Seminarausarbeitung:

Artificial Emotions

Christine Walter

s9cewalt@stud.uni-saarland.de

Universität des Saarlandes, Saarbrücken

Dozent: Michael Kipp

Seminar: Intelligent Virtual Characters

SS 2006

15. Oktober 2006

Abstract

This report is concerned with artificial emotions. First, example applications like emotionally intelligent learning programmes are introduced to motivate the interest in computers that can either recognise emotions or express emotions or both. Then different theoretical approaches to the modeling of emotions are shown, namely the discrete approach, the dimensional approach, the prototype approach and Plutchik's multidimensional approach. Finally, the model for emotions proposed by Ortony, Clore and Collins (OCC-model) is presented and its practical uses investigated.

Contents

1 Introduction

Imagine you come to the office one cold, wet winter Monday morning. You haven't slept enough and you really don't want to have to work. You log in to your computer by telling it "Good Morning!". From your voice your computer can tell that for you it is not a good morning at all and it reacts by showing you the cartoon page of your online newspaper first and telling the office coffee machine to start making coffee. Wouldn't that be nice?

The key to making futuristic scenarios like this one a reality is the research field of artificial emotions. The computer needs to somehow understand that the speaker is in a bad mood. In order to do that, it needs some concept of emotions.

To further motivate why it might be rewarding to research the subject of artificial emotions, in Chapter 2 I will present a few possible applications where computers deal with emotions in some way, either by understanding emotions and reacting to them or by acting in an emotional way themselves. Some applications exist in real life, others are still wishful thinking. All applications are taken from Rosalind W. Picard's book "Affective Computing" [3]. After a few vocabulary definitions, in Chapter 3 I will then present five different theoretical approaches to the modeling of emotions – a discrete approach buikt on the principle of basic emotions, two dimensional approaches that classify emotions by certain characteristics, the prototype approach that combines basic emotions with an hierarchic ordering, and Plutchik's multidimensional approach that combines the discrete and dimensional approaches – followed by the so-called OCC Model developed by Ortony, Clore and Collins, that not only models emotions, but also makes it possible to compute them. In the conclusion finally I will discuss the results and uses of the different approaches, including if and in how far they might help to solve the problems posed in the following applications.

2 Example Applications

2.1 Learning Programmes

Learning in any form is a highly emotional experience. Ideally, a student experiences curiosity, fascination and enthusiasm while learning. These emotions are necessary for the motivation to keep going. But learning can also be frustrating when the student has to work through difficult things. And a certain level of frustration is usually unavoidable to reach a new level of understanding in a subject.

A good teacher will understand this and help the student if the frustration becomes too great, to prevent the student from quitting altogether. And if students are caught up in their studies and really fascinated by a subject, good teachers will do their best to not interfere with or disturb them. So a teacher understands the emotions of the students and reacts to them in the best possible way to support the learning process.

This is one main reason why computer learning programmes can not as yet replace a good teacher. They usually offer help, but they have no way of knowing when help is needed and when it is just distracting and annoying. If the computer could somehow tell the emotional state of the learner, it could give or withhold its feedback depending on the perceived emotions.

2.2 Helping Autistic People

Autistic people have the problem that they often do not recognise emotions in other people. For example, they can not connect the observation of somebody crying with the conclusion that that person is probably sad. But they can learn to connect facial or other expressions with their corresponding emotions. However, this learning process takes a lot of time, as they need a lot of repetitions and also find it very difficult to abstract from one learned situation to another similar one. Therefore, the teachers going through the pictures or simulated situations with them need a lot of patience and time. Computers have both in unlimited amounts, so it would be very useful to have computer programmes that can teach autistic people to recognise emotions.

If computers were able to understand and explain the emotions connected to given expressions or situations, they could lead autistic students through simulated situations and their variations as long and with as many repetitions as needed. This application assumes, of course, that the learning programme would also be capable of responding to the student's emotions as described in 1.3.1 in order to provide support and motivation.

2.3 Using Small Talk to Understand Emotions

Apart from being a way to break uncomfortable silences, small talk can also serve as a sort of 'mood barometer'. For example, if somebody walks into the office in the morning, we can usually tell their current mood from the way they say 'Good Morning'. The better we know a person, the better we can differentiate between emotional states.

The idea is to use this fact to communicate our mood to our computer. Instead of using a user-name and password to log in to one's computer, one could use a verbal log-in, such as "Good morning!" or something similar. Speech recognition is difficult, but in this case the computer would already know the speaker and even the spoken words, so it is imaginable that it could be taught to recognise the emotions communicated in the way something is said.

As described in the opening paragraph in 1.1, the computer could then react to the user's current mood, such as putting up a joke and telling the coffee-machine to make coffee if it recognises a bad or tired mood, or scheduling all the long-delayed unpleasant tasks if it senses a good mood.

2.4 The Right Piece of Music

Often, people want to listen to music that fits their particular mood. A programme that had a way of mapping given pieces of music to the emotions they communicate could do that selection automatically. It would then be possible to tell the radio or CD-player to "play something sad" or "find a channel with lively music" and it would select a channel or piece or at least offer a reduced selection for our choice.

Apart from making it easier to find a suitable piece of music to listen to, such a programme could be very useful and instructive to music students or composers who also need to learn and know what kind of music will elicit what kind of emotions from the audience.

Similar applications might also be possible for images or even movie sequences.

2.5 Text-to-speech Systems

Some disabled people – e.g. famous scientist Dr. Stephen Hawking – rely on text-to-speech systems to communicate with others. In such systems, typed-in text is spoken by a computer voice.

Since it gets only plain text as input, this computer voice can only speak the given text, without inflections or emphasis. But communication usually is a lot more than the spoken words themselves: inflection, speed, emphasis or volume add important nuances of meaning, all of which are lost here.

It would be highly desirable to enhance text-to-speech systems to be able to communicate the emotions of a speaker in addition to the plain words. However, typing already takes much longer than talking, so time is a factor. Two possible solutions are mentioned: First, there could be additional buttons on the keyboard that act as 'emotion-shortcuts', to add things like a laugh, a sigh, an angry or exasperated sound or to indicate that the text should be spoken in a tired or enthusiastic or otherwise emotionally coloured fashion; Secondly, the system could directly observe the emotional state of the user through biometric data and add the corresponding inflections to the spoken words.

2.6 Animated Agents and Sensitive Toys

So far, the applications shown required the computer or machine to have some concept of emotions, to understand observed emotions – from images, observed behaviour or medical data – and act accordingly. On the other hand, machines can pretend to have emotions, i.e. express emotions in some fashion.

Animated Agents in computer games these days express emotions to some degree through body language, and the agents in the *Facade*-game [4] even express emotions through their facial expression, enhancing the playing experience and making the game-play more realistic.

That very limited emotional expression in a machine is enough to create some emotional attachment from humans was proved by the success of the Tamagocchi in the 90's, that had only two emotional states (happy and sad) expressed in a tiny black-and-white drawing. Research shows that having a pet may reduce stress and generally have positive effects on one's health. While electronic pets like RoboDog can not replace real flesh-and-blood pets, for people that suffer from loneliness and are incapable of keeping a pet – e.g. because of old age or mental disorder – sensitive toys like these might be a lot better than nothing.

3 Modeling Emotions

There have been several approaches to modeling and classifying emotions. A number of theoretical approaches are described in [1] and will be introduced in Chapter 3.2. A more practical approach that not only models emotions but even makes it possible to compute them to some extent is the OCC-model [2] that will be outlined in Chapter 3.3.

3.1 Definitions

In order to model emotions, we first need to define what an emotion is – and what it is not. In the following the definitions of affect, emotion and mood from [1] as explained below will be used, but it has to be noted that others may use different conventions – there are no universal definitions for these terms.

3.1.1 Affect

Affect is the most general of the terms defined here and is used to refer to the positive or negative valence of an emotional experience, i.e. to say whether an emotion is positive or negative. In other words, one of the characteristics of an emotion is its affective valence.

3.1.2 Emotion

An emotion itself is defined as an internal mental and affective state. By this definition, e.g. pain is not an emotion, because it is a physical state, not a mental state. Similarly, aggression is not an emotion because it is a behavioural state. Note also that there can be no such thing as a neutral emotion, emotions are always positive or negative.

It is interesting that some researchers argue that by this definition love or hate can not be emotions either, because they always have a social context – you love or hate somebody or something else outside of yourself – and consequently are not purely internal states. We will see in the following chapter that there is no complete consensus on precisely what are emotions and what are not.

3.1.3 Mood

The main difference between mood and emotion is that mood is an emotional state that lasts over a comparatively long time, while emotions might be short-lived. Also, moods can have less specific causes and are generally less extreme than emotions. As with emotions, moods can have a positive or a negative valence, but they can also be neutral.

3.2 Theoretical Approaches

3.2.1 Discrete Approach

The discrete approach to modeling emotions is founded on a distinction between so-called basic emotions and all other emotions.

Basic emotions are defined as emotions that:

- correspond to a unique and universal facial expression,
- form quickly and spontaneously and are automatically appraised and
- are connected with a unique state of feeling.

All other emotions are either mixes of several basic emotions – e.g. envy could be considered a mix of fear, anger and sadness – or they are special cases of basic emotions – e.g. hate would be a special case of anger.

The characteristic of a 'universal facial expression' also means that the more basic an emotion is, the more readily it will be recognised and similarly expressed across different countries and cultures. Smiles for happiness and tears for sadness will be recognised correctly around the world.

Researchers do not completely agree on what the basic emotions are, but the lowest common denominator everyone can agree on is the group of anger, happiness, sadness and fear.

3.2.2 2-Dimensional Approach: Circumplex model

The dimensional approach differentiates between emotions according to two or more characteristics that are assigned some value for all emotions. The Circumplex model has two such dimensions: valence and activity. Valence denotes how positive or negative a given emotion is, activity refers to how actively or passively the emotion is experienced. Consequently, the model results in a two-valued vector for each emotion and makes it possible to illustrate in a 2-dimensional graph how emotions compare with each other (Fig 1).

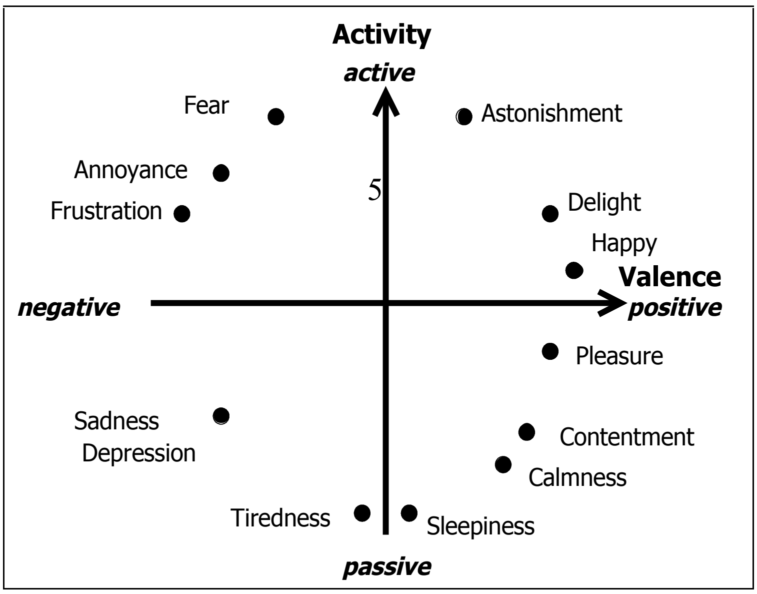

Figure 1: Circumplex model

For example, both happiness and pleasure have a positive valence, but happiness is an active emotion, while pleasure is a passive emotion. And fear and contentment are almost completely opposite emotions, fear being both active and negative, contentment on the other hand being passive and positive.

3.2.3 3-Dimensional Approach

The Circumplex model is a rather simple model, since it reduces emotions to only two characteristics. The 3-dimensional approach makes matters a bit more complex by adding a third characteristic or dimension, the emotional intensity. This new dimension permits the sorting of related emotions according to their intensity, e.g. the group *concern, nervousness, fear, dread, terror* – emotions related to fear – is sorted by increasing intensity.

3.2.4 Plutchik's Multidimensional Approach

In his approach, Plutchik combines the discrete idea of basic emotions with the dimensional idea of ordering emotions according to selected characteristics. The resulting relations between emotions can be illustrated in a cone similar to the RGB-colour-model (Fig. 2).

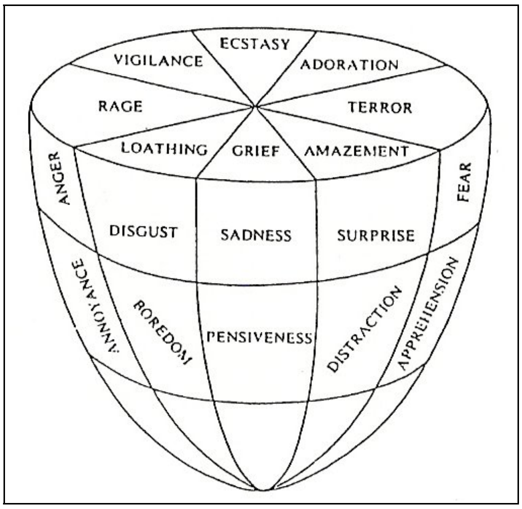

Figure 2: Plutchik's Multidimensional model

The eight emotions in the circular section in the middle of the cone are Plutchik's basic emotions. Neighbouring emotions are closely related, e.g. anger and disgust, while emotions on opposite ends of a circle section are opposed, e.g. fear and anger. The vertical dimension of the cone corresponds to the intensity of the emotion, so going down the cone inside one wedge we would, e.g., come across concern, fear and terror in that order.

3.2.5 Prototypes

The prototypes approach divides emotions into prototype-families, again using the principle of basic emotions. The ordering is hierarchical: The first division is into emotions with positive affect and emotions with negative affect, these two main categories are then divided into the individual basic emotions, and in the third layer below the basic emotions appear all the related emotions. For example, the non-basic emotions *pride* or *contentment* are related to the basic emotion *joy*, and basic emotions like *joy* or *love* fall under the category of *positive affect* (see Fig. 3).

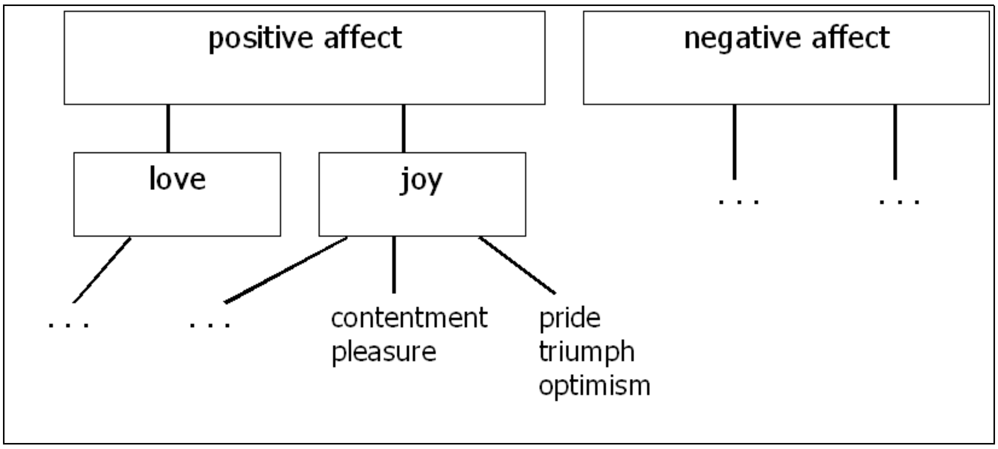

Figure 3: Prototypes model

3.3 The OCC Model

3.3.1 Emotions as Reactions

The OCC-model gets its name from Andrew Ortony, Gerald Clore and Allan Collins, the authors of *The Cognitive Structure of Emotions* [2], who introduce the model.

It is also a discrete approach to modeling emotions, but unlike to the approach seen in Section 3.2.1, it does not use the concept of basic emotions. In fact, the authors do not believe that there really are basic emotions. Instead, they interpret emotions as reactions to either:

- consequences of events, or
- actions of agents, or
- aspects of objects.

The goal of the model is to predict and explain human emotions, or more precisely, to say under which circumstances which emotions are most likely to occur. By dividing emotions into reactions to events, actions and objects and also differentiating between positive and negative reactions, we end up with different types of emotions.

3.3.2 Structure of Emotion Types

The structure of the OCC Model can be seen as a graph where each branching represents a decision regarding the particular situation (see Fig. 4).

The first branch deals with emotions that are reactions to events and their consequences. All emotions on this branch in some way express pleasure or displeasure with events and their consequences. A difference is made between consequences of events that concern oneself and consequences that concern others. If an event and its consequences concern someone else, we end up with emotions of the type labelled as *fortunes-of-others*. Emotions of this type are further divided into reactions to events that are desirable for the other and events that are undesirable and finally the emotion felt depends on how one feels about the other in question. For example, if somebody else wins the jackpot – generally a desirable event – and it was a good friend, you would be happy for him, but if it was your worst enemy, you would feel resentment. Similarly, if a friend had to pay a huge fine, one would probably feel pity, but if it happened to an enemy, one would feel more like gloating. Those four emotions represent the extremes of this particular type of emotions, anything in between that fits the described circumstances would also be part of the *fortunes-of-others*-emotion-type. On the other sub-branch, consequences for oneself can be divided into reactions to events where the prospects are relevant and reactions to events with irrelevant prospects. If the prospects of an event are irrelevant, the emotional reaction will be joy or distress, depending on the desirability of the event. This is the *well-being* type of emotions. For example, if I have nothing special planned for the day, I might still feel joy at wonderful sunshine or distress at endless rain. If the prospects are relevant, things get more complicated. First of all, I will approach an event that has relevant prospects with emotions, namely hope if I want it to happen or fear if I don't. If my expectations are confirmed, I will feel satisfaction – if the hopes were confirmed – or otherwise my fears will be confirmed. If my expectations are not confirmed, I will feel relief – if my fears prove needless – or disappointment – of the hope was in vain. This type of emotions is called *prospect-based*. For example, if someone gets to participate in 'Who wants to be a Millionaire?', they might approach the show with hope (of winning the million) or fear (of embarrassing themselves) and either expectation could be confirmed or not, resulting in the described emotions.

The next branch leads to emotions that are reactions to actions of agents. All emotions in this branch can be seen as approval or disapproval of an action. An agent can be the person itself, or another person or even a group of people or an organisation, or basically anything we ourselves might perceive as an agent, i.e. someone or something that acts in some way, as opposed to an object or event. Again the main differentiation is made between the self as the agent and something else as the agent. If one is the agent oneself, the emotional reaction varies between pride and shame. For example, if a professional athlete breaks her own record, she will feel pride in the action, or if she performs very badly, she might feel shame. If somebody else is the agent and we react to their actions, we will feel admiration or reproach, for example admiration for the success of a brilliant athlete or reproach for one that has been accused of doping. All four belong to the *attribution* type of emotions.

A special sub-branch occurs where *well-being* and *attribution* type emotions are mixed. This happens, for example, when we react to something as both an event and an action. For example, an athlete that has just broken her record might react to the action of breaking the record with pride and to the event of winning the tournament with joy, resulting in the combined emotion of gratification. Similarly, distress and shame mix into remorse, joy and admiration into gratitude and distress and reproach into anger.

The final branch of the graph describes emotions that are reactions to objects or aspects of them. These emotions are simply variations of liking or disliking something, called the *attraction* type of emotions, like love for an object that is especially attractive or hate on the other end of the spectrum.

3.3.3 Predicting and Explaining Emotions

Using the OCC Model, it is now possible to compute emotions, i.e. to predict likely emotions in a given situation or to explain why an emotion occurs. For example, consider this situation:

- Someone gets a parking ticket.
- The someone is not yourself, but someone else.
- Getting a parking ticket is not a desirable event.
- The person in question is somebody you don't like.

The OCC structure of emotions predicts that gloating is a very likely emotional reaction to the situation. Or in reversal, the model allows the reasoning that if somebody feels pity, it is very likely that an undesirable event has happened to somebody else that the person cares about.

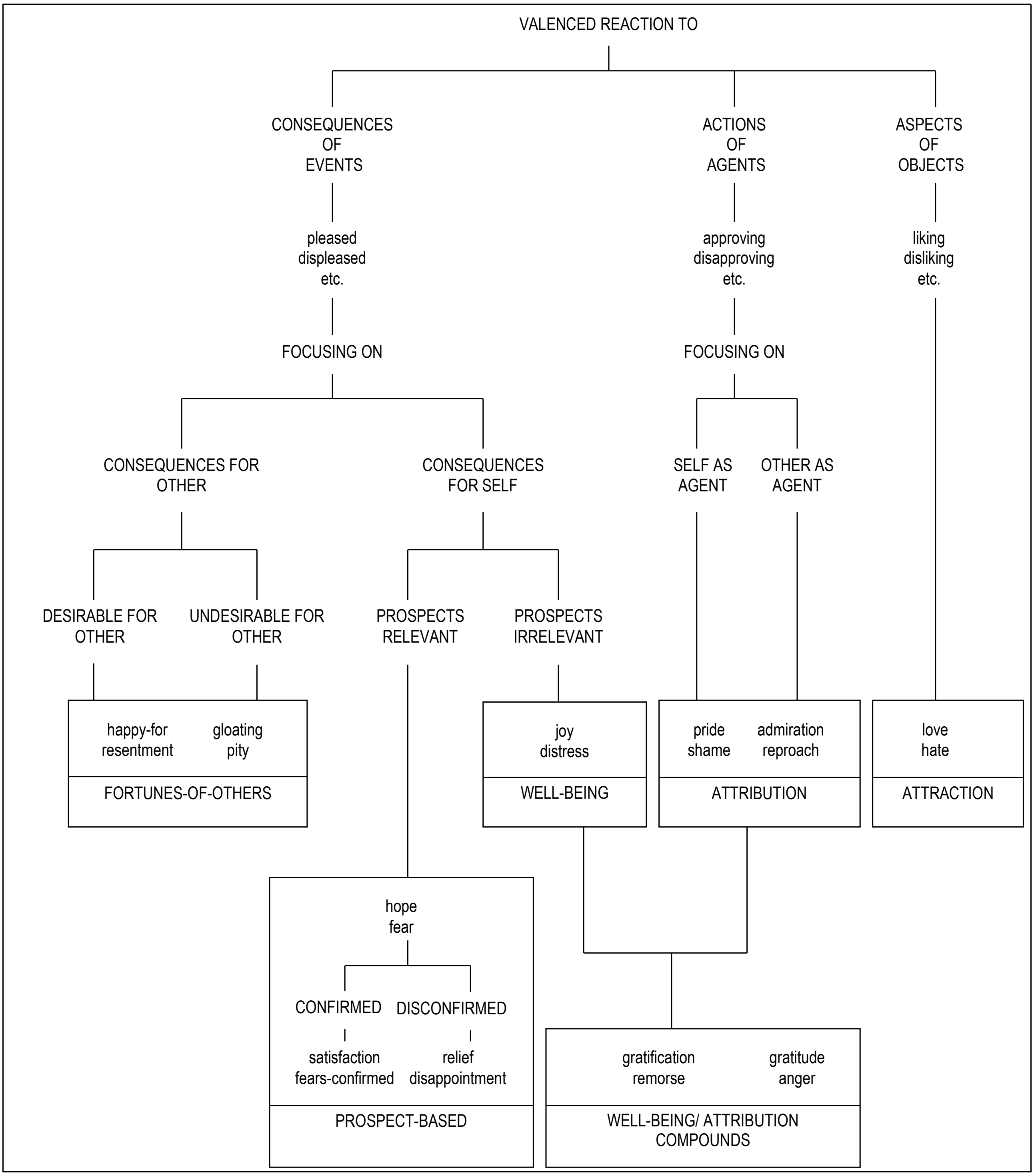

Figure 4: Global structure of emotion types in the OCC Model.

4 Conclusions

As we have seen, there are numerous approaches to modeling emotions. In particular, there are three different ways in which emotions can be classified. First there is the discrete approach as seen in both the OCC-model and the approach using basic emotions and mixes and special cases of basic emotions. Secondly, we have seen the dimensional approach, either with two dimensions as in the Circumplex model or the 3-dimensional approach adding intensity as third dimension to activity and valence. Thirdly, there is the prototype approach that classifies emotions in a hierarchical structure. There are also mixes of those three principles, for example Plutchik's Multidimensional model combines the discrete and the dimensional approach.

The theoretical approaches are mainly useful for categorising and classifying emotions and ordering them. Also, their diversity shows that there is no one perfect answer to the question of how to model emotions. Rather, it is probably a question of finding a model that works for a given problem. The OCC-model, by comparison, is more suited for practical use and makes it possible to compute emotions. It's main idea of understanding emotions as valenced reactions to events, actions or objects seems intuitively reasonable. It might not yet completely cover the entire emotional spectrum, and maybe there should be more compound emotional types such as, e.g., prospect-based and attribution compounds, but the authors themselves admit that the structure could probably be extended further.

The example applications in Chapter 2 have shown us that there are two kinds of practical applications in the field of artificial emotions: machines that understand or recognise emotions and machines that can express emotions or show appropriate emotional behaviour. Despite the example given in Section 3.3.3, it is questionable how useful the OCC Model is for the task of understanding emotions, simply because an observer rarely has all the context of a situation available (if he does, of course, the OCC model should work just fine). Similarly, we ourselves usually do not, e.g., recognise joy from the knowledge that somebody is reacting to a desirable event whose prospects are irrelevant, but rather from clues like facial expression, body language or the way somebody talks. But the reverse might work quite well, i.e. for example a virtual character in a game could choose an emotion based on the OCC structure of emotions by using it as a kind of decision graph. Generally, it is apparently a lot easier to teach a machine how to express emotions than it is to have a computer that recognises them.

Bibliography

[1] Guerrero, L., Andersen, P.A., and Trost, M.R.: Communication and Emotion: Basic Concepts and Approches. *In: P. Andersen and L. Guerrero (eds.) "Handbook of Communication and Emotion: Research, Theory, Application, and Contexts", San Diego, London: Adademic Press, 1998, pp. 49-96.*

[2] Ortony, A., Clore, G.L., Collins, A.: "The Cognitive Structure of Emotions", Chapter 2: The Structure of the Theory. *Cambridge: Cambridge University Press, 1988, pp. 15-33.*

[3] Picard, R.W.: "Affective Computing", Chapter 3: Applications of Affective Computing. *London: MIT Press, 1997, pp. 85-111.*

[4] Mateas, M. and Stern, A.: Facade: An Experiment in Building a Fully-Realized Interactive Drama. *In: Game Developer's Conference: Game Design Track, San Jose, California, 2003.*