USING ARTIFICIAL INTELLIGENCE TO ASSIST PSYCHOLOGICAL TESTING

Karthi Selvarajah and Debbie Richards
Computing Department
Macquarie University
richards@ics.mq.edu.au

ABSTRACT
Balancing ecological validity and control in psychological testing is a challenge. We have explored the use of Interactive Virtual Environment Technology to create an environment for psychological testing. Specifically we have created a Cocktail Party World to provide experimental control in the study of the phenomena of ostracism in its various forms. To address ecological validity we have created an emotion-based agent version of the Cocktail Party World to add realism to the experiment. In our experiments the results were statistically highly significant for the punitive ostracism condition revealing that the use of agents whose emotional response was depicted on their faces increased the potency of the situation.

KEY WORDS
Emotion-based agent architecture, ostracism, psychological testing, interactive virtual environment

1. Introduction

Psychology researchers have confronted the serious methodological problem of trying to find the correct balance between Ecological Validity and Experimental Control [1]. Experimental Control is a prerequisite of any experiment [1, 2]. Ecological Validity refers to the extent to which the condition simulated in the laboratory reflects real life conditions. Psychology researchers usually struggle between the two methodological problems, as concentrating on one compromises the other.

Interactive Virtual Environment Technology (IVET) increases the realism of the experience, by immersing the participant in the environment. It thus increases the ecological validity of the experiment, as the participant can now interact and behave in the world as if in physical reality. As such it increases the power of experimental research [1, 2].

Furthermore, IVET gives researchers the ability to manipulate aspects that are physically impossible in the real world, such as the size of certain objects, the distance between objects or the flow of events in a particular situation [1, 3]. One aspect, which is particularly powerful when studying social interaction, is the power to separate the non-verbal behaviour that one participant displays from what the other perceives [1].

In the study we conducted involving the Cocktail Party World paradigm, IVET allowed creation of an environment with twenty guests to interact with the participant in the virtual cocktail party. Hiring twenty actors and a suitable venue would be costly and control over exact facial expressions and body movements of the actors would be extremely difficult. Other biases such as appearance, gender, ethnicity, etc can also be minimized in a simulated world.

The studies described here are specifically concerned with the use of IVET and techniques from AI to study the powerful social, but little studied, phenomenon of Ostracism. The study of Ostracism has become more relevant in the past decade, especially after the tragic teenage high school shooting at Columbine High school. The teenage shooters attributed their behaviour to feelings of isolation and exclusion. Williams, on whose behalf this study was conducted, defines Ostracism as the "phenomenon of being excluded or ignored from the social interactions of those around you" [4].

For the study to be valid and the human participant to interact with the avatars in a manner that is similar to interacting with other humans, we need to model realistic looking avatars that display some form of human-like intelligence. Within an interactive virtual environment (IVE), it is not actual intelligence that is significant, rather it is the perceived intelligence [3]. Bailenson [3] found that humans respond to virtual avatars, even those avatars that lack photo-realism, in ways very similar to their response to other humans. However in early constructions, avatars were seen to be exhibiting stiff behaviour and "lacking emotions" [5]. So while appearance does not need to be realistic, it is evident that realistic non-verbal behaviour, including facial expressions and body gesture, significantly improves the realism of an avatar [3, 5].

In our study we have developed two separate worlds, one containing avatars (graphical representations of human beings following scripted scenarios) and agents (ie avatars with agent capabilities) in order to determine firstly whether the artificially intelligent software agents caused the human participant to perceive the Agent Cocktail Party World as more realistic than the Avatar Cocktail Party World. Secondly, as a natural follow on to show that psychological experimentation and testing could be enhanced by artificial intelligence. For the avatars and agents, we have chosen to model emotions through the use of facial expressions to increase their perceived realism. The difference being that the avatars display these expressions when programmed to do so, the agents display these expressions autonomously as they switch between different emotional states based on their environment.

1.1 The Role of Emotion in Intelligent Behaviour

There exists no standard definition for what constitutes an emotion or emotional phenomena. Ekman et al [6] identify three aspects of emotions: Emotions are feelings; Emotions are accompanied by physiological responses and produce a readiness to act in a certain way.

Ekman asserts that logically, psychologists can expect that facial expression should be related to changes in the physiology of emotion [7]. Rosenberg & Ekman [8] discovered that facial expressions are universal signs of emotion, and as such, are related to the subjective experience of emotions. More specifically, a facial expression is highly informative, but is not necessarily intended to be by the person making them [9].

Emotions can be used for activity selection in autonomously motivated agents [10, 11]. Emotions are also emergent behaviours in response to an agent's interaction with its environment [10]. In this sense emotions can be used like a feedback mechanism for a creature to evaluate its own performance in its environment [12]. Furthermore an agent can use emotions as a mechanism to form or prioritize goals [12].

Researchers have found that agents, who exhibited some emotional-type awareness, were treated like real human beings [3]. In a study conducted to measure interpersonal distance between a human participant and a human avatar, the human participant maintained a much larger interpersonal distance, similar to a real human, when the avatar maintained mutual gaze or smiled [3]. This suggests that a more realistic human avatar interaction response can be derived if the avatar displayed some form of emotional awareness, as emotions are a distinct characteristic of human-like intelligence [10, 13]. In his work in animation at Disney Bates [14] discovered that one aspect, which is essential for modeling realistic human behaviour, is to have appropriately timed and clearly expressed emotions in characters.

Since emotions affect not only the way human beings interact with their environment but also their abilities to process information and formulate plans, we needed an agent architecture that will model the complex interactions taking place. The approach we have taken builds on from the work of Ekman using facial expression to display emotion and the emotion-based agent architecture of Oliveira & Sarmento [12].

We next introduce the paradigm, first conceived by Williams, that we developed for testing of Ostracism. We then describe in Section 3 how emotions were modeled. Section 4 briefly describes the experiments and results of our initial studies. Our conclusions are given in Section 5.

2. Ostracism and The Cocktail Party World Paradigm

We've all experienced attending a social event where we did not know anyone. When we arrive everyone is busy chatting with others. We are anxious to fit in and be accepted. The Cocktail Party World paradigm was created to simulate a situation where the participant could be subjected to Ostracism.

The essential characteristic of Ostracism is that it involves some form of ignoring of the target and usually carries an explicit or implicit exclusion of the target individual or group [15, 16]. Furthermore, Williams [4] has identified different types of Ostracism. **Punitive Ostracism** is sometimes likened to a form of punishment. This is where the group deliberately ostracizes an individual by behaviours such as giving someone the silent treatment or excommunication. **Defensive Ostracism** is when a group or individual ostracize a target out of fear. Targets of both Defensive and Punitive Ostracism are able to benefit from the knowledge that the source has gone out of their way to exclude them. Targets of **Oblivious Ostracism** feel that they are too insignificant to be noticed. As a result, targets often feel "inconsequential, invisible and as if they do not matter" [4] as the source doesn't even realize they are ostracizing the target. This is the most potent form of ostracism.

The Cocktail Party World is a simulation of a cocktail party with many guests in a room, drinking and non-verbally interacting with each other. Guests respond to the experimental subject depending on whether the participant has been assigned to one of the three ostracism options or the inclusion (control) option. For example, in the Oblivious Ostracism condition, the participant, through the use of a head mounted display (HMD), is placed in the room where the guests do not react to the participant at all, leaving the participant feeling the invisibility that is characteristic of Oblivious Ostracism (Figure 1).

Figure 1: Screenshot of Avatar-Oblivious Experiment

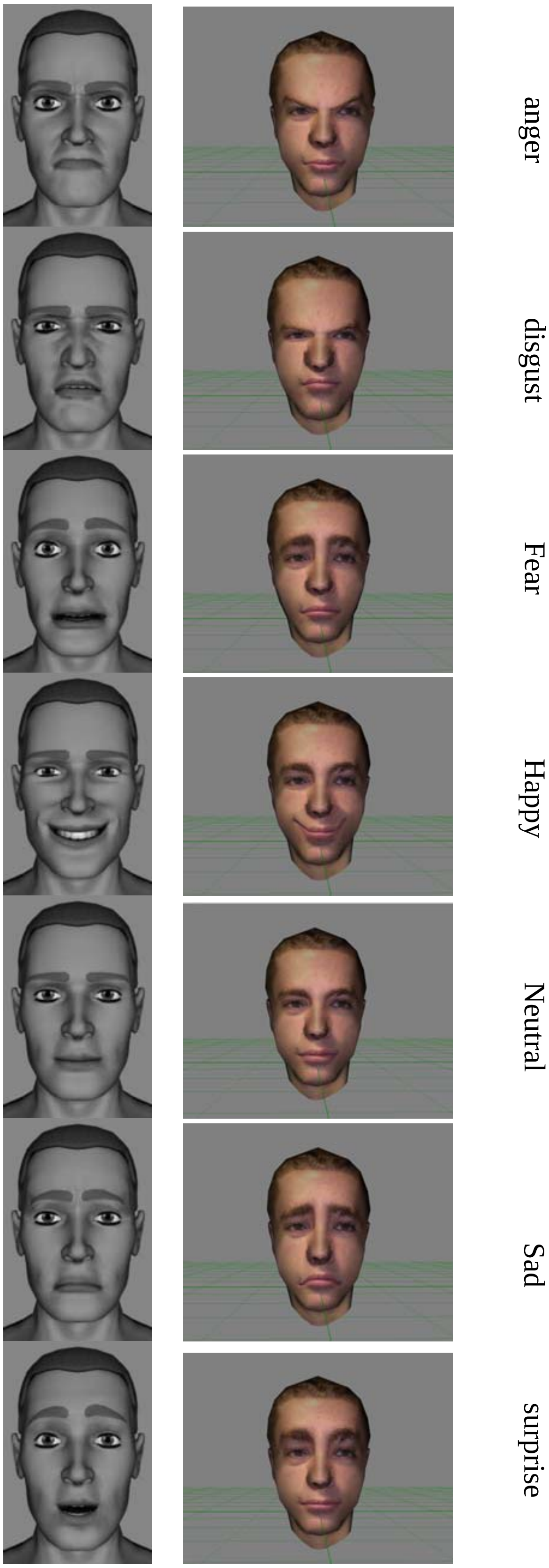

Figure 2: Paul Ekman's (left) and the Cocktail Party World's (right) faces depicting the seven basic emotions

3. Modelling Emotions

As stated we chose to use the facial expressions proposed by Ekman to display emotion and the architecture of Oliveira and Sarmento as the basis of our emotion-based architecture. Figure 2 shows Ekman's faces on the left-hand side with ours on the right. Ekman identified a set of seven globally identifiable emotions: anger, disgust, fear, happy, sad, neutral, sad and surprise. Note that some differences were unavoidable. We have modeled the shape of the mouth, eyes, eyebrows, wrinkles, etc but chose not to display teeth as the software we were using produced silly looking faces when teeth were shown.

The Oliveira and Sarmento architecture is based on four main factors that affect emotional phenomena. These factors are used when analyzing, evaluating or assigning particular emotional states to specific agents. Using these definitions emotions can be categorized into three distinct groups: specific emotions, moods and emotional dispositions.

Specific emotions are clearly identifiable emotional phenomena. They have well defined objects, have a strong intensity and last for a period of a few seconds to a few minutes. The agent is clearly conscious of the existence of specific emotions. Examples of specific emotions include anger and happiness. Oliveira & Sarmento describe two other emotional phenomena: moods and emotional disposition. This project was only concerned with specific emotion as we were trying to identify the reaction to short term Ostracism, and therefore we can only test emotions which have a short duration and a clearly identifiable antecedent so that a reaction is guaranteed.

We have adopted the notion of an Emotional Evaluation Function (EEF), formula (1) from Olivera and Saramento [12] to form the Basic Emotional Structure of our architecture. EEF is used for the process of emotional elicitation. The process involves evaluating the chances of achieving a certain goal while taking into consideration the state of the environment as well as the internal state of the agent. The term 'goal' is an explicit formulation about desirable future states that the agent seeks to achieve, as well as the implicit states that must be reached or maintained in the process.

The Emotional Evaluation Function (EEF) of a given goal 'G' is a function that receives as input: a vector 'E' containing the information perceived by the agent of the outside environment. A Vector 'I' containing information about the internal state of the agent.

It return s a scalar value 'V' that reflects the chances of the agent achieving the goal 'G'.

$$V = EEF_G(\langle E\rangle, \langle I\rangle) \qquad (1)$$

The Cocktail Party World has similar notions of the environment, the goals and the agent. Within the Cocktail Party World the environment is defined to be the hotel/conference room where the cocktail party is being held. More precisely, the twenty agents in the room constitute the environment. This is because otherwise the room contains only inanimate objects, whose state cannot be determined.

Initially each agent is assigned one random emotion. This is because at a real cocktail party, usually each guest arrives at the party with an emotional state that is (generally) not related to the emotional state of any other guest at the party.

Each agent has a very simple specific goal, which is to make a friend in the room achieved by interacting with other agents in the room, more specifically those in their vicinity. The vicinity region is approximately a circular region of radius half a meter in room dimensions. The vicinities of agents overlap when the agents come together and this further promotes the two-way interaction between agents. The varying degrees of success in the agents attempt to achieve the goal, is depicted by the emotional expressions on the face of the agent.

An agent leaves its mark on the environment by injecting its current emotional state into its vicinity. The agent receives feedback from the environment in terms of a new emotional state, which is calculated, based on the collective emotions of the other agents in its vicinity. Just as the mood of a group/party will affect each individual.

The algorithm we have developed assigns a weight to each of Ekman's seven emotions. The weights are initially assigned to each emotion by classifying the emotion as either positive or negative. Once the seven basic emotions were classified into positive and negative emotions (See Table 1), they were then ranked in order of intensity. Within the positive emotions, the happy emotion was thought to be more intense than the surprise emotion and was assigned a higher weight. Similarly the negative emotions were also ranked with sad being the least intense and disgust being the most intense. The weight of each emotion, positive or negative, is assigned so that both the positive and negative emotions in total will have the same value of 10. The emotions are ranked in this manner because it is still unknown how each individual emotion affects or interacts with the other emotions. The weights assigned to each emotion are shown in Table 1 and can be adjusted according to the experimental results in future studies.

	Emotion	Assigned Weight
Positive Emotions	Happiness	6
	Surprise	4
Neither	Neutral	0
Negative Emotions	Sadness	-1
	Fear	-2
	Anger	-3
	Disgust	-4

Table 1: The emotions and associated weights

The algorithm works as follows. Each agent is randomly assigned an emotion on initialization. Each agent is aware of all of the other agents in its vicinity. Each agent's current emotional state is affected by the emotional states of the agents in its vicinity. Furthermore, if the human participant walks into an agent's vicinity the participant's emotional state, which is gathered by asking the participant to choose how they feel at that particular time using the game pad, will affect the emotional state of the agent.

The revised EEV algorithm collects information from the agent's environment consisting of the other agents in its vicinity and determines what emotions the agent is currently feeling based on the agent's success in achieving its goal, given its capabilities in its environment. The algorithm calculates the aggregate emotion of all the agents in each vicinity including the participant. It then calculates the average of the aggregate emotion. This average is added to the current emotional state of each agent in the vicinity. Thus, the new emotional state of the agent is dependent on both its old emotional state as well the emotions of the other agents in its vicinity. As specified by the psychological experimenter, the group size, location and membership remain the same for all experiments.

For example, in our study there are three groups of 6 agents in the room that are clearly visible to the participant. Thus, there are three defined vicinities – one for each group. Figure 2 demonstrates how the next emotional state of Agent 1 in Group 1(Grp1Ag1) can be calculated. Equation (2) shows the emotional states of each agent in the vicinity being added together. To make the agents receptive to the participant, the algorithm, incorporates the participant's current emotional state (Participant_G) as part of the aggregate total for a particular vicinity. When the participant enters into the agent's vicinity, they are asked to choose which of the seven emotions best describes their current mood. Their choice is included into formula (2) The average is then computed (formula 3) and then added to the agent's current emotional state to determine the next emotion (formula 4). Changing formula 4 to an average would be better algorithmically but we found that the emotions and thus faces changed as we discovered in prepilot testing. To get around this we have to adjust the weights a bit, so that the participants do notice that the agents' faces are changing, which is the major difference between the Agent and the Avatar worlds. To avoid emotions being trapped at each minimum or maximum point, there is a lower and upper bound of 15 and -15. When these limits are reached, that agent's emotion value is set to neutral.

Aggregate_Grp1 = A + B + C + D + E + F + G----------(2)
Where
A = the Current Emotion of Agent 1 in Group 1
B = the Current Emotion of Agent 2 in Group 1
C = the Current Emotion of Agent 3 in Group 1
D = the Current Emotion of Agent 4 in Group 1
E = the Current Emotion of Agent 5 in Group 1
F = the Current Emotion of Agent 6 in Group 1
G = the Current Emotion of the participant
Average_Grp1 = Aggregate_Grp1 / 7-----------------------(3)
Grp1Ag1_next_emotion = Grp1Ag1 + Average_Grp1---(4)

Figure 3: Example of the Agent Cocktail Party World's adapted Emotional Elicitation Function

The emotional state is updated periodically every 30 seconds. Although most specific emotions take effect for a period of a few seconds to minutes, we update the emotions every 30 seconds because the agents are using a graphical means to depict their expressions, and it would be odd for an agent to display the same expression on their face for a period longer than a few seconds.

Additionally the agent's behaviour is modified depending on the world and the form of ostracism the participant has been assigned to. In the agent world the agents' faces change based on how the agents interact with each other- this happens for all of the four cases – punitive, defensive, oblivious and inclusion (null).

In the punitive condition, it is only when the agent turns around that the agent displays the angry or disgusted expression. However, as soon as the agent turns back to the group, it can displays its current emotion. For example, when the agent turns back to the group it may start smiling so that the ostracism is seen as deliberate, and we want to get a reaction out of the participants. This is similar for the defensive and inclusion conditions.

For the punitive and defensive conditions, when the avatars turn around to face the participant and display the relevant expression, they do this for a period of 4 seconds, then they turn back and face their group. In the inclusion condition, the agents turn around, and remain smiling at the participant until the participant leaves the reaction area.

4. Evaluation Studies

The purpose of this work was to allow Williams to test out his theories on Ostracism, particularly Oblivious Ostracism. The original experiments were to involve only an avatar world. To increase ecological validity we chose to also create an agent world using a emotion-based architecture to see if this made the agents more believable and the studies thus more realistic. While there were numerous psychology-based research questions to be answered via this study, we were concerned primarily with answering *Research Question 6*: Does the Agent Cocktail Party world seem more realistic? For space we only deal with our results related to this question.

The experimental studies involved 49 volunteers recruited only with the knowledge that they were to participate in a study involving IVET and AI. They were not informed about the Ostracism studies. The study had the prior approval of our university Human Ethics Committee. The participants were divided into three groups of 12 and one group of 13. Each group of 12 participants was randomly assigned to one of the Oblivious, Punitive or Defensive Ostracism Studies. With 6 participants in each group experiencing the Avatar Cocktail Party World as their first world, followed by the Agent Cocktail Party World. The remaining 6 participants in the group experienced the Agent Cocktail Party World first followed by the corresponding Avatar Cocktail Party World.

Paul Ekman stated that of his six expressions, the expressions for happy, anger and sad were the easiest to identify. He further documented that some subjects confused the emotions fear and surprise, as they are very similar. We recorded the following correct identification percentages: anger (71%), disgust (55%), Fear (8.5% - 68% confused fear with sadness), happy (78%), Sad (50%) and surprise (32% with 50% confusing surprise with the neutral emotion). These figures show that the only "positive" expressions (Table 1) identified was the happiness expressions. All other emotions identified are negative expressions according to Table 1. We will discuss the impact of these results on the rest of the experiment later.

Following the Emotion Recognition Test participants were asked to put the headset back on and were told that they could begin their first randomly assigned world when they liked, by pressing a button on the game pad. Once their first world was finished, they were instructed to remove the head-set, fill out a questionnaire about the completed world, whilst their second world was configured. They were asked to place the headset back on, complete the second world and then answer a questionnaire about the completed world. At the completion of both worlds, participants were asked to answer questions comparing the two worlds and finally some general technology questions. Participants were then de-briefed and paid for their participation.

The questions in the surveys related to how the subjects were feeling after each world to assess if they had felt the type of ostracism they were being exposed to. Certain emotions were associated with certain forms of Ostracism and these were displayed by the agents/avatars accordingly. In the Punitive Ostracism condition, the avatars and agents displayed either an angry or disgusted expression. In the Defensive Ostracism condition, the avatars and agents displayed either a surprised or fearful expression. In the Oblivious Ostracism condition, the avatars and agents in the room completely ignored and excluded the participant.

When we analysed the scores of how subject's felt after the agent and avatar world's for the condition they had been assigned to, we noted a statistically significant difference (confidence level greater than 99%) between those who experienced the Agent-Punitive condition, and those who experienced the Avatar-Punitive condition. The participants reported a lower mean score in the Agent Cocktail Party world regardless of whether the world was experienced first or second. We can thus conclude that participants feel worse in the Agent Cocktail Party World than in the Avatar Cocktail Party World.

Since the Emotion Recognition Test revealed that the only clearly identifiable positive emotion was happy and that the negative emotions (Anger, Disgust, Fear and Sadness) were more clearly identified the result has been that participants perceived more negative emotions than positive emotions. This result then significantly favours the Punitive Ostracism condition over the other ostracism conditions. As in the Punitive Ostracism condition, it was our intention to increase the feelings of negative mood within the participants, as we wanted participants to feel as if they are in fact being punished. As a result it can be said that in the Punitive Ostracism condition, participants in the Agent world felt worse than participants in the Avatar world.

By the same token, it can be seen that the other conditions (Oblivious, Defensive and Inclusion) could not be affected in the same way as the Punitive Condition. Instead the other conditions are invalidated because of this. This is evident when comparing the means for the Inclusion condition in both the Agent and the Avatar worlds. Finding more positive expressions to model can possibly rectify this situation. Alternatively, further studies may result in finding that William's Ostracism model needs to be modified if the feeling attributed to each type of ostracism, and the corresponding survey questions, are not valid.

On a qualitative level, of the 45 participants, 21 participants preferred the avatar world as they thought the agent world contained people that were very hostile and unwelcoming. Another 14 participants preferred the agent world as they thought they were interacting with more realistic people, or were fascinated by the agents changing their emotional state. The remaining 11 participants preferred either the agent or avatar world for miscellaneous reasons such as the music in the room, or the paintings on the walls even though both rooms were physically identical. Again these results support that the use of emotional agents provided a more engaging experience than the use of avatars whose emotions are scripted.

5. Conclusion

Thus the Artificial Intelligence component was able to aid the psychological study of Punitive Ostracism by creating an environment, which promoted feelings of negative mood in the participant. This confirms what Paul Ekman [11] suggested and what Bates [14] discovered with Disney animators, that humans actually are able to interpret and react to emotions being displayed to them in interaction even with a computer-simulated object. Moreover the human participant was affected by the emotions that they saw. This shows that inclusion of an Artificial Intelligence component based on emotions can add external validity to the psychological testing. These results justify the need to develop methods, which allow for more expressions to be clearly modeled.

Furthermore these results are in keeping with those identified by Paul Ekman [7, 11] and Bates [14], which suggest that emotions are a powerful communicative tool between creatures of the same species. Even though the human participant had no prior interaction with these agent, they were able to comprehend the emotions being displayed and were affected by the expressions depicted by the agent.

Research suggests that emotional states are a direct result of a person's interaction with their environment, if the environment is favorable to the person then the person exhibits a positive emotional state. The idea of incorporating these emotional mechanisms into autonomous software agents has been done in the past [12, 10]. What is novel about the research conducted is that the emotional mechanism has never been explored in the context of interactions with a real human participant. It is this merging of new research work into the emotions with the existing Immersive Virtual Environment Technology, which will be of significant importance to the IVE, AI and psychology communities.

6. Acknowledgements

Thanks to Prof. Kip Williams for his knowledge, guidance and inspiration in this work. Thanks to Dr. Manolya Kavakli and Andy Beall for their support, expertise and input.

7. References

[1] J. Loomis, J. Blascovich & A. Beall, Immersive Virtual Environment Technology as a basic research tool in psychology. *Behavior Research Methods, Instruments & Computer*, 31(4) 1999, 557-564.

[2] J. Blascovich, Immersive Virtual Environment Technology as a tool for Psychological Science, *Psychological Science Agenda*, November/December 2001.

[3] J. Bailenson, A. Beall, J. Blascovich, M. Raimundo & M. Weishbush, Intelligent agents who wear your face: User's reactions to the virtual self, Technical Report, Center for the Virtual Environment and Behaviors, Department of Psychology, University of California, Santa Barbara, 2000.

[4] K. Williams, *Ostracism- the Power of silence*, Salovey (Ed) Emotions and Social Behavior series, Guilford Press, 2001.

[5] J. Cassell & H. Vilhjalmsson, Fully embodied conversational avatars: making communicative behavior autonomous, Autonomous Agents and Multi- Agent Systems, 2, 1999, 45-64.

[6] P. Ekman, R.W Levenson, & W. V.Friesen (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221*, 1208-1210.

[7] P. Ekman, Facial Expression. *In*: Dalgleish & Power (Eds) Handbook of Cognition and Emotions, chapter 3.Wiley & Sons, Sussex, UK, 1999.

[8] E. L. Rosenberg & P. Ekman, Facial expression and emotion. *Neuroscience year: Supplement 3 to the Encyclopedia of Neuroscience*, `993, 51-52.

[9] P. Ekman, Should we call it expression or communication? *Innovations in Social Science Research*, Vol 10, No5, 1997, 333-344.

[10] D. Canamero, Designing emotions for activity selection, Technical Report, Lego Lab, Department of Computer Science, University of Arhus Aabogade, Denmark 34 DK-8200, 1997.

[11] P. Ekman, Basic Emotions. *In*: Dalgleish & Power (Eds) Handbook of Cognition and Emotions, chapter 3.Wiley & Sons, Sussex, UK, 1999.

[12] E. Oliviera. & L. Sarmento, *Emotional advantage for adaptability and autonomy*: In: Proceedings of 2nd International joint conference on Autonomous-Agents and Multi-Agent Systems, Melbourne, Australia, 2003.

[13] D. Rousseau & B. Hayes-Roth, Interacting with Personality-rich Characters, Technical Report, Knowledge Systems Laboratory, KSL 97-06, 1997.

[14] J. Bates, The Role of Emotions in Believable Agents, *Communications of the ACM, Special Issues on Agents*, July 1994, 94-136.

[15] M. Gruter & R. Masters (eds). Ostracism: A social and biological phenomenon. *Ethnology and Sociobiology*, 7, 1986, 149 – 395.

[16] J. D. Snoek, Some effects of rejection upon attraction to a group. *Journal of Abnormal and social psychology*, 64, 1962, 175-182.