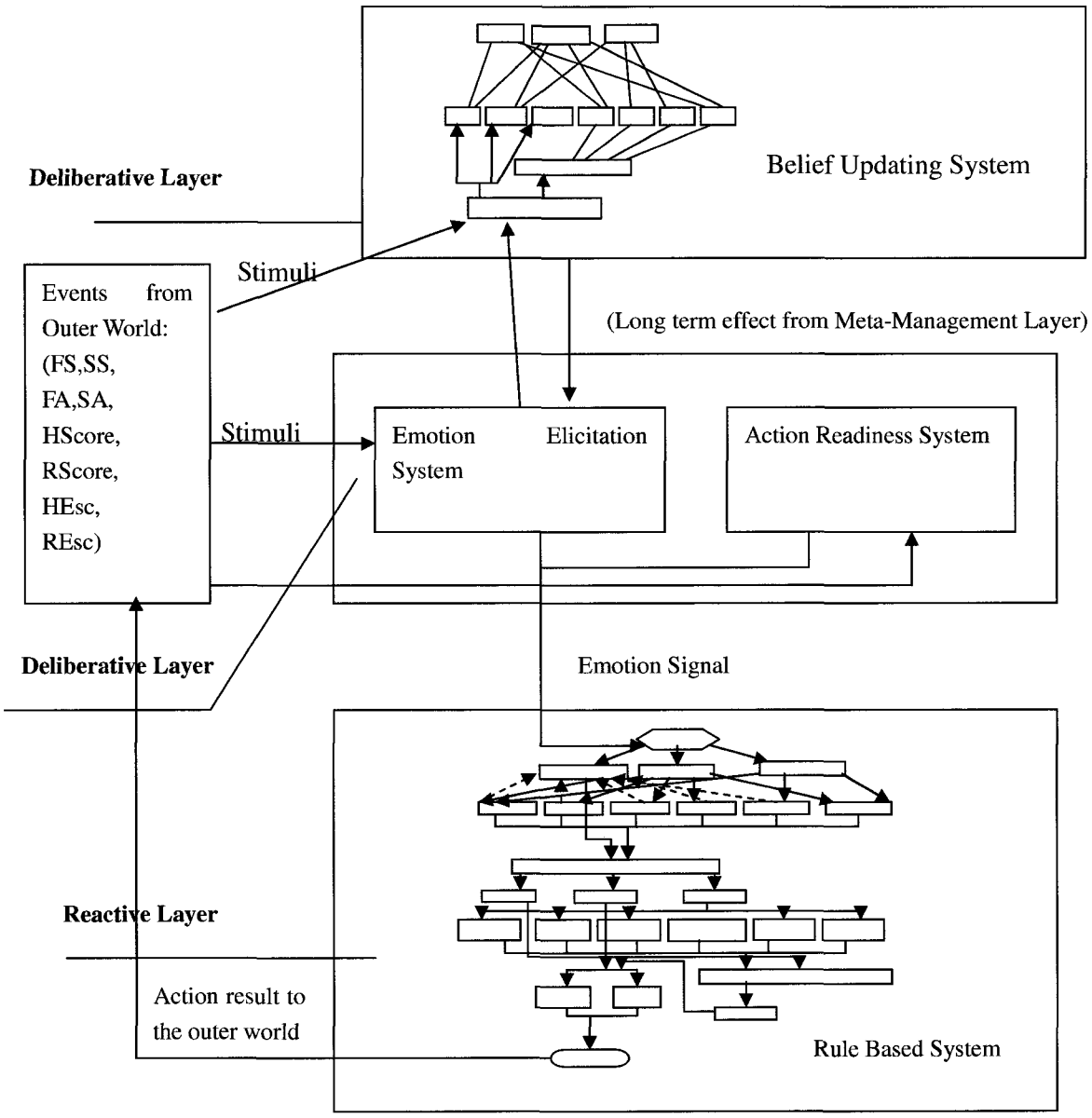

Figure 3.8: Complete Agent's Mind Architecture. The three layers emotional agent architecture is clearly presented under my implementation.

Experiment Design

Experiment Purposes

To validate the functionality of each layer of the agent's mind architecture proposed previously, one experiment was set up with twenty human subjects. The experiment was set up for two purposes: one is to test if agents wired with emotion component(s) are able to behave much "better" than those without in some or all testing aspects. The word "better" here refers to the significant enhancement in certain experiment measure which can be computed under the statistic method ANOVA (Analysis of Variance) at the significant level of 0.05. In other words, the first purpose is to test if emotion is really helpful to enhance the agent's performance. The second is to test if the three layers proposed in the last chapter are all necessary to enhance the agent's performance. The above two hypotheses were tested by five experiment measures which will be mentioned soon.

For the first aim, it is simple to deal with, i.e. we can compare the performance between the full structure agent and the agent only with the reactive layer. However, to be compatible with the second aim, the first aim will be extended to make comparisons between the emotionless agent and three other types of "emotional" agents. If the performance from the experiment demonstrates one or more agents with emotion components behave better than the agents without, we can conclude the certain emotion architecture is useful. Otherwise we may say the testing emotional

agent architecture is unable to embody its superiority over the regular rule based system. Notice "better" here will be measured in five different types of measures, which will be expounded in the next section.

The reason to create three types of emotion agents is for the second aim, i.e. to test if the full structure agent has the best performance over the emotional agents with only either part of the two upper layers. Again, if the full structured agent does not have better performance than the other emotional agents, we may conclude that some layer may be not necessarily added in, or some other factors could be analyzed that they hold back the good performance from the layer.

As a result, four types of agents are created: Reactive Agent (RA) which is emotionless due to only reactive layer wired, Emotion Only Agent (EOA) which is the agent with reactive layer and deliberative layer, i.e. the one can only be affected by incidental emotions, Belief Only Agent (BOA) which is equipped with reactive layer and meta-management layer, i.e. the one can only be affected by the expected emotions, and the Full Agent (FuA) which possesses all three layers introduced in the early of this chapter. Certainly, we expect FuA is able to bring outstanding performance over the other three types of agents.

Experiment Process Introduction

The general experiment process is as follows: Twenty subjects, ten dyads in total are invited. The experiment is formed of five sessions, in each of which two dyads among ten will be invited without duplication. One session consists of five phases,

each of which is finely divided into two stages: challenging stage and rating stage in sequence. In each phase, each subject will be randomly assigned with an opponent, without knowing whom they will be played with beforehand; the opponent can be his or her partner in the same dyad or one of the four types of agents mentioned previously. The opponents' appearing sequence for one subject will be created according to the Latin Square Order. The sample sequence order can be referred to Appendix B.

One challenging stage will last for 9 minutes in which the subject will fight his or her opponent in the uniform experiment map; the goal is simple: the subject is asked to seek the opponent and eliminate it when encountering it. Following the stage is the rating stage in which each subject will be asked to rate their score on five types of measures by filling out the question form within 3 minutes. The form contains questions or columns regarding the five measures which will be explained in the next section. The question form can be referred to the Appendix D. Since the challenging stage and the subsequent rating stage will be repeated five times to form a complete session, one session will last for exactly one hour.

All five sessions are all held in the same office room where four computers having the exactly same configuration are connected within the same local area network. They are placed in two rows of the tables onto each of which lines two computers; the monitors of the computers in different lines are positioned face to face. Each computer of the four is installed with the testing agents and the trial version of Quake2. The four participants will be asked to sit in front of the four computers

respectively, and the two people in the same dyads are seated back to back, i.e. the computers they use are not in the same row; doing so can prevent people from recognizing his or her opponent as a human only because the subject discerns the actions from the screen can be mapped to the operations performed by the person sitting next to him or her. Rather, we expect subjects to judge their opponents' performance by conceiving what they perceive from the game.

Each session there are six persons in locale: four subjects, one operator (the author) and one coordinator. The operator is responsible to claim start or end of a stage, and also for setting up the correct opponent for each subject in each challenging stage. The coordinator helps to collect marked question forms and replies necessary questions raised by subjects, such as how to customise personal controls before game starts. When in challenge stage, nobody will be allowed to talk with each other unless somebody decides to quit the experiment or the instructor claims the end of the stage. Marked question forms will be temporarily kept by the coordinator until the entire session is over; during one session, the instructor will not be allowed to know any information from the marked question form.

All the subjects in one session were paid 10 dollars after the session as described above was done. In the conducted experiment, they were all between the ages of twenty and twenty five and they were composed of four females and sixteen males. All five sessions were all composed of the subjects of the same gender, and all dyads were randomly paired within a session. No subject ceased his or her participation before the session ended. More details can be found in the Appendix C, the instruction

script.

Experiment Measures

To test all the designed agents' performance in the game Quake2, five parameters have been chosen as measures, human believability, effectiveness and preference, long term effect and incidental effect. They were all collected from the question form: human believability, long term effect and incidental effect and preference map to the question 1 to 4 respectively and data for effectiveness is from the score table (See Appendix D for details). Among them, except the effectiveness, all the other four measures are subjective measures, i.e. they were obtained through subjects' rating result. Effectiveness is objective measure and it was actually represented by the scores between the subject and his or her opponent in a battle.

Human believability here refers to the degree of what an agent behaves closely to a human. To test it, subjects will be asked to rate their last opponent in terms of its general performance. In other words, subjects will guess how possible their last opponent was actually acted by a human according to their impression. Their rating score should fall within the range 1 to 10. 1 means one subject fully believes his or her last opponent was a robot, while 10 means the subject fully believes the opponent was a human.

Preference refers to the degree of what an agent is favoured by a subject. Similarly to the Human believability, it will be obtained by asking subjects to score after one play; and the mark is also bounded between 1 and 10. 1 means the subject

does not like the opponent at all, and 10 means he or she appreciates the opponent very much.

Effectiveness refers to the measure of one agent's fighting performance. The data about it will be collected from the net income after one subject's score subtracting his or her opponent's score in one battle:

$$Eff(A_i, Sub_n) = Scr(A_i) - Scr(Sub_n) \quad (4.1)$$

In (4.1), A_i denotes agent i, and Sub_n denotes the n th subject, Scr is the abbreviation of "score", and Eff for "effectiveness".

To be directly related to the design of the agent architecture, two more auxiliary measures are chosen: long term effect and incidental effect.

Long term effect refers to one agent's ability of keeping coherent to the past experience as human have. Since human are able to change their fighting attitudes according to their general fighting performance, such as human may behave more aggressive if they outperform their opponent much in the previous battles. The measure will be helpful to check if my designed meta-management layer could have some equivalent performance as human does. To test it, subjects will be asked to measure the coherent degree of changes in their last opponent's behaviours or strategies within the entire nine minutes' challenging stage. 1 means the subject does not observe any coherent change from their opponent in a challenging stage, 10 means the subject thinks what the opponent behaved is perfectly coherent to what an average person could do in a challenging stage.

Incidental effect refers to one agent's ability of making human-like adaptation

during a short term, say in a battle encounter. Since human are able to adjust their fighting skill according to the stimuli they receive in a battle, the measure will be helpful to check if the designed deliberative layer could have some equivalent adaptation as human does. Similarly to the above measure, 1 means the subject does not observe any human-like adaptation made by their opponent in battles, 10 means the subject thinks opponent's adaptability in fight is totally like what a human does in fight.

Since the other measures except effectiveness are all subjective ones, the data collected about them should be converted to the values relative to the human opponent's corresponding scores before performing statistical process[1]. The main reason to do so is it is to normalize the scores to a subject's "base-line" of what they consider human. Therefore, we need to convert different ratings into the relative score before doing any statistics. One participant's relative feed back to a certain type of agent can be computed as follows:

$$Fb(A_i, M_j) = Rt(A_i, M_j) - Rt(H, M_j) \quad (4.2)$$

In the above formula, $Fb(A_i, M_j)$ means one subject's final feedback to the agent A_i about the measure M_j is the difference between his or her rating to the agent $Rt(A_i, M_j)$ and his or her human opponent $Rt(H, M_j)$ in the same measure..

After the above conversion, the data for five types of measures will be processed by single Repeated Measures ANOVA (Analysis of Variance).

[1] It is originally suggested by my supervisor Dr. Joesoph MacInnes.

Summary of Methodology

In this chapter, the methodology was introduced on how to design an emotional agent by using Sloman's three layers mind architecture, and it was followed by the validation component, the experiment design.

Regarding the agent design, we started with a discussion on the deficiency a rule based system has, i.e. rigid and reflexive only mechanism, which is incapable of reacting in diverse ways as human does. Then, a potential solution towards the problem is proposed, adoption of the emotion theory to the rule based system. The main framework of the agent architecture is based on the Sloman's Three Layers conjecture about human's mind, but with simplification and improvement tailored to the game Quake2. Two highlights can be identified with the design: one is adding the self-adaptation mechanism to emotion system which follows Damasio's Somatic Markers Hypothesis but with more specification: each emotion may have its own feature to mark objects or events, instead of simply grouping the emotional effects by positive or negative ones. Such adaptation mechanism also demonstrates that the emotional decision should not be made only according to its current emotional state, but also based on the action readiness in an object or event. The combination of considering the above two factors could generate more smooth and more realistic decisions as human does. The other highlight is to make use of beliefs in the meta-management layer, as it is expected to guide agent to make decision not only in terms of the current generated emotional signals, but also taking the account of past experience or impression.

Regarding validation on the agent architecture, the experiment design, has been presented with detailed expound. The experiment will be used to mainly testify two hypotheses under the proposed agent architecture: one is the emotional agent should behave more human-like than the emotionless agent; second, the agent with the full of proposed agent architecture should outperform any other types of agents with only part of emotional architecture. In other words, if both hypotheses can be proved true, the agent architecture is definitely meaningful to the future emotional agent design. To make the two hypotheses measurable, five types of parameters were picked up as measures for all types of testing agents: believability, effectiveness, preference, long term effect and incidental effect. The method of how to measure them is presented later with explanation. Finally, a sufficiently described experiment process was given.

Results and Analysis

This chapter will present the experiment result with explanation. It is followed by the conclusions regarding the validation result of the proposed agent architecture.

After running the experiment and collecting the data for those five measures introduced in the "experiment design" of the last chapter, we will first test if there is any correlation between the four measures; if any correlation is found, we may infer the experiment result may not be objective as it may be influenced by the correlation to some extent. And then we will make Repeated Measures ANOVA on those experiment measures in order to see any significant difference among them. If some significant difference in a measure is observed, the further pair wise comparisons among those five types of opponents will be adopted by using Turkey's HSD (Honestly Significance Difference) algorithms.

Correlation Analysis

The correlations between each pair of the five measures have been examined by running SPSS as the following table shows:

		Beli	Prf	LTE	STE
Beli	Pearson Correlation	1	**.544**	**.439**	**.484**
Prf	Pearson Correlation	**.544**	1	**.411**	**.638**
LTE	Pearson Correlation	**.439**	**.411**	1	**.496**
STE	Pearson Correlation	**.484**	**.638**	**.496**	1

Table 4.1: Correlations between Each Pair of the Five Measures. A bold figure in the table denotes a significant correlation between the measure of the row and the one of the column at the level of 0.01 (Beli: Believability, Prf: Preference, LTE: Long Term Effect, STE: Short Term Effect).

Since the significant correlations have been found between each pair of those four subjective measures, it indicates each measure was probably influenced by the other three measures. For example, the highest correlation 0.638 between STE and Prf implies that subjects preferred the agents with good fighting skills most, but they paid less concerns on the agent's long term behaviours and strategies (0.411 between LTE and Prf). Besides, we also find that the correlation between Believability and Preference is 0.544, the second highest one among the six correlation values. This implies that to some certain extent subjects who liked playing with some opponent tended to rate a high score for their opponent's Believability, and vice versa. The final finding is that the correlation between STE and LTE is 0.496, which implies the subjects may rate STE and LTE in the same trend, i.e. either high marks in or low marks in the two measures.

Statistics Results for Believability

First, let us watch the graph about the standard means of the five types of opponents (including human):

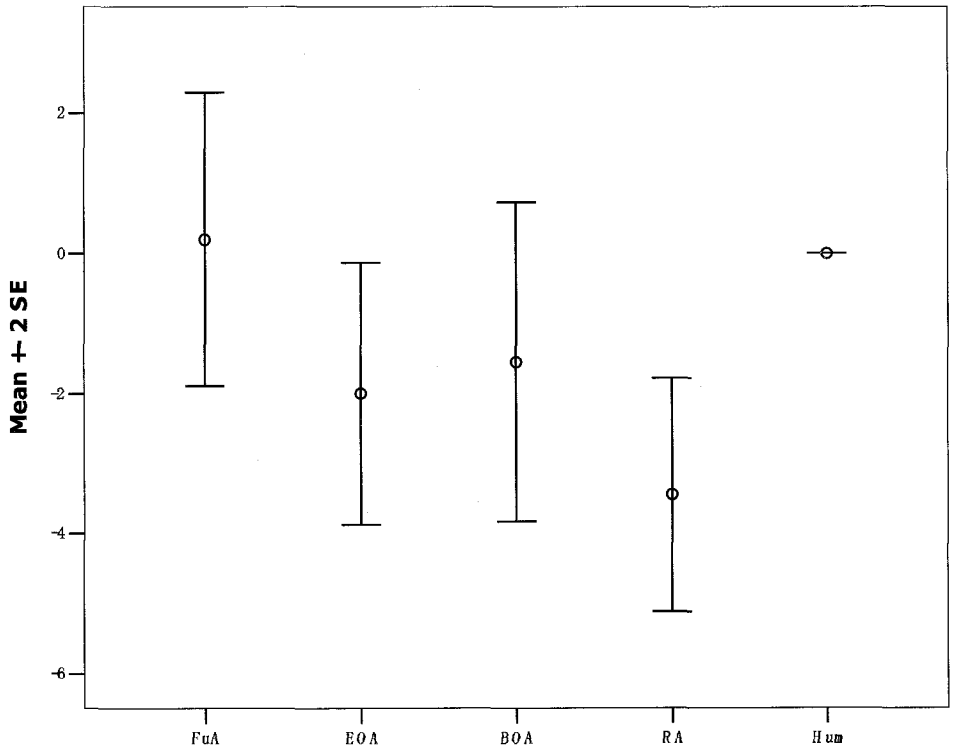

Figure 4.1 :Standard Means for Five Types of Agents on Believability

From the above figure, it is manifest to see the FuA has the highest mean value in believability over the other four including human opponent. The result seems surprising as FuA even surpasses the humans in believability although the former is not significantly better than the latter (see table 4.3 below). It is rational to see this result; it is not only because relative complete emotion dealing mechanism is wired in

FuA, but also because not all human participants were proficient in game playing and judgement; some inexperience subjects may behave poorer than FuA or make wrong judgement on their opponent's overall performance.

By running SPSS to perform the Repeated Measures of ANOVA with the confidence interval of 0.05, we could obtain the following statistic result which extracted from the raw tables generated by SPSS:

The variable Believability is significant as $f(4,76)=5.13,\ p<0.001$.

Therefore, we may further figure out which pair or pairs of objects have such difference by using Turkey's HSD as table 4.2 shows:

Pairwise Comparisons by HSD

Measure: Believability ("Beli" in the following table)

Beli (i)	Beli (j)	Mean Difference (i--j)	Significance
FuA	EOA	2.200	
	BOA	1.750	
	RA	3.650*	
	Hum	.200	
EOA	FuA	-2.200	
	BOA	-.450	
	RA	1.450	
	Hum	-2.000	
BOA	FuA	-1.750	
	EOA	.450	
	RA	1.900	2.65
	Hum	-1.550	
RA	**FuA**	-3.650*	
	EOA	-1.450	
	BOA	-1.900	
	Hum	-3.450*	
Hum	FuA	-.200	
	EOA	2.000	
	BOA	1.550	
	RA	3.450*	

Table 4.2: Pairwise Comparisons for Believability by HSD.

Table 4.2 indicates that FuA and Human opponents both have the significantly better performance in Believability than RA. The finding suggests the emotional agent, at least FuA, is able to enhance the believability for game agents.

Statistics Results for Effectiveness

Again, similarly to the above procedures, the standard means of the five types of opponents (including human) for Effectiveness are:

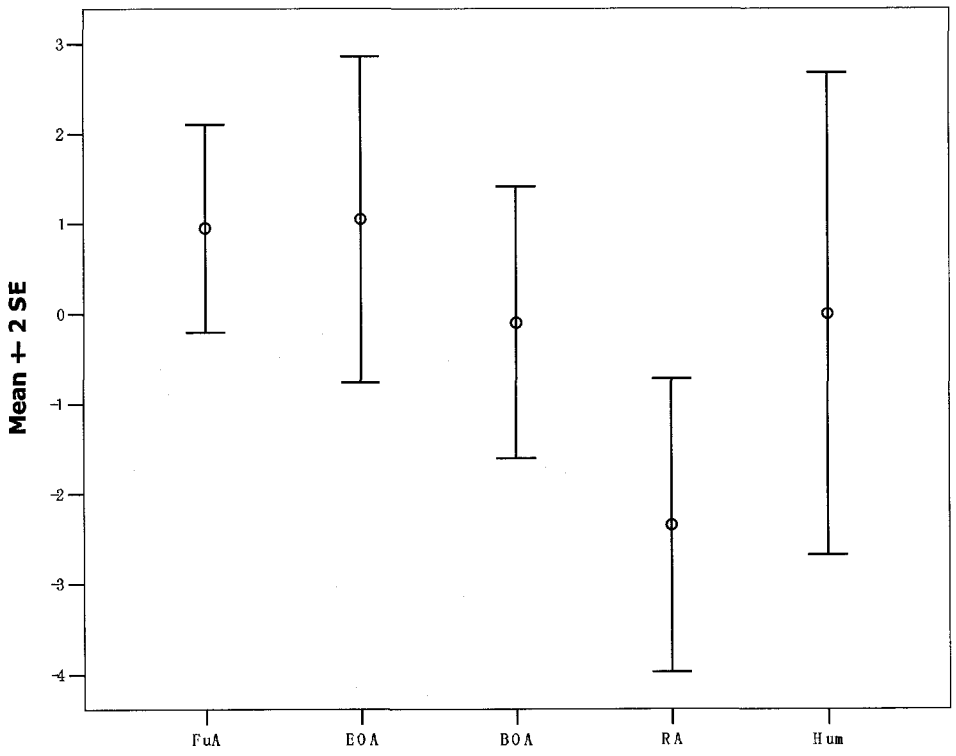

Figure 4.2 :Standard Means for Five Types of Agents on Effectiveness

From the above graph, we can find FuA did not behave ideally; it only had

significantly better performance than RA. It is intuitive to see EOA has the highest standard mean on this category.

Again, by running SPSS in repeated measures ANOVA, we can obtain the result that the variable effectiveness is significant as $f(3.05, 57.89) = 3.55$, $p < 0.02$.

Therefore, we may further figure out which pair or pairs of objects have such difference as table 4.3 shows:

Pairwise Comparisons by HSD

Measure: Effectiveness ("Eff" in the following table)

Eff(i)	Eff (j)	Mean Difference (i--j)	Significance
FuA	EOA	-.100	
	BOA	1.050	
	RA	3.300	
	Hum	.950	
EOA	FuA	.100	
	BOA	1.150	
	RA	3.400*	
	Hum	1.050	
BOA	FuA	-1.050	
	EOA	-1.150	
	RA	2.250	3.36
	Hum	-.100	
RA	FuA	-3.300	
	EOA	-3.400*	
	BOA	-2.250	
	Hum	-2.350	
Hum	FuA	-.950	
	EOA	-1.050	
	BOA	.100	
	RA	2.350	

Table 4.3: Pairwise Comparisons for Effectiveness by HSD.

Table 4.3 concludes that only EOA have significantly better fighting performance than RA, and no significant difference among others, although we can discover that FuA has "almost significant" better performance than RA. The finding illustrates some type of emotional agent, at least EOA, can help improve game agent's fighting performance.

Statistics Results for Preference

Again, similarly to the above procedures, the standard means of the five types of opponents (including human) for Preference are:

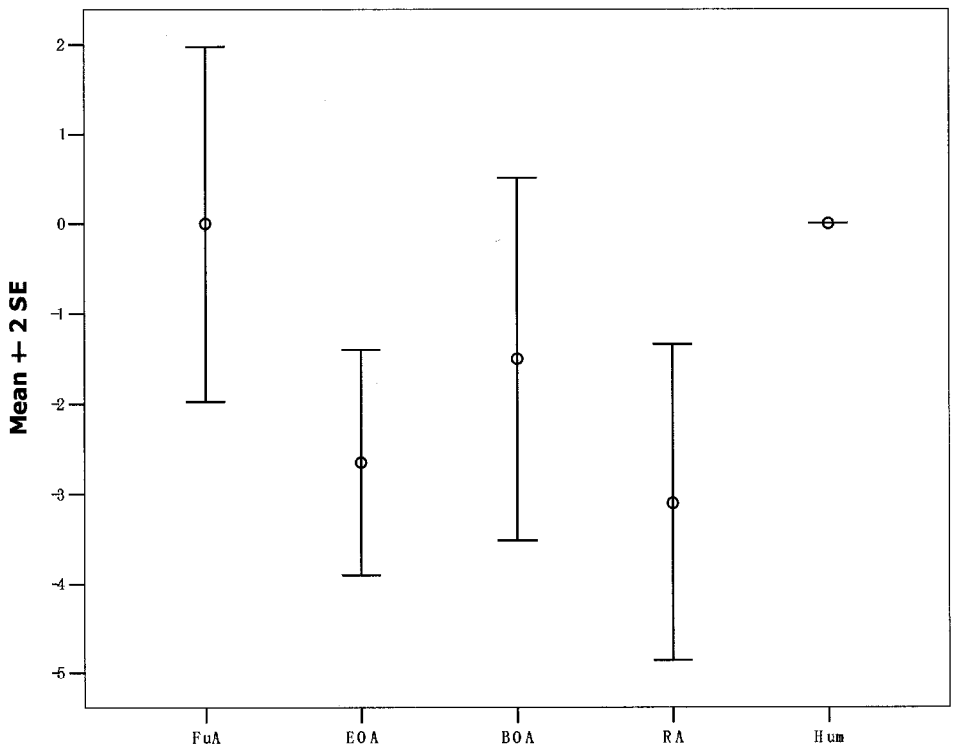

Figure 4.3 :Standard Means for Five Types of Agents on Preference

From the above graph, we can find FuA won the favour from subjects as it has the highest standard mean on this category. BOA is also outstanding but less fancied by subjects than FuA.

Again, by running SPSS in repeated measures ANOVA, we can obtain the result that the variable Preference is significant as $f(4,76)=5.17$, $p<0.001$.

Therefore, we may further figure out which pair or pairs of objects have such difference as table 4.4 shows:

Pairwise Comparisons by HSD

Measure: Preference ("Pref" in the fcllowing table)

Pref (i)	Pref (j)	Mean Difference (i--j)	Significance
FuA	**EOA**	2.650*	
	BOA	1.500	
	RA	3.100*	
	Hum	.000	
EOA	**FuA**	-2.650*	
	BOA	-1.150	
	RA	.450	
	Hum	-2.650*	
BOA	FuA	-1.500	
	EOA	1.150	
	RA	1.600	
	Hum	-1.500	2.53
RA	**FuA**	-3.100*	
	EOA	-.450	
	BOA	-1.600	
	Hum	-3.100*	
Hum	FuA	.000	
	EOA	2.650*	
	BOA	1.500	
	RA	3.100*	

Table 4.4: Pairwise Comparisons for Preference by HSD.

In table 4.4, the analysis of preference implies that FuA has significantly better performance than EOA and RA, and human opponents performed significantly better than EOA and RA. The finding from table 4.4 suggests FuA is mostly favoured by human subjects as it possesses the emotional architecture while the RA does not, and FuA also wins more favour than EOA and BOA, especially much more than EOA. By comparing the difference among the three types of agents' architectures, and also some evidence from correlations discussed before (since the STE and LTE are both positively correlated to Preference), it could infer that the combination of the two types of emotions within an agent's architecture is necessary to enhance the agent's performance, at least for the increase of human players' favourite degree to the game agent.

Statistics Results for Long Term Effect

The standard means of the five types of opponents (including human) for Long Term Effect are:

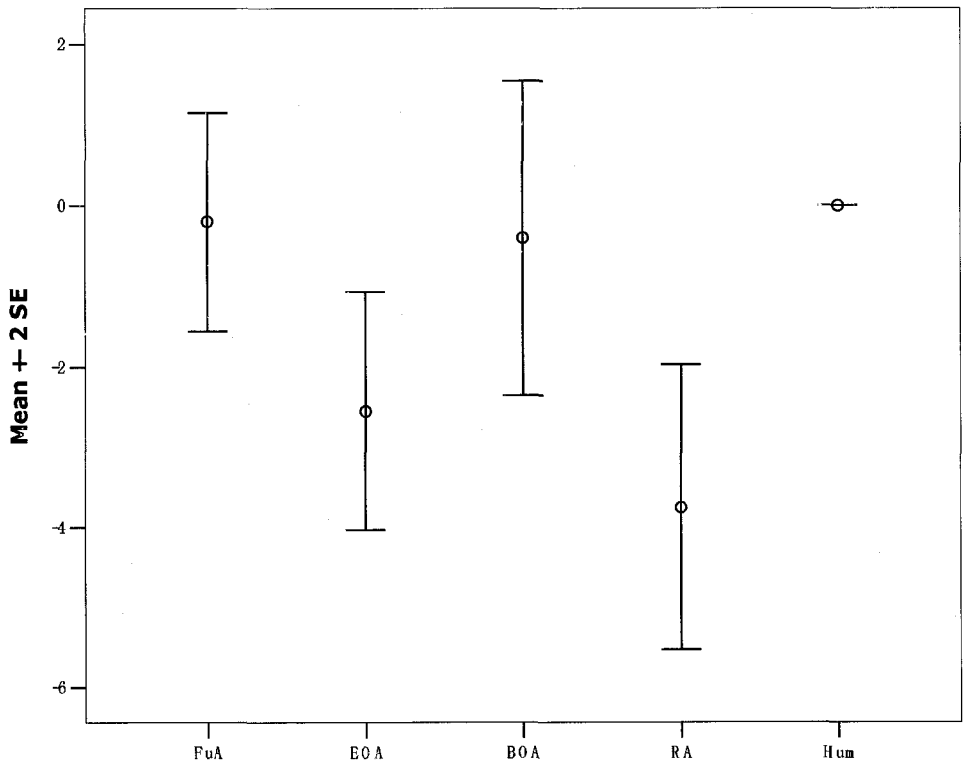

Figure 4.4: Standard Means for Five Types of Agents on Long Term Effect

From the above graph, we can find FuA and BOA are both outstanding on this category.

Again, by running SPSS in repeated measures ANOVA, we can obtain the result that the variable Long Term Effect is significant as $f(4,76)=8.45$, $p<0.001$

Therefore, we may further figure out which pair or pairs of objects have such difference as table 4.5 shows:

Pairwise Comparisons by HSD

Measure: Long Term Effect ("LTE" in the following table)

LTE (i)	LTE(j)	Mean Difference (i--j)	Significance
FuA	**EOA**	2.350*	
	BOA	.200	
	RA	3.550*	
	Hum	-.200	
EOA	**FuA**	-2.350*	
	BOA	-2.150	
	RA	1.200	
	Hum	-2.550*	
BOA	FuA	-.200	
	EOA	2.150	
	RA	3.350*	
	Hum	-.400	2.30
RA	**FuA**	-3.550*	
	EOA	-1.200	
	BOA	-3.350*	
	Hum	-3.750*	
Hum	FuA	.200	
	EOA	2.550*	
	BOA	.400	
	RA	3.750*	

Table 4.5: Pairwise Comparisons for Long Term Effect by HSD.

Table 4.5 suggests that FuA evidently has more consistent performance than EOA and RA as the former is equipped with the third layer where long term belief updating system resides in. Human opponents are certainly significantly better than EOA and RA, too.

Since FuA has satisfactory performance on the Long Term Effect, we may attribute the result to their possession of belief component. It further demonstrates Belief component is helpful to enhance the agent's coherence in the long term run.

Statistics Results for Incidental Effect

The standard means of the five types of opponents (including human) for Incidental Effect are:

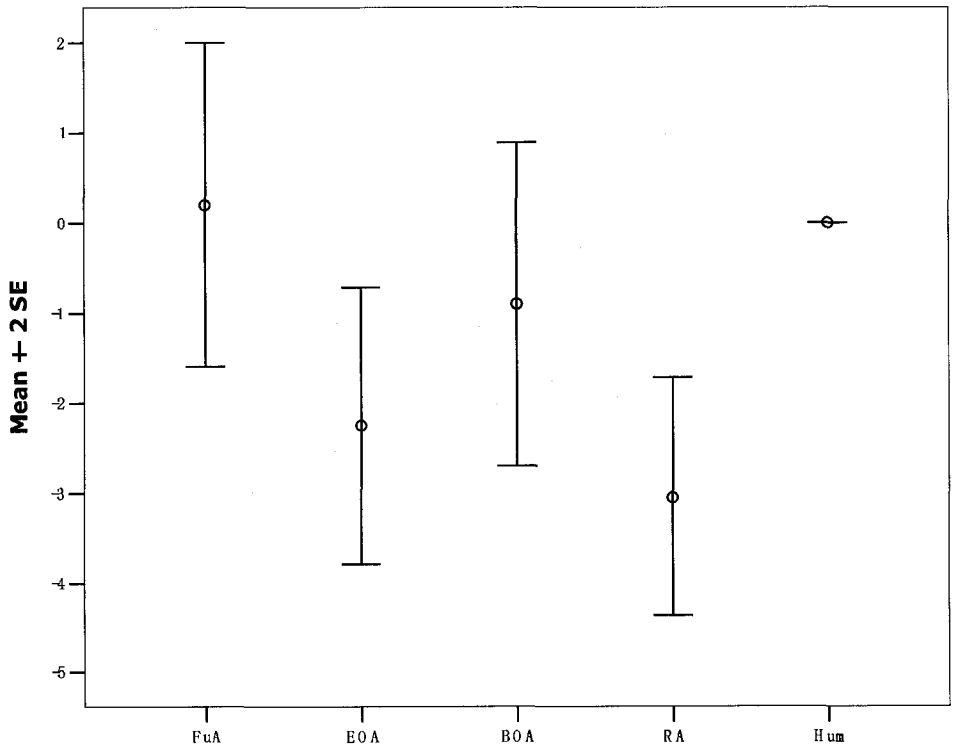

Figure 4.5 :Standard Means for Five Types of Agents on Incidental Effect

From the above graph, we can find FuA is outstanding on this category. It also has the slightly higher means than Human.

Again, by running SPSS in repeated measures ANOVA, we can obtain the result that the variable Incidental Effect is significant as $f(3.10,58.87)=4.87$, $p<0.004$

Therefore, we may further figure out which pair or pairs of objects have such difference as table 4.6 shows:

Pairwise Comparisons by HSD

Measure: Incidental Effect ("IE" in the following table)

IE (i)	IE(j)	Mean Difference (i--j)	Significance
FuA	EOA	2.450	2.94
	BOA	1.100	
	RA	3.250*	
	Hum	.200	
EOA	FuA	-2.450	
	BOA	-1.350	
	RA	.800	
	Hum	-2.250	
BOA	FuA	-1.100	
	EOA	1.350	
	RA	2.150	
	Hum	-.900	
RA	**FuA**	-3.250*	
	EOA	-.800	
	BOA	-2.150	
	Hum	-3.050*	
Hum	FuA	-.200	
	EOA	2.250	
	BOA	.900	
	RA	3.050*	

Table 4.6: Pairwise Comparisons for Incidental Effect by HSD.

Table 4.6 only suggests FuA and Human have the significantly better performance in Incidental Effect than RA.

Statistics on the Overall Believability of the Agents

By synthesizing the above four measures, we may further figure out their overall performance based on those measures in order to obtain a rough rank order for them. The procedure is to work out the averages of the four measures rated by each subject, and then make Repeated ANOVA to analyze their overall ratings for those five types of testing objects:

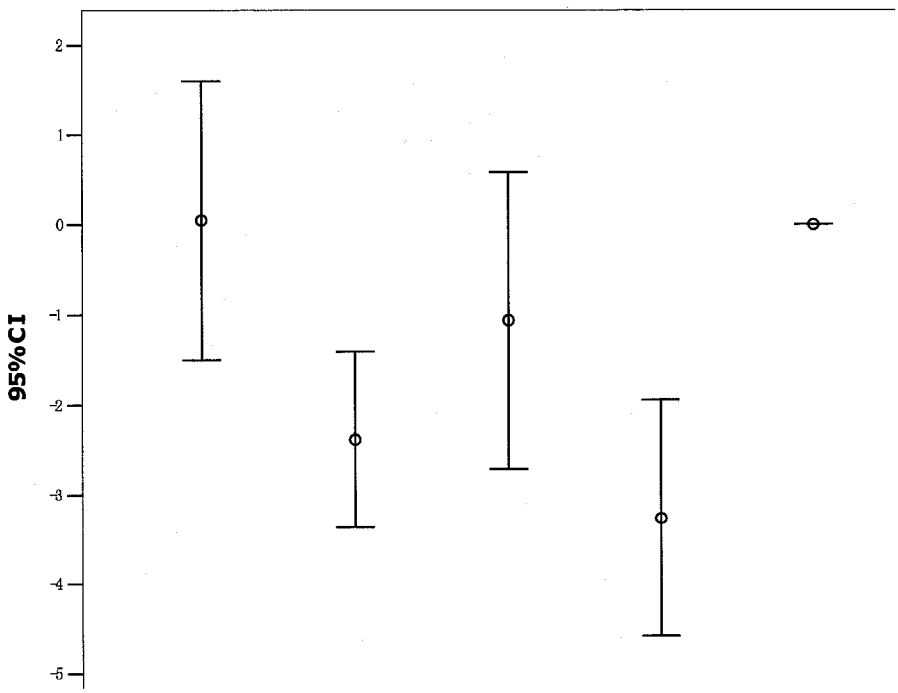

Figure 4.6 :Standard Means for Five Types of Agents on Overall Believability

From the above graph, we can find FuA is outstanding on this category. It also has the slightly higher means than Human.

Again, by running SPSS in repeated measures ANOVA, we can obtain the result that the Overall Believability is significant as $f(4,76)=9.515,\ p<0.001$

Therefore, we may further figure out which pair or pairs of objects have such a significant difference as table 4.7 shows:

Pairwise Comparisons by HSD

Measure: Incidental Effect ("OB" in the following table)

OB (i)	OB(j)	Mean Difference (i--j)	Significance
FuA	EOA	2.438	
	BOA	1.113	
	RA	3.313*	
	Hum	.050	
EOA	FuA	-2.438	
	BOA	-1.325	
	RA	.875	
	Hum	-2.388	
BOA	FuA	-1.113	
	EOA	1.325	
	RA	2.200	
	Hum	-1.063	2.65
RA	**FuA**	-3.313*	
	EOA	-.875	
	BOA	-2.200	
	Hum	-3.263*	
Hum	FuA	-.050	
	EOA	2.388	
	BOA	1.063	
	RA	3.263*	

Table 4.7: Pairwise Comparisons for Incidental Effect by HSD.

From the overall performance, we may discover that FuA and Human both behaved significantly better than RA. It further confirms the hypothesis one in overall that emotion with appropriate agent architecture is able to perform much better than emotionless ones.

Conclusions on the Experiment Result

By synthesizing the conclusions from the above five measures by HSD, we can produce the following table which reflects each types of opponent's performance relative to the others.

	Believability (better than)	Preference (better than)	Effectiveness (better than)	Long Term Effect (better than)	Incidental Effect (better than)	Overall Believability (better than)
FuA	RA	EOA, RA		EOA, RA	RA	RA
EOA			RA			
BOA				RA		
RA						
Human	RA	EOA, RA		EOA, RA	RA	RA

Table 4.8: Synthesis of the Statistic Result

From the above table, it is intuitive to see FuA possesses sufficiently good performance in the experiment as it almost behaved the same as what human participants did. On one hand, it is worth noting because it demonstrates most of the

hypothesis one proposed in the third chapter:

One is emotional components are basically necessary to enhance the agents' playing performance in games. As HSD suggests, FuA, which is fully equipped with emotion components, have remarkably high ratings than the emotionless agent (RA) in four of the five measures. EOA and BOA, the agents wired with only part of emotional components, show fewer positive results than FuA did, i.e., EOA only behaved significantly better than RA in Effectiveness and BOA in Long Term Effect. But EOA and BOA at least reflects their certain enhancement in agent's playing performance by adding emotions. According to the above facts, it is rational to conclude that only the combination of the two types of emotional components is able to bring the most satisfactory enhancement in agent's overall performance.

Another is table 4.8 testifies the issue of the emotional agent architecture design. As HSD suggests, FuA outperforms EOA in Long Term Effect and Preference. The finding demonstrates the necessity to take into account the functionality of long term effect when modelling emotions; FuA has the same settings as EOA has but possesses one more layer, the meta-management layer, which would be considered the main reason why FuA performs much better than EOA in the Long Term Effect. Furthermore, the fact that FuA is remarkably favoured by most of subjects over EOA also suggests human players prefer the setting of the long term emotions. On the other side, the finding that the BOA alone was unable to embody its superiority over EOA in Long Term Effect again implies the necessity of using the combination of the two different types of emotions.

On the other hand, hypothesis two is failed to be demonstrated.

We can discover that FuA did not show any significantly better than EOA or BOA, and the above is what hypothesis two is designed to test. By comparing the structures between FuA and EOA and between FuA and BOA, we may infer that the interaction between the meta-management layer and the deliberative layer holds back the performance of FuA in Effectiveness.

The explanation to the above deficiency and possible solutions will be discussed on the sub section "Possible Amelioration for Testifying Unproved Hypothesis" under the next chapter "Conclusion and Future Work".

Contributions and Future Work

Contributions

This thesis partially tested the two hypotheses proposed in the abstract section and thus the following contributions can be taken into account:

First, emotion theory can be considered useful in enhancing the playing performance for agents in computer games. As was stated in the first hypothesis, FuA possesses significantly better performance than emotionless agent RA in most of the five measures. Other emotional agents also behaved much better than RA in some measures.

Second, the rationality of the proposed agent architecture (FuA) is supported by the evidence collected in the end of the preceding chapter. That is to say, Long term emotions are necessary in the emotional agent design, and only the combination of the two types of emotions can obtain the most satisfactory enhancement for agent's gaming performance. As a result, this thesis suggests that when building emotional agents, the belief system is an essential addition to the emotional agent architecture since emotions released from beliefs are able to make agents' decisions coherent to their past experience.

Third, the thesis develops a simple but feasible interaction mechanism between two types of emotions (see the section related to the design of the meta-management layer), which once received little interest in emotion research or only theoretic frameworks were presented (**Damasio 1994, Loewenstein and Lerner 2003, Wright**

et al. 1996). The lack of consideration and implementation of the higher emotions' working mechanism may be used to explain why some researchers claimed they could only create more diverse but still less human-like emotional decisions (**Henninger et al. 2003**).

Last, this thesis made an innovative attempt to understand the adaptability of emotions as outlined in Damasio's Somatic Markers Hypothesis (**Damasio 1994**) and Frijda's motivation theory (**Frijda 2004**). The proposed emotional updating mechanism is easily adoptable. It is also distinct from other classical artificial intelligence learning algorithms, (reinforcement learning for example), as it is based on experiment findings in neuroscience and thus more closely emulates the human thinking process.

Future Work

The current emotion modelling work is still far from complete. Many aspects must still be improved in the future in order to create more believable agents and more objective experimental results. The rest of this section will outline a possible solution to the unproved hypothesis, and then propose future improvements for either the agent's architecture or the experiment design.

Possible Amelioration for Testifying Unproved Hypothesis in Future

As mentioned in the last chapter about the conclusions from the experiment, FuA

did not outperform BOA and EOA in a significant way on some certain measures as was expected by hypothesis two (See "experiment purpose" under the chapter "Methodology"). By examining the structures between FuA and BOA or EOA, it is easy to tell FuA possesses one more different layer than BOA or EOA, the deliberative layer or the meta-management layer, respectively. The rest of FuA is the same as BOA. Therefore, it is easy to assume that some deficiency in the interaction mechanism between the deliberative layer and the meta-management layer held back the FuA making significantly better performance than them.

This problem or deficiency may be due to the fact that some settings in the agent architecture may overly emphasize long term effects but overlook incidental effects. By recalling the introduced interaction procedure between the deliberative layer and the meta-management layer (see page 104 to 105), the long term effect exerts its own influence on the emotional state in the deliberative layer so as to keep the produced emotional signal being coherent to the past experience. Since a belief can always strengthen its linked emotions, and it is not weakened by other units within the deliberative layer (see page 85), its influence may be sometimes overly powerful compared to other regular units. This prevents the agent from taking any other contradictive information into account. For example, an agent holding the strong belief that its opponent is weak may always select fighting aggressively without considering the successions of negative evidence such as FA and FS. From this point we can find the emotional signals produced by FuA are close to the ones by BOA because the incidental effect in FuA is sometimes too feeble and unconcerned during

the revised coherence calculation. Thus, the two types of agents may produce a similar decision if given the same situations.

There are a number of approaches possible in order to solve this problem. The most direct one is to allow the beliefs to be updated in the deliberative layer. The difficulty is that this violates the hypothesis from **Frijda (2004)** which states that belief is a kind of dispositional emotion so that it should exert the stable emotional influence. Another approach is to modify the linking weights between emotions and beliefs to be weaker in the deliberative layer so that beliefs cannot exert their influences as greatly as before. Yet, such an approach holds the risk of insufficient consideration for the long term effect.

The feasible approach is to build up a mechanism which could rationally involve the long term effect only when it is necessary. That is to say, the involvement of the long term effect may not always be required by the deliberative layer. Rather, in some urgent situation or certain tertiary emotional states (grief for example), the meta-management layer may also possibly lose its control over its bottom layer (**Wright et al. 1996**). Thus, any future work could focus on how to build up a kind of urgency mechanism in order to involve the long term effect in a selective way. For example, the mechanism may specify that if a certain urgent level is reached, the agent should give up considering the long term effect. This kind of consideration also parallels the claims found in most works of emotion scholars who agree that making quick decisions to deal with urgent situations required less deliberation (**LeDoux 1996, Ventura 2000, Sloman 2001**). The above possible solution also

Possible Improvements on the Agent Architecture

Although the agent architecture provides a basic framework to model human emotions, there is still a lot of room for improvement in each of the three layers.

First, all the symbols defined in the reactive layer, including the symbols within the fighting theme, should be given emotion adaptability. Doing so would make the agent more adaptable to any other domain dependent environments where it has little emotional experience.

Second, other intelligence components need to be added in the deliberative layer. As Damasio described in the first chapter, the emotion itself can not substitute the position of intelligence. If we want to gain high believability for agents, we still need to empower it with the ability to perform reasoning, planning or other intelligent activities similar to humans.

Third, it is necessary to extend the ability of the meta-management layer, (i.e. we could add more global control mechanisms over the entire agent architecture). In addition to the urgency mechanism mentioned in the previous section, the issue of how improve control or evaluation on the higher deliberation process, (as suggested by (**Wright et al. 1996, Sloman 2001**)), may also pose as a challenging topic for future implementation on the meta-management layer.

Possible Improvements on the Experiment Design

In this section, a few of experiment design deficiencies are identified below with possible solutions for future ameliorations:

First, we may do some improvement to better test those two hypotheses in future. For the hypothesis one, we can put much more focus on "measuring enhancement" from the emotional agents to emotionless agents instead of sheer comparisons between them. That is to say, in the future experiment, we can further figure out what and how much actually benefit we could gain if we add some emotional component to an emotionless agent; for example, if we find FuA is able to improve the agent's believability, we may further ask how much it is able to improve for emotionless agents. For hypothesis two, we should waive some unnecessary comparisons. In other words, if some emotional agent cannot embody their advantage over emotionless in some experiment measure, it seems not necessary to make further comparisons between this type of emotional agents with the others. For example in the current experiment, it seems EOA is unable to improve the agent's believability, the comparison between FuA and EOA in believability is therefore not necessary.

Second, since it has been found high correlations between those four subjective measures, it may not be sufficient to use only one single question to measure each of them (see Appendix D). A single question may not cover all the important aspects of a measure. For example, during the rating stage of the experiment, some subjects raised questions about the difference of STE and LTE, which implies they did not understand those two measures very well or they could not tell any dissimilarity between them. As a result, more questions need to be designed for each of those measures. Since those questions are able to better embody the characteristics of a single measure, they are believed to reduce the correlations between measures and present more

convincing experimental results.

There are some other reasons to explain the high correlations between the four subjective measures. One is some questions in Appendix D were not well written so that some subjects did not understand the purpose of setting the question. The other is subjects did have their own biases on ratings. From the correlation result table 4.1, we can find people tended to rate high scores on both Preference and STE simultaneously. For the first reason, more clear statement in questions of the Appendix B needs to be made in future. For the second reason, the next point of improvement which suggests recruiting more people seems necessary to reduce the biases produced by individuals.

Third, the quantity of experiment subjects may not embody the diversity for the experiment, i.e. twenty subjects may not be sufficiently to represent most of the people's opinions on my designed agent. Therefore, recruiting more people for the future experiment is necessary.

A fourth improvement would be to categorize the subjects according to their level of playing skill before the experiment. As mentioned before, some subjects were unskilled at Quake2 and so their behaviours were judged as non-human, along with the robot. Thus, it was difficult to distinguish between the behaviours performed by humans from robots. This point is also the reason why even human themselves cannot embody some superior characteristics in long term effects and incidental effects over FuA and BOA. Their opinions should be rectified by recruiting the same percentage of subjects who are in the advanced level of playing skills.

One criterion for such a categorization might be to check if playing scores have

reached some specified mark before the experiment begins. Those players scoring above the mark could be grouped as advanced players, and the others could be grouped as basic players. More refined categorizations according to score levels are also possible, i.e. high, medium and low. This kind of categorization would be expected to result in more objective ratings of the five measures used in the experiment.

A fifth improvement would be to better balance certain experiment criteria within given conditions and increase the experimental conditions tested. In the current experiment settings, there was a gender imbalance among invitees in either the entire experiment or between groups, (i.e. only four participants were females compared to the sixteen male participants for the experiment, and only one group of female subjects compared to the other four groups of male subjects). Nor were there any groups that consisted of an equal number of mixed genders, (i.e. two male subjects against two female subjects in a session). Reasonably, it is necessary to test the equal number of the three types of groups, (i.e. the groups of all males, the groups of all females and the groups of the mixture of the equal number of males and females). Doing this would collect rating results for the agents from a more generalized testing condition.

Next, the statistics methods applied should allow for between-subjects factors such as "gender", "skill level", "male group", "female group" and "mixture group".

Last, in order to stand out the LTE from STE, we may consider extending the duration of an experimental session to be longer, say 1 hour and 15 minutes, As the

LTE is a kind of less observable effect and it may require more time to be aware of by the subjects.

References

Anderson, J. R. 1991. *The adaptive nature of human categorization.* Psychological Review(98), pages 409-429.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. 2004. *An Integrated Theory of the Mind.* Psychological Review (111), pages 1036 - 1060.

Barnes, A. and Thagard, P. 1996. *Emotional decisions.* Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society, pages 426-429. Erlbaum (Accessible at http://cogsci.uwaterloo.ca/Articles/Pages/Emot.Decis.html 2006).

Belavkin, R. 2001. *The Role of Emotion in Problem Solving.* In C. Johnson (Ed.), Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing, ISBN 1-902956-19-7, pages 49–57.UK.

Belavkin, R. 2003. *On Emotion, Learning and Uncertainty:A Cognitive Modelling Approach.* PhD Dissertation. The University of Nottingham, Nottingham, UK.

Bozinovski,S. 1999. *Training a Football Playing Robot Using Emotion Based Learning Architecture.* In Proceedings of Affect Interaction Workshop.Germany.

Botelho, L.M. and Coelho, H. 1998. *Adaptive Agents: Emotion Learning.* Grounding Emotions in Adaptive Systems, pages 19-24.

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C., Kazemzadeh, A., Lee, S., Cacioppo, J., Larsen. J., Smith N. and Berntson.G. 2004. *The Affect System: What Lurks below the Surface of Feelings?* From *Feelings and Emotions[Book]*, pages

223-242. Cambridge University Express 2004.

Cañamero, D. 1997. *Modelling motivations and emotions as a basis for intelligent behavior.* ACM Press, pages 148-155. New York.

Cañamero, D. 2003. *Designing Emotions for Activity Selection in Autonomous Agents.* Emotions in Humans and Artifacts [Book], pages 115-148. The MIT Press. London. England.

Champandard, A.J. 2003. *AI Game Development: Synthetic Creatures with Learning and Reactive Behaviors [Book].* Book of New Riders Publishing.

Conati, C. and Zhou X. 2002. *Modelling Students' Emotions from Cognitive Appraisal in Educational Games.* Intelligent Tutoring Systems 2002, pages 944-954.

Damasio, A.R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain.* Penguin Books, Gosset/Putnam Press. New York.

D'Mello, S.K., Craig, S.D., Gholson, B., Franklin, S., Picard, R.W. and Graesser, A.C. 2005. *Integrating Affect Sensors in an Intelligent Tutoring System.* In Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International conference on Intelligent User Interfaces, pages 7-13. AMC Press. New York.

Ekman, P. 1992. *An Argument for Basic Emotions.* In: Stein, N. L., and Oatley, K. eds. Basic Emotions, pages 169- 200. Lawrence Erlbaum. UK.

Ekman, P. 2004. *What We Become Emotional about.* Feelings and Emotions[Book], in Manstead, A.S., Frijda, N.H., and Fischer, A. (Eds.), pages 119-135. Cambridge University Express 2004.

El-Nasr, M., Yen, J. and Ioerger, T. 2000. *FLAME-Fuzzy Logic Adaptive Model of*

Emotions. Autonomous Agents and Multi-Agent Systems 3(3), pages 219-257.

Elliott, C. 1992. *The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System.* PhD dissertation. NorthWestern University.

Estrada, C., Isen, A. and Young, M. 1997. *Positive Affect Facilitates Integration of Information and Decreases anchoring in Reasoning among Physicians.*Organizational Behavior and Human Decision Processes (72), pages 117-135.

Fernandez, R. and Picard, R.W. 2003. *Modelling Driver's Speech under Stress.* Speech Communication (40), pages 145-159.

Frijda, N.H. 1986. *The Emotions [Book].*Cambridge University Press. USA.

Frijda, N. H. and Swagerman, J. 1987. *Can Computers Feel? Theory and Design of an Emotional System.* Cognition and Emotion, pages 1235-1257.

Frijda, N.H., Manstead, A.S.R., and Bem, S. 2000. *The Influence of Emotions on Beliefs.* Emotions and Beliefs—How Feelings Influence Thoughts [Book]. Cambridge University Press.U.K.

Frijda, N.H. 2004. *Emotions and Action.* Feelings and Emotions[Book], in Manstead, A.S., Frijda, N.H., and Fischer, A. (Eds.), pages 158-173. Cambridge University Express 2004.

Gadanho, S. and Hallam, J. 2001. *Robot Learning Driven by Emotions.* Adaptive Behaviour 9: No.1.

Gadanho, S. 2003. *Learning Behaviour-Selection by Emotions and Cognition in a Multi-Goal Robot Task.* Journal of Machine Learning Research (4), pages 385-412.

Glasser, W. 1999. *Choice Theory: A New Psychology of Personal Freedom.* Harper

Paperbacks.

Gratch, J. and Marsella, S. 2003. *Fight the Way You Train: The Role and Limits of Emotions in Training for Combat.* In The Brown Journal of World Affairs 10(1), pages 63-76.

Gratch, J. and Marsella, S. 2004a. *Evaluating the modelling and use of emotion in virtual humans.* In Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems. New York.

Gratch, J. and Marsella, S. 2004b. *A Domain-Independent Framework for Modelling Emotion.* Journal of Cognitive Systems Research 2004 V5 (4), pages 269-306.

Hebb,D.O. 1949. *The organization of behaviour[Book].* Wiley. New York.

Henninger, A.E., Jones, R.M., and Chown, E. 2003. *Behaviors that Emerge from Emotion and Cognition: Implementation and Evaluation of a Symbolic-Connectionist Architecture.* AAMAS 2003, pages 321-328.

Hudlicka, E. 2004. *Beyond Cognition: Modelling Emotion in Cognitive Architectures.*In proceedings of the Sixth Internation Conference on Cognitive Modelling, pages 118-123.

Isen, A. 1993. *Positive Affect and Decision Making.* Handbook of Emotions [Book], In M.Lewis and J.M.Haviland (Eds.), pages 261-277. Cuilfor Press. New York.

Isen, A. 2004. *Some Perspectives on Positive Feelings and Emotions.* Feelings and Emotions[Book], in Manstead, A.S., Frijda, N.H., and Fischer, A. (Eds.), pages 263-281. Cambridge University Express 2004.

Izard, C. 1991.*The Psychology of Emotions*. Plenum Press. New York.

LeDoux, J. E. 1989. *Cognitive-Emotional Interactions in the Brain.* In C. Izard editor, *Development of Emotion-Cognition Relations*, pages 267-290. Lawrence Erlbaum Associates, U.K.

LeDoux, J. E. 1996. *The Emotion Brain[Book].* Simon&Schuster Press.

Lerner, J. S. and Keltner, D. 2000. *Beyond valence: Toward a model of emotion-specific influences on judgment and choice.* Cognition and Emotion(14), pages 473-493.

Lerner, J. S. and Keltner, D. 2001. *Fear, Anger, and Risk.* Journal of Personality and Social Psychology 2001(81), pages 146-159.

Lerner, J. S., and Tiedens, L. Z. 2006. *Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition.* Journal of Behavioral Decision Making (*19),* pages 115-137.

Liao, W., Zhang, W., Zhu, Z. and Ji, Q. 2005: *A Decision Theoretic Model for Stress Recognition and User Assistance.* AAAI 2005, pages 529-534.

Loewenstein, G. and Lerner J. S. 2003. *The Role of Affect in Decision Making.* In *Handbook of Affective Science.* Oxford University Press, pages 619-642.

Marinier, R., and Laird, J. 2004. *Toward a Comprehensive Computational Model of Emotions and Feelings.* International Conference on Cognitive Modelling 2004. (Accessible at http://ai.eecs.umich.edu/people/laird/recent-research.html 2006).

Mehrabian, A. 1995. *Framework for a comprehensive description and measurement of emotional states.* Genetic, Social, and General Psychology Monographs (121),

pages 339-361.

Mellers, B. 2004. *Pleasure, Utility and Choice.* Feelings and Emotions [Book], in Manstead, A.S., Frijda, N.H., and Fischer, A. (Eds.), pages 263-281. Cambridge University Express 2004.

Merriam-Webster 2007. http://www.m-w.com/dictionary/emotion . Accessible on Jan 7, 2007.

Minsky. M. 1986. *The Society of Mind [Book].* Simon&Schuster Inc.

McCarthy, J. 1995. *Making Robots Conscious of Their Mental States.* Machine Intelligence 1995(15), pages 3-17.

McCauley, L. and Franklin, Stan. 1998. *An Architecture for Emotion.* AAAI Fall Symposium Emotional and Intelligent, pages 122-128.

Neumann, U. and Narayanan, S. 2004. *Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information.* ICMI 2004, pages 205-211.

Picard, R.W. 1997. *Affective computing.* Cambridge: MIT Press.

Ortony, A., Clore, G. L., and Collins, A. 1988. *The Cognitive Structure of Emotions.* Cambridge University Press. UK.

Reilly, W.S.N. 1996. *Believable social and emotional agents.* Ph.D. Dissertation. Carnegie Mellon University.

Russell, S.J, and Norvig. P. 2003. *Artificial Intelligence—A Modern Approach(2nd Edition)[Book].* Pearson Education Inc. New Jersey.

Scheutz, M. 2002. *Agents with or without Emotions?* FLAIRS Conference 2002, pages 89-93.

Scheutz, M. 2004. *Useful Roles of Emotions in Artificial Agents: A Case Study from Artificial Life.* AAAI 2004, pages 42-48.

Silva, D.R., Siebra, C.A., Valadares, J.L., Almeida, A.L., Frery, A.C., Falcão, J.R., and Ramalho, G.L. 2001. *Synthetic Actor Model for Long-Term Computer Games.* Virtual Reality (5), pages 1–10.

Simon, H. A. 1967. *Motivational and emotional controls of cognition.* Reprinted in Models of Thought, Yale University Press 1979, pages 29-38.

Sloman, A. and Croucher, M. 1981. *Why Robots Will Have Emotions.* IJCAI 1981, pages 197-202.

Sloman, A. 1998. *Damasio, Descartes, Alarms and Meta-management.* IEEE International Conference on Systems 1998, pages 2652–2657.

Sloman, A. 2001. *Beyond Shallow Models of Emotion.* Cognitive Processing 2001(2), pages 177–198.

Sloman, A. 2004. *What are Emotion Theories about?* Invited talk at cross-disciplinary workshop on Architectures for Modelling Emotion at the AAAI Spring Symposium at Stanford University in March 2004. (Accessible at **http://homepages.feis.herts.ac.uk/~comqlc/ame04/ 2006**)

Smith, C.A., and Ellsworth, P.C. 1985. *Patterns of cognitive appraisal in emotion.* Journal of Personality and Social Psychology (48), pages 813-838.

Tanguy, E., Willis, P., Bryson, J. 2003. *A Layered Dynamic Emotion Representation for the Creation of Complex Facial Expressions.* IVA 2003, pages 101-105.

Thagard, P. 1989. *Explanatory coherence.* Behavioral and Brain Sciences, 12(3),

pages 435-502.

Thagard, P. 2001. *How to Make Decisions: Coherence, Emotion, and Practical Inference.* In Varieties of Practical Iinference. MIT Press.

Thagard, P. 2002. *Emotional Gestalts: Appraisal, Change, and the Dynamics of Affect.* Personality and Social Psychology Review 2002, pages 274–282.

Thagard, P.2003. *Why wasn't O.J. Convicted? Emotion Coherence in Legal Inference.* Cognition and Emotion, 17(3), pages 361-383.

Toda, M. 1993. *TheUrge Theory of Emotion and Cognition.* SCCS technical report, pages 93-101, Chuyko University, Japan.

Tomkins, S. 1984. *Affect theory.* In Scherer, K. R. and Ekman, P. (Eds.), Approaches to Emotion. Lawrence Erlbaum, U.K.

Velásquez, J.D. 1997. *Modelling Emotions and Other Motivations in Synthetic Agents.* AAAI 1997, pages 10-15.

Velásquez,J.D. 1998. *When Robots Weep: Emotional Memories and Decision-Making.* AAAI 1998, pages 70-75.

Ventura. R. 2000. Emotion-Based Agents. Msc. Thesis. Instituto Superior Técnico, Lisboa, Portugal.

Wang, H. 1998. *Order effects in human belief revision.* PhD dissertation. The Ohio State University.U.S.A.

Watson, J.B. 1929. *Psychology from the Standpoint of a Behaviorist (the 3rd edition).* Lippincott. Philadelphia, USA.

Wikipedia 2007. http://en.wikipedia.org/wiki/Emotion_%28disambiguation%29.

Accessible on Jan 7, 2007.

Wright,I., Sloman, A. and Beaudoin, Luc. 1996. *Towards a Design-Based Analysis of Emotional Episodes.* Philosophy Psychiatry and Psychology (3.2), pages 101-126. U.K.

Wright, I. 1997. *Emotional Agents.* PhD dissertation. University of Birmingham, U.K.

Young, P.T. 1943. *Emotion in man and animal[Book].* Wiley. New York.

Appendix

Appendix A (Typical Game Scenario)

A common fighting picture from the game Quake2 can be found in Appendix A.

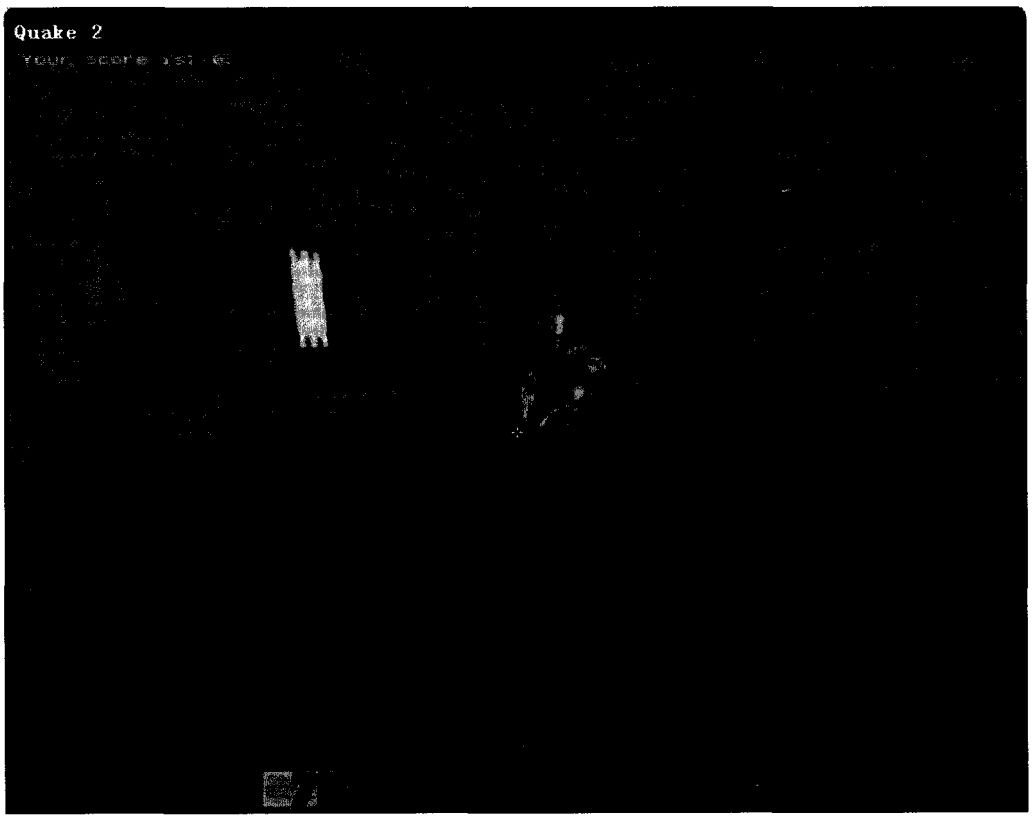

Figure A1: A Common Fighting Scenario in Quake2

Appendix B (Sample Appearance Order)

Sample Appearance Order in One Session according to the Latin Square order:

No.15:	Hum	FuA	RA	BOA	EOA
No.16:	RA	FuA	Hum	EOA	BOA
No.17:	BOA	EOA	Hum	RA	FuA
No.18:	Hum	EOA	RA	FuA	BOA

Four computers in the experiment room are named No.15, No.16, No.17 and No.18 according to the last two digits of their IPs in the local area network.

From the above displayed order, we may find the subjects in No.15 and No.18 belong to the same dyad as they are arranged to fight in the first phase, and No.16 and No.17 belong to another.

Appendix C (Instruction Script)

Introduction to Participants

Welcome to the exciting Quake2 game world. In the forthcoming experiment, you will be asked to play with each of five different types of opponents first for 9 minutes; your opponents could be either human players or computer robots. After you finish one play, you will be asked to rate the subsequent three questions (see the answer form in your hand) in 3 minutes, by recalling your memory of the last play. The above process, i.e. playing with one opponent (called "Fighting stage") and then rating the questions ("Rating stage"), will loop for five times until all of your opponents have played once with you. For this game, your only task is to earn as many points as you can. The rule is simple: When you eliminate your opponent once, you will be awarded 1 point. While you are killed once, you will lose 1 point (The score board can be triggered by pressing F1 on your keyboard or you can watch it on the top left corner of the screen during game playing). The entire experiment is estimated to take up your following 60 minutes.

Instruction Script

1. Please read the above introduction part if you feel interest to this experiment design. Also, please make sure you agree the content in the consent form, and sign it before you start playing.
2. [After each subject signs on the consent form, speak to all subjects:] Please turn

away from your facing monitor until the experiment instructor (EI) tells you to turn back; EI will connect your opponent to you during this period. Please do NOT turn back until you are told to do so. Thanks for your cooperation.

3. [Start typing the name of one agent or connect to another player's server for each subject, (Note: the appearance order of the opponents is different for every player, but it sticks to the Latin Square Algorithm). And then speak to all of the subjects:] you can turn back now. Let us start the fighting stage. Please wear your ear phone and keep silent during the game playing. [After all subjects wear their earphones, speak to all subjects:] Please start your game by pressing "Enter" in your keyboard.
4. [After 10 minutes playing, speak to all subjects:] The fighting stage ends. Please take off your earphone and start rating the questions in your form. You have 3 minutes to copy your current scores V.S. your opponents to your form, and please also mark the three questions for the last opponent below the score table; you can also make use of the rest of time for break. [After three minutes, speak to all subjects if this is not the last time to play with opponents:] Please turn yourself from your facing monitor; the EI will connect your opponent to you during this period. Please do NOT turn back until you are told to do so. Thanks for your cooperation.
5. Repeat the step 3 and 4 in the same order for five times in total.
6. (Speak to all subjects :) Time is up. Please hand in your question form. Thanks a lot for your cooperation.

Appendix D (Question Form)

Name _____

Sex ___

Age ___

You VS. OpponentX: ____:____

Please circle an appropriate number which is the best fit your thought for the question. You are NOT allowed to modify your previous rating result(s) when you are working on your current rating process, e.g. you are NOT allowed to modify your rating result for opponent 1 to 4 when you are working on rating the opponent 5.

1. Do you think your opponent was a human (1 means it's certain a computer agent, 10 means it is certain a human)?

   Opponent X: 1 2 3 4 5 6 7 8 9 10

2. What do you think your opponent's long term behaviour? Is it close to what an average person does? For example, you may expect him or her to behave more aggressively (e.g. more likely to attack you) or more conservatively (e.g. more likely to flee from you) after many battles, but what did your opponent actually do from your expectation? Please circle a number from the following ten figures to indicate the close degree your opponent to a human according to his or her long term behaviours (1 indicates you did not notice any long term behaviour performed by your opponent, 10 indicates you believe what your opponent behaved is what an average person does in the long run).

   Opponent X: 1 2 3 4 5 6 7 8 9 10

3. During a fight encounter, what do you think of your opponent's performance? Is it close to what an average person performs? Please circle a score from the following ten figures to indicate the close degree your opponent to a human according to his or her fight performance (1 indicates you do not think the opponent expressed any human-like behaviour in fight, 10 indicates you believe for sure that your opponent's fight performance is totally like what a human could do).

   Opponent X: 1 2 3 4 5 6 7 8 9 10

4. Rate your current opponent to indicate your favourite degree: (1 is you did not like to play with the opponent at all, 10 means you like playing with the opponent the most).

   Opponent X: 1 2 3 4 5 6 7 8 9 10

(Note: "X" appeared in the above can be substituted by the number 1 to 5. Since the question form for one subject has five copies, each of which will ask the subject to rate on only one type of opponent. Doing so is to prevent the subject from revising the previous rating result when working on the current one.)

Saint Mary's University
Halifax, Nova Scotia
Canada B3H 3C3

Patrick Power Library
tel 902.420.5534
fax 902.420.5561
web www.stmarys.ca

Research Ethics Board Certificate Notice

The Saint Mary's University Research Ethics Board has issued an REB certificate related to this thesis. The certificate number is: 06-009

A copy of the certificate is on file at:

Saint Mary's University, Archives
Patrick Power Library

Email:

For more information on the issuing of REB certificates, you can contact the Research Ethics Board at