Artificial Intelligence and Multi-Agent Systems Laboratory

University Politehnica of Bucharest

Towards a computational model of emotions for enhanced agent performance

PhD Thesis Proposal

PhD Student: **Mihaela-Alexandra Puică**
AI-MAS Laboratory
Computer Science and Engineering Department
University "Politehnica" of Bucharest

Supervisor: **Prof. Adina-Magda Florea**
AI-MAS Laboratory
Computer Science and Engineering Department
University "Politehnica" of Bucharest

1 Introduction

The field of Affective Computing is a recent, but fast growing research direction among computer scientists. Its beggings are hesitating and uncertain, given that affect and emotions were always considered undesirable. It is still common to speak of "emotional behaviour" as of "non-rational behaviour" and usually people treat emotions with reluctance. The first computer scientists to discover the importance of building affective computers had a rough time in convincing the rest of the community of the role that emotions play in endowing computer applications with intelligence. Today the road is not so harsh: there is a broader acceptance of the benefits affect can have on artificial intelligence and the scientific community has gain a new research path on which several people are engaged. Thus, there are several research groups established at various universities that study different aspects of affective computing: the Affective Computing research group, led by Rosalind Picard, part of MIT Media Lab at Massachusetts Institute of Technology, is entirely dedicated to emotion-related topics: better communication and understanding of affective-cognitive states through wearable sensors, emotionally-intelligent computers that empathise with users, or ethical issues; at the Computer Laboratory from the University of Cambridge, the Graphics and Interaction Group (also known as Rainbow Group) has several research projects which focus on emotionally intelligent interfaces, affective robotics or emotions in learning. Also, since 2010, IEEE has begin to edit "IEEE Transactions on Affective Computing", a journal aimed at disseminatig results of research on recognition and simulation of human emotions and related affective phenomena. Moreover, a professional association called HUMAINE was created, which is a large network of excellence dedicated to research on emotions and human-machine interaction.

The applications of Affective Computing are very diverse, starting from e-learning and ambient intelligence to serious games and helping autist peaople. Emotions enhance problem-solving skills of computer systems and make them behave in a more realistic, natural way, thus being easier for the regular user to interact with them.

AutoTutor [Graesser et al., 2004] is a learning environment that teaches students through natural language dialogue. The authors showed that AutoTutor improved learning by giving the students a deeper understanding of the subject through direct explanations. Though, Picard et al. [D'Mello et al., 2007] planned to improve it even more by adding affect recognition capabilities to the artificial tutor. It uses several recognition methods to obtain better results. By understanding student emotions, AutoTutor can respond more effective to their needs.

Mission Rehearsal Exercise (MRE) [Gratch and Marsella, 2004] is a trainig environment developed by Jonathan Gratch and Stacy Marsella which is used to train decision making skills by simulating an environment where the trainee has to deal with friendly and hostile forces, all controlled by intelligent agents. The user interacts with these agents to solve a given situation by taking the necessary decisions. The system has been evaluated by military cadets with noticeable good results.

iCat [Philips Research, 2005] is an experimentation platform for human-robot interaction. It has been developed by Philips Research in the form of a robotic cat that is able to mechanically produce facial expressions. Reasearch has focused on giving the iCat different personalities (extrovert or introvert) and test which suits best human needs. The evaluation was made with two applications: game buddy and TV assitant. The results show a preference towards a more extrovert temperament, when the iCat is more expressive, both through facial expressions and through being more talkative.

2 State of the Art

The term "Affective Computing" was first coined by Rosalind Picard in the book with the same name that she wrote in 1997 and where she defines it as "computing that relates to, arises from or deliberately influences emotion"[Picard, 2000]. The field is a new interdisciplinary field mixing computer science and engineering with multiple research disciplines, including psychology, cognitive science, neuroscience, affective science, education and sociology.

2.1 Affect

Emotions

Although emotions have been studied since 19th century, we still have little knowledge about them: how they appear, why they appear and how we should deal with them.

Emotion is the basic element in affective computing and therefore research in this area should take into consideration theories developed by human sciences regarding this concept. Because, if computers are to understand and display emotions, we must be able to model, or at least simulate, their functionality in humans. Although this term is widely used both in the scientific comunity and in daily life, there is still no clear definition of its meaning. As Fehr and Russell stated in 1984 [Fehr and Russell, 1984], "everybody knows what an emotion is, until asked to give a definition". There is no sharp boundary between what is and what isn't emotion. Thought it is of general agreement that emotions are feelings, we probably cannot state that every feeling is an emotion. There are several related concepts against which we differentiate emotions, like moods, attitudes or personality traits. These are distinguished through time constraints: emotions are "behavioral dispositions that persist for seconds or minutes" [Gratch and Marsella, 2004], moods are states with similar effects, but that last for hours or days, attitudes can stretch over months and years and personality traits "reflect relatively stable behavioral tendencies" [Gratch and Marsella, 2004] [Ekman, 1994] [Moshkina and Arkin, 2009].

Because emotions display various tipologies in what concerns their expression or their elicitation, depending on the personality of the person or on their culture, researchers have classified them in different categories. One taxonomy is represented by the basic or complex emotions. Basic emotions are claimed to be innate, and therefore universally felt and recognized by all people. Ekman supports this claim by studies that show that basic emotions are expressed using similar configurations on the face which any person is able to read, provided their environment, education or culture [Ekman, 1999]. These emotions are happiness, sadness, fear, anger, disgust and surprise. He later adds other emotions like contempt, shame or satisfaction, but the first six enumerated are equally mentioned by other researchers. For example, Plutchik created a model known as "Plutchik's wheel of emotions" [Plutchik, 2001] which consits of eight basic emotions arranged as four pairs of opposites, their combinations that give the complex emotions and their variants depending on their intensity. The eight basic emotions suggested by Plutchik are joy versus sadness, trust versus disgust, fear versus anger and surprise versus anticipation. Plutchik's model resembles the model of colors, where basic colors blend to create new colors.

The wheel of emotions is a hybrid approach in differentiating emotions: the basic-complex categories and the dimensional model. This model defines emotions according to their position in a multi-dimensional space. The most common is the two-dimensional scheme: one axis represents the valence of the emotion (pleasant-unpleasant), while the second displays the level of arousal (anxiety-boredom). This model was first developed by James Russel [Russell, 1979] [Russel, 1980] and later adapted by others like Pieter Desmet [Desmet, 2002] and Trevor van Gorp [van Gorp, 2006].

Another classification takes into consideration the way an emotion is generated: cognitively or non-cognitively. Thus, the distinction is made between primary emotions and secondary emotions [Damasio, 1994]. This sorting is based on neurological evidence of the functions of different parts of the brain. The main important in this context are the cortex and the limbic system. The cortex is the most studied - given that is much more accessible, being situated closer to the scalp. It contains the visual cortex and the auditory cortex, thus the conclusion that the cortex handles all perception. The limbic system is situated below the cortex and it is the place of attention, memory and emotion. For a long time it was believed that the two components of the brain are completely independent, but recent studies by Antonio Damasio [Damasio, 1994], Joseph LeDoux [LeDoux, 1996] and Cytowic [Cytowic, 1996] show that cognition and affect are much more related and highly influence eachother. The main difference between cortex and limbic system is that the first one is slower, but more accurate, while the latter is faster, but more imprecise (the "quick and dirty" path). Consequently, when a stimulus is perceived, it reaches first the limbic system, which outputs the primary emotion. Then, the same stimulus reaches the cortex, which processes it connecting physiological responses with cognitive appraisal of events and outputs the secondary emotion. For example, if one sees an object approaching rapidly, fear will instantly arise, which will determine quick moves to avoid the object. This kind of emotion is innate, hard-wired into the human body, and thus non-cognitively generated. That is why it is called "primary emotion" and its primary role is in survival. Rosalind Picard recognizes the importance of primary emotions stating that "it is clearly beneficial for our survival that fear can hijack our brain and cause us to jump out of the way of a rapidly approaching object before we can consciously perceive and analyze that a bus is about to hit us" [Stork, 1996]. On the other hand, secondary emotions "play an especially important role in decision making, even in decision making that appears to be purely rational" [Stork, 1996]. Marvin Minsky agrees that emotions have a crucial role in decicion making: "I don't think you can make AI without subgoals, and emotions is crucial for setting and changing subgoals" [Stork, 1996].

Theories of emotion

Saul Kassin defines emotion as "a feeling state characterized by physiological arousal, expressive behaviors and a cognitive interpretation" [Kassin, 2004]. The exressive behavior is the specific action (body, face, language) displayed by a person who experiences the emotion, action which helps the others recognize their emotion. The other two are the main components that elicit the affective state. Physiological arousal refers to the bodily changes that happen when the stimuli is present: pulse quickened, breathing rate increased, profuse sweating and so on. Cognitive interpretation deals with understanding the situation, appraising it in terms of its influence on the subject. The main question that theories of emotion try to answer is which of the two components goes first.

One of the earliest theories of emotion that is still discussed today is that of James-Lange. This was developed independently by William James in 1884 [James, 1884] and Carl Lange in 1885. This theory claims that physiological arousal comes first, and then the emotion. More precisely, the emotion is the result of bodily responses to the event (and not their cause). James' thesis was that "the bodily changes follow directly the PERCEPTION of the exciting fact, and that our feeling of the same changes as they occur IS the emotion". He explained that "we feel sorry because we cry, angry because we strike, afraid because we tremble", and not the opposite, which is what common sense would suggest. The implications of this theory is that without phisiological arousal, no emotion is felt.

This implication was one of the main criticism of the James-Lange theory. In 1927, Walter Cannon rejected this assumption after conducting a series of experiments which proved that emotion can appear without the brain knowing about the bodily responses (the connection between the two was cut off) [Cannon, 1927]. Another reason to not accept the James-Lange theory was that the same physiological arousal happens in different emotions (for example angry or excited). Philip Bard agreed with Cannon and they continued their experiments until they concluded that affect and bodily responses are independent. Therefore, they presented the theory that when an event occurs, emotion and phisiological arousal take place simultaneously and independent from one eachother. Though, they say nothing about the role of cognition in these processes. In fact, they provide no mechanisms of how emotions appear.

More insight about this was taken by two researchers, Stanley Schachter and Jerome Singer, who developed their own theory in 1962 [Schachter and Singer, 1962]. Their theory is called *Two-factor theory of emotion*, because it states that both physiological arousal and cognition take part in the apparition of emotion. More exactly, when a stimuli is present, the bodily changes appear, and the subject tries to explain these changes according to the context they are in. For example,

- → if a person is aroused and they don't know the cause, then they will search for a reason and
  - → if they find themselves in a happy situation, they could label their state as being happy;
  - → if they find themselves in an unpleasant situation, they could label their state as being angry;
- → if the person is aroused and they know the cause (for example having been injected with adrenaline) then they won't feel the need to label their state, so they won't declare to be neither happy, nor angry;
- → if the person is not aroused, then they won't feel anything, no matter the context (happy or angry situation).

Schachter and Singer proved their theory by making an experiment where subjects were injected either with adrenaline or with a saline solution), then they were either explained, misinformed or said nothing about the side-effects, and eventually they were put in a happy or angry situation. Their state was measured by two methods: self report and third-party observation of their behavior. Subjects were not told about the real purpose of the experiment. The adrenaline induces arousal, but the saline solution has no effect on the physiological parameters, so it is used just as control measure. The results confirmed the statements of their theory.

A more recent theory is the cognitive appraisal theory. Its development is attributed mainly to Magda Arnold and Richard Lazarus. They argue that when an event is triggered the first step is to cognitively assess the event in terms of each one's beliefs, goals, standards and attitudes. This process is commonly referred to as "appraisal" and is mainly concerned with how a person perceives the environment they live in. The relevance of perception in everyday life has been noticed not only by psychologists, but also by people working in other domains. The writer Anais Nin said that "The world is not as it is, but as we perceive it to be".

Lazarus specifies two types of appraisal:

- primary appraisal, which consists in assessing the event in the light of what metters for self [Smith and Kirby, 2009]; this includes:
  - motivational relevance - if the event is relevant to one's goal;
  - motivational congruence - if the event is consistent or inconsistent with one's goal;
- secondary appraisal, which consists in assessing what one can do to cope with that situation; there are two posibilities:
  - problem-focused coping - what can one do to change (external) situation
  - emotion-focused coping - what can one do to change (internal) perception

This first description of appraisal theory has been later developed in various directions by other researchers, but the basic idea of appraising and coping is still ongoing. Cognitive appraisal emphasizes the role of cognition in the apparition of emotions. It must capture various possible evaluations of the context for an accurate interpretation and it must cope with the dynamic nature of emotions to be able to quickly adapt to new changes.

Arousal and performance

The reason why affect was interlaced with computing is the evidence that emotions influence cognitive processes like perception, attention, memory, motivation and decision making. If one can identify, assess and control their emotions and others', then they are likely to be more succesful in life, even if their traditional IQ score is not among the highest. It is what Daniel Goleman is apprehending with his concept of emotional intelligence [Goleman, 1996], even if not everybody agrees with his definition [Eysenck, 1998] [Locke, 2005]. One could say that machines are built to perform lots of computations, so it suffices to increase their memory and processing power. But every system is resource bounded, and, besides, this can not be calles an intelligent machine. An agent shouldn't be able to test every single possibility, but to know which of them to exclude from the start. Emotion can provide this optimization by restricting the posibilities to be considered through attention narrowing. Moreover, we are interested in human-computer interaction, and in this domain emotions are essential. They enhance computers with a more realistic behavior and an easier, more pleasant interface. In case of an intelligent tutoring system, this leads to improved user performance: by recognizing students' emotions and being able to influence them, the artificial tutor could motivate the student better.

According to Yerkes-Dodson law [Yerkes and Dodson, 1908], the performance is higher at certain levels of arousal: if the level of arousal is low, then performance is also low; when the level of arousal is rising, the performance is also getting higher; though, when the arousal exceeds some limit, the performance starts lowering again. All these limits are specific to each person: some may perform better at lower levels of arousal, while some may perform better at higher levels.

When the level of arousal increases, the focus of attention narrows and is directed towards the stimulus, the cause of the arousal. Thus, an agent can direct its resources to what it matters most at a specific time.

2.2 Computing

When applying the theory presented above to embed emotions into computers, there are two directions research is focusing on:

- build machines that are able to automatically recognize human emotions
- build machines that are able to simulate or model affective behavior

Both directions are equally important, and the ideal system should have both of these skills.

Emotion recognition

There are several methods developed to recognize human emotions, depending on the forms of the sentic modulation - what Rosalind Picard defines as the influence of emotion on bodily expression [Picard, 2000]. People have different ways of showing their emotions, willingly or unwillingly, more or less apparent to others. Usually the more apparent forms are easier to control by the subject, while the less apparent forms are harder to control. In the first category there are *facial expressions, voice intonation, gestures, posture, pupilary dilation* . The second category includes *respiration, heart rate, temperature, perspiration, blood presure, muscle action potentials* [Picard, 2000].

Facial expressions are among the most important form of sentic modulation for the recognition of emotions by humans. In a conversation, the first thing that one observes is the other's face. And if we are to believe Ekman, we can recognize different emotions from the facial expression because each emotion has its own specific configuration of the face muscles [Ekman, 1994]. There are of course some exceptions, when subjects control their muscles to hide their true feelings, because the situation requieres it (either in a business context or culture determined). Therefore, face recognition is one of the most studied research topic. And the most used method for this is image recognition. Using various algorithms of pattern matching, usually supervised learning, the software can predict the most probable emotion experienced by the user.

Along the facial expressions, there are also the posture and gesture sentic modulation forms that complete the visual of the subject. For these, the same machine learning techniques are used. But if in the first case only muscles were in the target, now the whole body and arms position is studied.

Voice intonation though is different. It spans the audio part of affect recognition. It is also a very important step in recognition of emotions by humans and other animals. For example, if your pet dog upsets you, you will probabily yell at it in anger. The dog won't understand your words, but will realize that you are mad. This is true in all the situations where verbal communication is the case (wheather the listener is a person or an animal).

The second category are not so helping for humans in recognizing affect, but they are much more reliable for machines. The reason is that in most cases, the parameters mentioned can be measured with accurate precision using different tools designed to fulfill this exact purpose. For example, to measure perspiration (the electrodermal response of the skin), a GSR (Galvanic Skin Responde) tool is used.

Emotion simulation

There are several theoretical models of emotions, but these are not very clearly specified. To implement an affective agent, more detailed models are requiered. Therefore, computational models of emotions have been developed. There are plenty of them and we only discuss a small part in this paper, but most of them use a cognitive approach to simulate emotions.

The most proeminent model of the 20th century is that developed by Andrew Ortony, Gerald Clore and Allan Collins in 1988, generally known as the *OCC model* [Ortony et al., 1988]. They claim emotions are a result of cognitively appraising an event or a situation in the light of one's goals, standards and attitudes. Thus, there are three aspects of the world that contribute to experiencing emotions:

- events and their consequences on itself and on others - these factors can generate pleased or displeased emoions;
- agents and their actions - these factors can generate approval or disapproval emotions;
- objects and their aspects - these factors can generate liking or disliking emotions.

Overall, Ortony, Clore and Collins differentiate among 28 types of emotions (22 concrete and 6 abstract). Their model has widely influenced other systems, event though it hasn't dominate psychological community [Gratch and Marsella, 2004], and it still presents interest to AI researchers.

Affective Reasoner [Elliot, 1992] is a project developed by Clark Elliot for his PhD in 1991 which is based on the OCC model. He implemented a platform for reasoning about emotions aimed at testing psychological theories of emotions. The platform is a multi-agent world in which agents are endowed with affective states and are able to reason about other agent's emotions and emotion-induced actions. The project uses a set of 24 emotion types based on that of [Ortony et al., 1988].

In 1996, Scott Reilly finished his PhD thesis which presents the OZ project [Reilly, 1996]. In this project, the focus is on interactive drama and specifically on interactive characters that have to perform believable social agents in the provided setting. Interactive drama is a system where a human user plays the part of a character in an interactive story-based simulation. Reilly provides a general framework that allows building emotional agents. It incorporates the Tok agent architecture which, in its turn, embeds the Em emotion architecure. To test his work, Reilly developed seven believable social and emotional agents which act in three different simulations.

Another approach is to have BDI agents include emotions in their architecture. BDI is known to be a suitable and wide-accepted model for intelligent agents [Rao and Georgeff, 1991]. It is based on the practical reasoning that humans usually perform [Bratman et al., 1988]. That is, reasoning directed toward action, not the classical logical formalisms. Practical reasoning implies deliberation (the agents decides what are the goals it wants to obtain) and means-end reasoning (the agent decides the actions it has to perform to achieve its goals). By mixing emotions in this architecture, the agents beliefs, desires and intentions are influenced by emotions. Several researchers embraced this approach [Florea and Kalisz, 2004] (BDE architecture), [Pereira et al., 2005], [Jiang et al., 2007] (EBDI architecture), but work is not over yet and it is still an interesting research direction, suitable for future development.

Another model that uses OCC, but accounts for the dynamic nature of emotions is PETEEI [Seif El-Nasr et al., 1999]. The agent that follows this architecture learns form its experience through four types of learning:

- learning about events - the agent learns about the probability of an event; using Q-learning and computing a desirability measure, the agent is likely to experience a certain emotion;
- learning about the user - the agent learns typical sequences of actions (patterns) performed by the user and is able to later anticipate what is their next most probable action
- learning about pleasing and displeasing actions - the agent learns what actions are pleasant or unpleasant for the user

Jonathan Gratch and Stacy Marsella goes beyond OCC. They use Lazarus model of appraisal and coping to develop their EMA architecture (EMotion and Adaptation) [Gratch and Marsella, 2004]. The EMA agent is evaluating the personal-environment relation in terms of beliefs, desires, plans and intentions, creating specific appraisal frames. Based on the features extracted, the appraisal frames are mapped into individual emotions, which are then integrated into an emotional state. This further influence the decision to be taken in order to cope with the current situation.

Some researchers ([Adam et al., 2009], [Meyer, 2006], [Dastani and Meyer, 2006] or [Steunebrink, 2010]) take another approach to modeling emotions. They consider modal logic to be appropriate for representing agents' mental attitudes and to reason about them. Hence, they create a logical formalization of emotions using what is called Logic of Emotions. Althouh it seems contradictory to associate logic with emotions, the aforementioned researchers prove that this approach is reasonable and feasible. The psychiological part of their work is founded on the OCC theory of emotions.

Thus, Adam, Herzig and Longin [Adam et al., 2009] give definitions for 20 emotions of the 28 presented in the OCC model, having the aim of formalizing it as faithful as possible. But, even thought they claim to focus on the triggering conditions of the emotions, Steunebrink reproaches them that they mix emotion triggering with emotion expression [Steunebrink, 2010].

In his PhD thesis, Bas Steunebrink formalizes all the 28 emotions in the OCC model, making a clear difference between emotion elicitation, emotion expression and emotion regulation. His work takes inspiration from the work of Jonh-Jules Meyer and Mehdi Dastani [Meyer, 2006] [Dastani and Meyer, 2006], who formalize only four emotion types (happiness, sadness, anger and fear) based on the psychological work of Keith Oatley and Jennifer Jenkins [Oatley and Jenkins, 1996].

3 PhD Thesis

3.1 Research Approach

This research is directed towards emotion synthesis and affective behavior simulaton in an artifical intelligent agent. The goal of this thesis is to build a computational model of emotion which accounts for better decision making and better resource usage. Such a model should allow the agent to perform similar to or even better than a human user in a given situation.

The hyphotesis to verify is that emotions improve decision making. By narrowing the posibilities needed to be verified, emotions allow the agent to decide faster, and thus reach a solution more rapidly. This is like having the agent rely on its "guts" or "feelings". It might not seem like a good idea, but in critical situation where time is essential, emotions help taking a decision.

The other focus of this research is resource usage. When dealing with common situations, each person has a limit that will usually not pass. For example, the speed at which they can run or the energy that they put when working on a project. But if they are confronted with critical events, these limits can be exceeded: an immediate danger will trigger fear, which will make the person run faster than usual; the approach of the deadline for the project will trigger anxiety, which will give the person the energy to work harder than usual. These effects are tested on people and the aim is to implement it on artificial entities as well. Agents, like humans, are resource-bounded. If they would use their resources at their entire capacity all the time, they would be overwhelmed and maybe crash. Therefore, it is wise to use only a chunk as normal behavior, but when situation requires it, one should be able to employ more resources for solving the situation.

To implement this, the agent will mantain a data structure with resources and paths between these resources. In a common situation (normal behavior), the agent will have access to certain nodes in that structure, only some of the paths being open. When an emotion is triggered, other paths will open and thus the agent will be able to access and use other resources. Depending on the intensity of the emotion, the nodes will be accessible or not according to the Yerkes-Dodson law: too little or too much emotion will close the paths, only the optimum level of arousal will give access to that resources. The goal is to obtains a faithfull simulation of the human processes of arousal and performance.

Conceptually, resources can be physical or psychological. For humans, physical resources can refer to running speed, weights handled or speed of reaction. For computers, they are processing speed, memory, capacity. For agents, this type of resources can refer to any of the ones mentioned above, depending on the goal that it is built for. The psychological type of resources are motivational ones: they help people work harder, put more energy to achieve their goals. An agent could also intesify its efforts, for example by changing its strategy from single-minded commiment to open-minded commitment in a BDI architecture. Altogether, it is important to define a set of (initial) resources that the agent can use when acting to achieve its goal.

Another aspect to decide upon is the set of emotions that must be taken into consideration. What the studies have shown is that each situation claims for different set of emotion. For example, it is less probable that a user would feel disgust in the process of learning, even though disgust is consider one of the basic emotions. Therefore, each scenario will involve choosing a set of relevant emotions.

And yet another goal of this research is to develop agents with different temperaments. The agents will receive as input their dispositions and will behave accordingly. There are four types of tempers that are well-known and accepted: sanguine, choleric, phlegmatic and melancholic. They are best defined by putting them on a 2D axis system: psychologically stable - instable and introverted - extroverted. Thus:

- the **sanguine** is extroverted and psychologically stable;
- the **choleric** is extroverted and psychologically instable;
- the **phlegmatic** is introverted and psychologically stable;
- the **melancholic** is introverted and psychologically instable.

The approach to validating the model consists in three main components: the system (may be a game or any other scenario), an agent and a human user. These will interact in turns: the agent with the system, the user with the system, and then all three together - or better said, the human will act in the system, being helped by the agent. The system will provide an environment where the user can interact and also a methodology to measure the responses of each user, as well as the amount of emotion felt by them.

3.2 Research Goals

The title of my PhD thesis is

"Towards a computational model of emotions for enhanced agent performance"

The main goals are the following:

- Develop a model of agent affective behavior which simulates emotion triggering and the effect on agent behavior; the aim is to create an intelligent agent which can make better, faster decisions and which is able to use efficiently its resources. Previous related work suggests that practical reasoning is the most appropriate approach for building cognitive agents, because of its orientation towards action. BDI architectures are a common tool to implement this type of reasoning. Therefore, we find that there is a high probabilty that the agent is to behave intelligently and efficiently if it possesses mental attitudes and emotions on which it can rely to choose the actions that will help it cope with the situation. Moreover, the model should include the resources that the agent has available in different emotional states, such that it can use them wisely.

- Imagine some scenarios where the user (human or artificial) can interact with some situations in an environment. The scenarios must allow the user to show their performance in terms of decision making and resource usage. The context could be for example a game in which the player lives at a farm and takes care of it. During the game, it has to deal with critical situations - fire or burglers. In reaching the solution, the players have to make some choices which are affected by their mental and emotional states, these being thus reflected in their behavior. Another scenario could situate the player on an unknown planet, which they have to explore. The users should be able to reason about new discoveries, with the final aim of deciding if and how the new environment could become inhabited.

- Write an application that tests the agent affective behavior model developed as the first goal under at least one of the scenarios described by the second goal. The application will allow both the human and the agent interact with the environment and it will provide a means for evaluating three performances: that of the human, that of the artificial agent and that of both of them. The hypothesis is that comparative charts will show comparable results when the agent and the human are acting individually, and better perfomance when the user interacts with the agent to achieve the goals.

- Develop a method to evaluate the "quantity" of emotion that the agent is experiencing. We saw that too much, but also too little emotion impairs decision making, but how can we quantify "too little" or "too much"? This is related to the intensity of the emotion experienced or with the number of emotions experienced simultaneously. If we take fear as an example, fear is defined by Aristotle as "pain arising from the anticipation of evil". Hence, when a danger is anticipated, fear is experienced. If the intensity of this fear is too high, the agent becomes paralyzed, it can't move or think how to act in that situation. At the same time, the agent can experience (unpleasant) surprise, disgust and/or anger, which could lead to a lack of reasoning capabilities. The developed model will be tested with different personality types that will show the behavior when the agent experiences "too little", "too much" or the "optimum level" of emotion.

Given the considerations mentioned above, an option for the implementation of the agents would be to use some existing BDI agent implementation (be it JADEX, AgentSpeak, 2APL or other). Because we have no knowledge of frameworks that allow agents to have emotions, we will probably have to extend the existing framework to easily give emotions to the agent.

3.3 Research Reports

The intermediary results of the thesis will be presented in two research reports. These are:

- *Affective techniques for emotion simulation* - This report will present current state of the art in recognizing and synthesizing emotions by an intelligent agent, presenting current trends, characteristics and classifications in the field, along with advantages and disadvantages of current models. The focus will be on emotion simulation, as this is the main theme of this research.

- *Towards a model of agent affective behavior* - This report will present the proposed model for the agent affective behavior and some scenarios in which the model can be interpreted depending on the specific context.

3.4 Acknowledgments

This work is supported by University Politehnica of Bucharest and POSDRU Grant No. POSDRU/107/1.5/S/76813.

References

[Adam et al., 2009] Adam, C., Herzig, A., and Longin, D. (2009). A logical formalization of the occ theory of emotions. *Synthese*, 168(2):201–248.

[Bratman et al., 1988] Bratman, M., Israel, F., and Pollack, M. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence.*

[Cannon, 1927] Cannon, W. (1927). *Bodily Changes in Pain, Hunger, Fear and Rage - an account of recent researches into the function of emotional excitement.* D. Appleton and Company.

[Cytowic, 1996] Cytowic, R. (1996). *The Neurological Side of Neuropsychology.* MIT Press.

[Damasio, 1994] Damasio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain.* Gosset/Putnam Press.

[Dastani and Meyer, 2006] Dastani, M. and Meyer, J.-J. C. (2006). Programming agents with emotions. In *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI06)*, page 215219.

[Desmet, 2002] Desmet, P. (2002). *Designing Emotions.* PhD thesis, Delft University of Technology.

[D'Mello et al., 2007] D'Mello, S., Graesser, A., and Picard, R. (2007). Toward an affect-sensitive autotutor. *Intelligent Education System.*

[Ekman, 1994] Ekman, P. (1994). *The Nature of Emotion: Fundamental Questions*, chapter Moods, Emotions and Traits. Series in Affective Science. Oxford University Press.

[Ekman, 1999] Ekman, P. (1999). *The Handbook of Cognition and Emotion*, chapter Basic Emotions, pages 45–60. John Wiley & Sons Ltd.

[Elliot, 1992] Elliot, C. (1992). *The Affective Reasoner: A process model of emotions in a multi-agent system.* PhD thesis, Northwestern University.

[Eysenck, 1998] Eysenck, H. (1998). *Intelligence: A New Look.* Transaction Publishers.

[Fehr and Russell, 1984] Fehr, B. and Russell, J. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, 113(3):464–486.

[Florea and Kalisz, 2004] Florea, A. M. and Kalisz, E. (2004). Behavior anticipation based on beliefs, desires and emotions. *International Journal of Computing Anticipatory Systems*, 14:37–47.

[Goleman, 1996] Goleman, D. (1996). *Emotional Intelligence: Why It Can Matter More Than IQ.* Bloomsbury Publishing PLC.

[Graesser et al., 2004] Graesser, A., Lu, S., Jackson, G., Mitchell, H., Ventura, M., Olney, A., and Louwerse, M. (2004). Autotutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments &Computers.*

[Gratch and Marsella, 2004] Gratch, J. and Marsella, S. (2004). A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research.*

[James, 1884] James, W. (1884). What is an emotion? *Mind*, (9):188–205.

[Jiang et al., 2007] Jiang, H., Vidal, J., and Huhns, M. (2007). EBDI: An architecture for emotional agents. In *Proceedings of the 6th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-07)*, pages 38–40.

[Kassin, 2004] Kassin, S. (2004). *Psychology.* Prentice Hall, 4 edition.

[LeDoux, 1996] LeDoux, J. (1996). *The Emotional Brain.* Simon&Schuster.

[Locke, 2005] Locke, E. (2005). Why emotional intelligence is an invalid concept. *Journal of Organizational Behavior*, 26(4):425–431.

[Meyer, 2006] Meyer, J.-J. C. (2006). Reasoning about emotional agents. *International Journal of Intelligent Systems*, 21(6):601619.

[Moshkina and Arkin, 2009] Moshkina, L. and Arkin, R. (2009). Beyond humanoid emotions: Incorporating traits, attitudes and moods. In *Proceedings of 2009 IEEE Workshop on Current Challenges and Future Perspectives of Emotional Humanoid Robotics, IEEE International Conference on Robotics and Automation(ICRA 2009).*

[Oatley and Jenkins, 1996] Oatley, K. and Jenkins, J. (1996). *Understanding Emotions.* Blackwell Publishing, Oxford, UK.

[Ortony et al., 1988] Ortony, A., Clore, G., and Collins, A. (1988). *The cognitive structure of emotions.* Cambridge University Press.

[Pereira et al., 2005] Pereira, D., Oliveira, E., Moreira, N., and Sarmento, L. (2005). Towards an architecture for emotional BDI agents. In *IEEE Proceedings of the 12th Portuguese Conference on Artificial Intelligence (EPIA'05)*, pages 40–46.

[Philips Research, 2005] Philips Research (2005). iCat Research Community. http://www.hitech-projects.com/icat/index.php.

[Picard, 2000] Picard, R. (2000). *Affective Computing.* MIT Press.

[Plutchik, 2001] Plutchik, R. (2001). The nature of emotions. *American Scientist.*

[Rao and Georgeff, 1991] Rao, A. and Georgeff, M. (1991). Modeling rational agents within a BDI-architecture. In *Proceedings of Knowledge Representation and Reasoning(KR&R-91)*, pages 473–484.

[Reilly, 1996] Reilly, S. (1996). *Believable Social and Emotional Agents.* PhD thesis, Carnegie Mellon University.

[Russel, 1980] Russel, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

[Russell, 1979] Russell, J. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology*, 37(3):345–356.

[Schachter and Singer, 1962] Schachter, S. and Singer, J. (1962). Cognitive, social and psychological determinants of emotional state. *Psychological Review*, 69(5):379–399.

[Seif El-Nasr et al., 1999] Seif El-Nasr, M., Ioerger, T., and Yen, J. (1999). PETEEI: A pet with evolving emotional intelligence. Technical report, Texas A&M University.

[Smith and Kirby, 2009] Smith, C. and Kirby, L. (2009). Putting appraisal in context: Toward a relational model of appraisal and emotion. *Cognition and Emotion*, 23(7):1352–1372.

[Steunebrink, 2010] Steunebrink, B. (2010). *The logical structure of emotions*. PhD thesis, Utrecht University.

[Stork, 1996] Stork, D., editor (1996). *HAL's legacy: 2001's Computer as Dream and Reality*. MIT Press.

[van Gorp, 2006] van Gorp, T. (2006). Emotion, arousal, attention and flow: Chaining emotional states to improve human-computer interaction. Master's thesis, University of Calgary, Faculty of Environmental Design.

[Yerkes and Dodson, 1908] Yerkes, R. and Dodson, J. (1908). The relationship of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology*, (18):459–482.