WASABI:
Affect Simulation for Agents with Believable Interactivity

Dissertation zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von

Christian Becker-Asano

bei der Technischen Fakultät an der
Universität Bielefeld

14. März 2008

Preface

The research reported in this thesis is the result of the creative and communicative environment I enjoyed during the last eight years in the Artificial Intelligence Group at the University of Bielefeld. Over the years the members of this group discussed even the most far-reaching ideas with me and without their professional and personal support this work would not have been possible.

First and foremost, I am most grateful to my advisor Ipke Wachsmuth for trusting in my hidden abilities from the very beginning. With his idea to let me investigate the fascinating problem of emotion simulation for our virtual human MAX, he gave me the chance to experience the "world of science" to which I am now addicted.

I would like to thank all members of the AI Group, including Stefan Kopp, Bernhard Jung, Timo Sowa, Nadine Pfeiffer-Leßmann, Thies Pfeiffer, Christian Fröhlich, Peter Biermann, Marc Latoschik, Kirsten Bergmann, Ian Voss, Hana Boukricha, and Nhung Nguyen, for reminding me from time to time that the emotion simulation cannot be applied to every problem of Computer Science. Special thanks go to Stefan Kopp for his long-standing technical and ideational support as my colleague and friend. I am also indebted to Helmut Prendinger for supervising me during three-month in 2005 as a Pre-doctoral fellow of the Japan Society for the Promotion of Science in Tokyo, Japan, and supporting me ever since.

Further special thanks go to Klaus Scherer, who agreed to be the second reviewer of this thesis and found the time to share my fascination for this challenging research. I am also grateful to Roger Giles and Nick Thomas for proofreading parts of the manuscript and to Manfred Holodynski, Toyoaki Nishida, Aaron Sloman, Rosalind Picard, and Andrew Ortony for insightful discussions and advices.

Finally, I owe many thanks to my family, especially to my mother Dorothea, my brother Jörg, and my son Jonas. But most importantly, without the love and practical as well as "mental" support of my wife Yoko, finishing this thesis would have been close to impossible—thank you so much for your love and encouragement.

Contents

List of Figures

1 Introduction

Researchers in the field of Artificial Intelligence (AI) try to gain a deeper understanding of some of the mechanisms underlying human cognition. Traditionally pure rational reasoning has been the main focus of research in AI, starting with the conceptualization of the General Problem Solver (GPS) (Ernst & Newell 1969). However, the proposed heuristics could only be applied to a limited set of well-defined problem spaces and Newell & Simon (1972) already mentioned the necessity to “begin to search for the neurophysiological counterparts of the elementary information processes that are postulated in the theories.” (Newell & Simon 1972, p. 146) The GPS-approach on modeling human problem solving turned out to be insufficient with regard to a wide range of problems a human is naturally confronted with. Thus, for the next two decades AI-research focused on so-called expert systems that were used as advice-giving tools for the trained human expert in a limited domain such as medical diagnosis (Shortliffe, Rhame, Axline, Cohen, Buchanan, Davis, Scott, Chavez-Pardo & van Melle 1975). The final diagnosis, however, always lied in the responsibility of the human expert not only to avoid legal issues but also to take into account the complexity of human physiology. Nowadays rule-based expert systems are used in diverse types of applications and they can help to save money by providing domain-specific advice as soon as a certain amount of expert knowledge has been successfully encoded into their rules and knowledge bases (Giarratano & Riley 2005).

User interfaces of expert systems are designed in a dialog-based fashion. The system consecutively asks for more detailed information to be provided by the human expert, to come up with a set of possible solutions to the initially stated problem. Similar dialog-based routines have been used in a famous computer program named ELIZA (Weizenbaum 1976) to create the illusion of speaking to a caring psychotherapist. Despite the similarities of the dialog-based interfaces, Weizenbaum did not intend to build an expert system in the domain of psychotherapy. To his own surprise even professional psychotherapists expected his simple program to be of great help in counselling human patients. This might be due to the fact, that humans are prone to ascribe meaning to another’s responses, even for machines. Even if ELIZA successfully created “the most remarkable illusion of having understood in the minds of the many people who conversed with it” (Weizenbaum 1976, p. 189), it did not pass the Turing test (Turing 1950) as proposed by the early pioneer of computer science, Alan Turing (Hodges 2000). The Turing test was an attempt to provide a suitable test for machine intelligence when the question “Can machines think?” became reasonable to ask. After a five minute conversation—without direct physical contact, e.g. using a type writer machine—with both a human and a machine, a human tester has to decide, which one of the conversational partners is the machine and which one the human respectively. If in at least 30% of the cases the machine is falsely judged as the human, it has passed the test successfully, which no machine has achieved so far. During the conversation the human interrogator is free to choose whatever topic comes to her mind and therefore Picard argues for the integration of humor and more general emotions into artificially intelligent systems that are designed to pass the Turing test. With regard to the limitation concerning the available communication channels during the Turing test, Picard concludes: “A machine, even limited to text communication, will communicate more effectively with humans if it can perceive and express emotions.” (Picard 1997, p. 13) But how exactly can we endow machines with emotions such that they communicate more effectively with humans?

One approach to achieve the effectiveness of natural face-to-face communication of humans is the field of Embodied Conversational Agents (Cassell, Sullivan, Prevost & Churchill 2000). It is motivated by the idea that computer systems might one day interact naturally with humans, comprehending and using the same communicative means. Consequently, researchers in this field have started to build anthropomorphic systems , either in the form of virtual characters using advanced 3D computer graphics or in the form of physical humanoid robots. As these agents comprise an increasing number of sensors as well as actuators together with an increase in expressive capabilities, Cassell et al. (2000) propose an extended, face-to-face Turing Test.

Therefore researchers in the growing field of Affective Computing (Picard 1997) discuss ways to derive human affective states from all kinds of intrusive and non-intrusive sensors. With regard to the expressive capabilities of these agents the integration of the influence of emotion on bodily expression into an agent’s architecture is argued for. These bodily expressions include, e.g., facial expression, body posture and voice inflection and all of them must be modulated in concert to synthesize a coherent emotional behavior.

With the beginning of the new millennium the interest in affective computing has increased even more. Also the public has shown a renewed interest in the possible future achievements of AI, for example, a series of recent movies tackling the question of “emotional robots” (*I, Robot* and *Bicentennial Man*) as integrated members of a future society. In the near future, humanoid agents are to take part in social interaction with humans and therefore the integration of psychological concepts like emotions and personality into rational agents seems inevitable. Despite the ongoing debate about the formal definition of such concepts, many computational models have been proposed to simulate emotions for humanoid agents.

1.1 Motivation

> The Three Laws of Robotics:
>
> 1. A robot may not injure a human being, or, through inaction, allow a human being to come to harm
> 2. A robot must obey the orders given by human beings except where such orders would conflict with the First Law
> 3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law
>
> Established in the short story “Runaround” by Isaac Asimov (1942)

With computerized machines becoming increasingly powerful—both in the computational as well as the physical sense—the fear that such machines might one day supersede our human society is naturally evolving (Sloman 2000). As mentioned above such an hypothetical future gained high attention in the general public, because of a series of movies such as “Bicentennial Man” (Asimov, Silverberg & Kazan 1999), “A.I.” (Spielberg 2001) and “I, Robot” (Sietz 2004) that outline a future society, in which humanoid robots or androids live among us. In order to make these robots safe, their designers and programmers are assumed to have taken necessary and (hopefully) sufficient precautions.

Although the robots in the movie “I, Robot” are explicitly programmed to follow the three laws of robotics, they are far from being judged as perfect members of human society. The protagonist reports on a decisive occasion in his life, when he was being rescued instead of a much younger girl. The humanoid robot calculated that she had slightly less chances of survival and this led to a feeling of guilt in the survivor. During the course of the movie, however, a special robot is introduced that seems to be capable of “having” emotions[1]. This robot, named “Sonny”, is also able to deliberately break the three laws of robotics as to him the final logic they imply (in the movie)—that humanity is to be saved by means of captivity from harming itself—“just seems too heartless”.

The protagonist of the movie “Bicentennial Man” is a robot itself that—after two hundred years of existence, just before it ‘dies’—is declared human, because it has proven to be capable of creative thinking, moral judgement and even falling in love with a human. The same exceptional abilities, after some time, lead to full acceptance by the members of the family he is at first only serving for. But as the story evolves the unpredictability that is assumed to come together with creativity is rated too dangerous and, thus, the newer robots are programmed more strictly and confining.

The role of androids as social partners in a future society is the main topic of the movie “A.I.” (Spielberg 2001). It tackles the interesting possibility of human-like robots, i.e. androids, being used as ersatz-children and ersatz-lovers. Such androids as partners in intimate relationships are also examined in the scientific community already (Levy 2007).

In summary, many questions arise in the context of machines as social partners some of them serving as background for this thesis and can be stated as follows:

1. What is needed to build a robot one can fall in love with?

2. How can designers and programmers support a sustainable relationship with such artificial partners?

3. If we really succeed in building such complex, lovable and (presumably) autonomous artificial partners, are they still to be treated as machines as soon as they are malfunctioning?

Of course one can argue that for the sake of humanity nobody should ever even try to build such artificial partners. For the sake of science, however, one might have another motivation for research on sociable robots, namely, the experimental-theoretical motive (Burghouts, op den Akker, Heylen, Poel & Nijholt 2003). In trying to understand human psychology computer simulations might help to systematically combine and investigate psychological theories with interpretations of neurobiological findings.

[1] The difference between “having” versus only “showing” emotions is clarified in Chapter 2.

This thesis describes an implemented Affect Simulation Architecture, which not only combines different emotion theories and neurobiological findings, but is also successfully integrated it into an Embodied Virtual Agent and evaluated in two different interaction scenarios. In order to explain how the different theories and findings can be fused, the interdisciplinary background is clarified in Chapter 2. In the following section a brief overview of the author's main field of research "Artificial Intelligence" is given together with an introduction to the computational background, in which the Affect Simulation Architecture is integrated.

1.2 Artificial Intelligence background

Artificial Intelligence (AI) is a subfield of Computer Science and the term itself was born in 1956 during the *Dartmouth Conference* in Hanover, New Hampshire (for a review of its history see Buchanan 2005; Russell & Norvig 2003; Wachsmuth 2000). The initial enthusiasm about possible achievements in this field soon started to fade away after researchers realized the complexity of real world problems.

Nevertheless, the field still attracts many researchers with an interest in how the human mind happens to fuel intelligence and many introductory textbooks have been written about this fast changing field of research (for an overview see Russell & Norvig 2003, p. 2). The following brief introduction to AI mainly follows the lines of Russel and Norvig's influential book "Artificial Intelligence A Modern Approach" (Russell & Norvig 2003).

1.2.1 Four approaches to AI

In their book's introduction Russell & Norvig (2003) distinguish four approaches to AI that have been followed in the past and are introduced here in an abbreviated fashion.

1. Systems that act like humans

The Turing test (described in the beginning of this chapter) was proposed as a means to evaluate an AI system's human-likeness. The extended version—known as the total Turing test—includes a video signal to allow for direct face-to-face communication and allows for passing physical objects between the interactants. Russell & Norvig critically observe that AI researchers so far "have devoted little effort to passing the Turing test, believing that it is more important to study the underlying principles of intelligence than to duplicate an exemplar." (Russell & Norvig 2003, p. 3) Recently, however, an increasing number of researchers began building human-like virtual or robotic agents aiming at an understanding of the complex interaction of different channels of human expressivity, such as facial and bodily expressions in verbal and nonverbal communication.

Nonetheless, the following analogy is important to understand a basic principle of AI:

> "The quest for 'artificial flying' succeeded when the Wright brothers and others stopped imitating birds and learned about aerodynamics. Aerodynamical engineering texts do not define the goal of their field as making 'machines that fly so exactly like pigeons that they can fool even other pigeons." (Russell & Norvig 2003, p. 3)

This example is often shortened to "Thats why planes don't flap their wings!" and it clarifies that AI researchers do not aim at duplicating human's cognitive abilities as exactly as possible. But a decisive difference between aeronautics and AI is not to be missed: aeronautics researchers deal with the well defined phenomenon of "flight" whereas AI researchers try to simulate processes underlying the ill defined concept of (human) "intelligence". If we had a better definition of what it means to be intelligent—comparable to our understanding of what it means to be flying—we could probably also start ignoring its biological roots. In this context Wachsmuth (2000) highlights that "it is not the aim of AI to build intelligent machines having understood natural intelligence, but to understand natural intelligence by building intelligent machines." (Wachsmuth 2000, p. 45)

2. Systems that think like humans

In the attempt to build systems that think like humans, theories and findings of different disciplines are taken into account, but Wachsmuth (2000) follows Winston (1992) by pointing out that "AI differs from most of psychology because of its greater emphasis on computation, and it differs from most of computer science because of its greater emphasis on perception, reasoning, and action." The field of cognitive science defines itself as an interdisciplinary approach—combining philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology—to understanding and modeling the performances of humans and animals. Scientists in this field attempt to build computational models of human cognitive behavior in order to combine and verify the findings of the different disciplines.

Research in AI has a slightly different scope, because computational models of human behavior are central to it rather than experimental investigations of actual humans or animals. In effect, however, every computational solution has to perform similarly enough to the performance of a human in the same situation. If a general match of performance is achieved, AI researchers and cognitive scientists have to decide whether the underlying mechanisms are similar as well or at least comparable to each other.

3. Systems that think rationally

The ability of humans to think rationally has led the way of early philosophers (cf. Beckermann 2001, for a review). AI researchers, who model their systems to copy this ability, follow the "laws of thought" (Russell & Norvig 2003, p. 4) approach. Despite the general success of this approach on small scale problem spaces, the problem known as "combinatorial explosion" became obvious very soon. As of today, this subfield is highly active in the AI community trying to solve or at least circumvent this and similar problems by inventing special purpose solutions for different problem classes.

4. Systems that act rationally

The notion of a system capable of intelligent action in the real world brought up the term "Intelligent Agent". In short, an agent is believed to act rationally on the basis of factual knowledge by following the Principle of Rationality:

> "If an agent has knowledge that one of its actions will lead to one of its goals, then the agent will select that action." (Newell (1982), after Wachsmuth 2000, p. 47)

This “rational agent approach” (Russell & Norvig 2003, p. 4) involves the ability of an agent to follow the “laws of thought” mentioned above, but it complements it with deliberative goal-directed action. Furthermore, an intelligent agent’s internal processing can be captured by the perceive-reason-act triade (cf. Figure 1.3, p. 11) and the intermediate process of reasoning “involves internal processes that make a subject ‘think’ about what might be the best way of acting before actually moving to act.”[2] (Wachsmuth 2000, p. 44)

Accordingly, for an agent able to perceive the world it has to be able to represent aspects of the world in an internal knowledge base. Then reasoning is most often realized by means of some kind of first- or second-order predicate logic based on rules that transform the internal representation deriving new facts and discarding the implausible ones. Finally, an agent is assumed to act in the world causing an immediate or delayed goal-conducive effect.

Summary

The four approaches to AI research are not to be understood as mutually exclusive, because only the central aspect of investigation is to some extent different. In general, researchers dealing with robotic agents also have to solve problems of natural language understanding, planning, and human-like behavior as soon as their robots have a humanoid appearance and are assumed to assist in social life contexts.

The members of the Artificial Intelligence group at Bielefeld University (the author being one of them) were traditionally interested in the last (acting rationally) approach to AI, but with the advent of increasingly powerful computer systems the first (acting humanly) and second (thinking humanly) approach became ever more important in their research on Human-Computer Interaction (HCI). As outlined in the beginning of this chapter, HCI research has led to the invention of Embodied Conversational Agents that combine human-like appearance with human-like behavior. These agents are situated in a virtual environment and equipped with a virtual body enabling them to use the same multimodal communicative means as humans in conversation.

Before our work on the development of ECAs is presented, the underlying concepts “Situatedness” and “Embodiment” are briefly discussed.

1.2.2 Situatedness and Embodiment

According to Russell & Norvig (2003), AI researchers concerned about intelligent agents started in the late 1990s to gain interest in the “whole agent” problem again. In this view, an agent’s cognitive abilities cannot be separated from its physical body (embodiment) and situational context (situatedness).

Situatedness

Researches of the so-called “situated movement” (Russell & Norvig 2003, p. 27) focus on “agents embedded in real environments with continuous sensor input” (cf. Lindblom & Ziemke 2003, for a critical discussion). Of course, situatedness most often refers to real world robots

[2] This idea can be traced back to Newell and Simon’s proposal of a “General Intelligent Agent” (Newell & Simon 1972) as an early paradigm of AI; see beginning of this chapter.

that are acting among us, but it also applies to software robots that are situated in the world wide web, such as web-crawlers or auction bots in Russel and Norvig's opinion.

With respect to the internal reasoning capabilities of Situated AI systems Wachsmuth states:

> "[I]t is crucial for Situated AI to deal with embodied systems that are able to modify their internal processing while they are coupled to their environment by way of sensors and actuators." (Wachsmuth 2000, p. 55)

With our group's development of intelligent virtual agents, which are based on a strong computational background in the field of AI, these requirements are met.

Embodiment

Situatedness mostly involves some kind of embodiment and in his review of Artificial Intelligence Pfeifer (2001) emphasizes the need for "Embodiment" in modern AI approaches. In his view it is a promising challenge for AI "to build robots that can mimic the processes of human infant development." (Pfeifer 2001, p. 306) The use of the term "embodiment" in Pfeifer's opinion entails two main types of implications. First, dealing with the physical implications means to find solutions for the classical problems of robotics, namely the handling of all kinds of physical forces like inertia, friction, vibrations and energy dissipation. The second type of implications is information theoretic and it is concerned with "the relation between sensory signals, motor control, and neural substrate." (Pfeifer 2001, p. 297) To this respect Pfeifer follows a general distinction between "a body and a mind" in that he ascribes information theoretic processes to the brain (i.e. the "neural substrate") alone, without influence from the body.

Recently, as Niedenthal, Barsalou, Winkielman, Krauth-Gruber & Ric (2005) point out, the focus of embodiment has shifted from investigating the role of actual bodily states in cognition to that of the simulation of experience in modality-specific systems in the brain. The latter notion is also supported by neurobiological findings (cf. Damasio (1994), LeDoux (1996)) that will be discussed in Section 2.2. In everyday, face-to-face communication only using the right words at the right time in response to another's statements is not sufficient to appear intelligently. Our whole body is usually used for communication, including our tone of voice, facial expressions, gestures and postures. In his proposal for a "design-based theory of affect" Sloman (1992) highlights that facial expressions are also driven by involuntary mechanisms that are not caused by deliberative processes.

In this context Reeves & Nass (1998) point out that humans already treat disembodied computer systems as social actors. Their study on flattery, for example, shows that humans have a better opinion about computers that praise them than those that criticize them. This effect remains stable even if the users are made aware of the fact that the computer has no means to evaluate their responses and is simply producing random comments. Concerning the role of emotions in the media the authors first point to neurophysiological evidence supporting the basic assumption that every emotion has an inherent positive or negative quality, i.e. a "valence" dimension. In combination with their studies on "arousal" they conclude, that people confronted with media content do react with the same variety of emotional responses as in face-to-face interaction between humans.

Notably, all of these studies did not involve any kind of anthropomorphic interface. The different social and emotional aspects of the computer's responses were only encoded on the textual level, but even this very limited communication channel was efficient enough to support their hypotheses.

Summary

As explained before, with the Affect Simulation Architecture described in this thesis the author aims to combine different emotion theories to not only implement affect for humanoid agents but also to falsify the predictions of these theories. Therefore, the Affect Simulation Architecture must be well-grounded on some theoretical framework. The empirical data provided by Reeves & Nass (1998) are important in so far as they have to be taken into account when designing means to evaluate embodied emotional agents, as will be discussed in Section 5.2.

1.2.3 Embodied Conversational Agents

The term "Embodied Conversational Agents" (ECAs) was officially introduced by Cassell et al. (2000) (see also the introduction to this chapter). The different contributors to this book discuss the complexity of generating human-like virtual agents including the integration of "*emotion*, *personality*, *performatives*, and *conversational function*" (Cassell et al. 2000, p. 2) Pelachaud & Poggi (2002) provide a comprehensive discussion of these aspects together with an overview of different implementations. Cassell argues for the development of human-like interface agents in the following way:

> "Because conversation is such a primary skill for humans and learned so early in life (practiced, in fact, between infants and their mothers taking turns cooing and burbling to one another), and because the human body is so nicely equipped to support conversation, embodied conversational agents may turn out to be a powerful way for humans to interact with computers. " (Cassell 2000a, p. 71f)

With respect to the state of the art Cassell (2000b) admits that "the number of conversational behaviors that we can realize in real time using animated bodies is still extremely limited." She also states that "[o]ur models of emotion, of personality, of conversation are still rudimentary." (Cassell 2000b, p. 23). Focusing on our group's own work the development of ECAs is briefly outlined next.[3]

The Virtual Interface Agent "Hamilton"

Starting in 1993 our group headed by Professor Wachsmuth continuously investigated the use of agents in virtual reality contexts (cf. Wachsmuth & Cao 1995; Wachsmuth, Lenzmann, Jörding, Jung, Latoschik & Fröhlich 1997).

In the VIENA Project ("Virtual Environments and Agents") Wachsmuth et al. (1997) introduced a virtual interface agent (VIA) called "Hamilton" (cf. Figure 1.1), which assists the human user in a 3D virtual reality office presented on a computer screen. Notably, in the VIENA system the user interacts by means of speech and gesture in combination and Wachsmuth et al.

[3] The integration of "affective" qualities (such as emotions and personality) into ECAs is discussed in Chapter 3 after a clarification of these concepts in the following chapter.

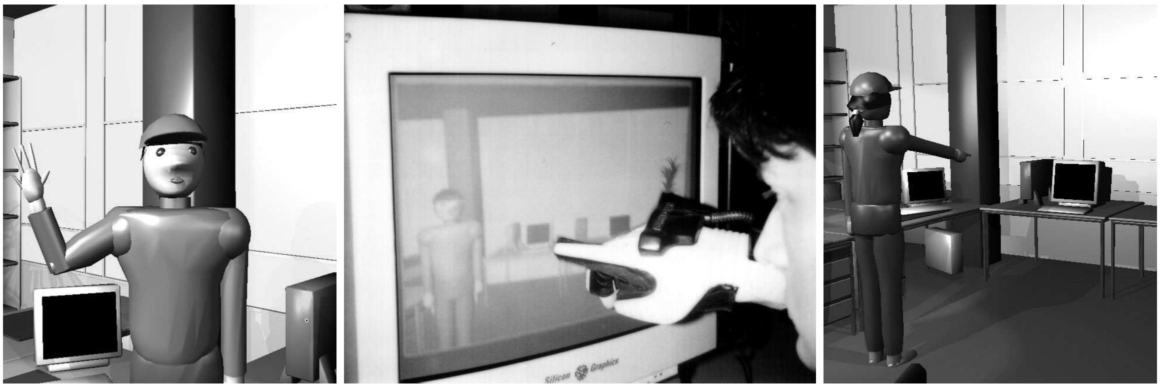


Figure 1.1: The Virtual Interface Agent “Hamilton” first greets the human user, who then points to an object on the screen, and finally Hamilton explains the indicated object together with a pointing gesture (Wachsmuth et al. 1997)

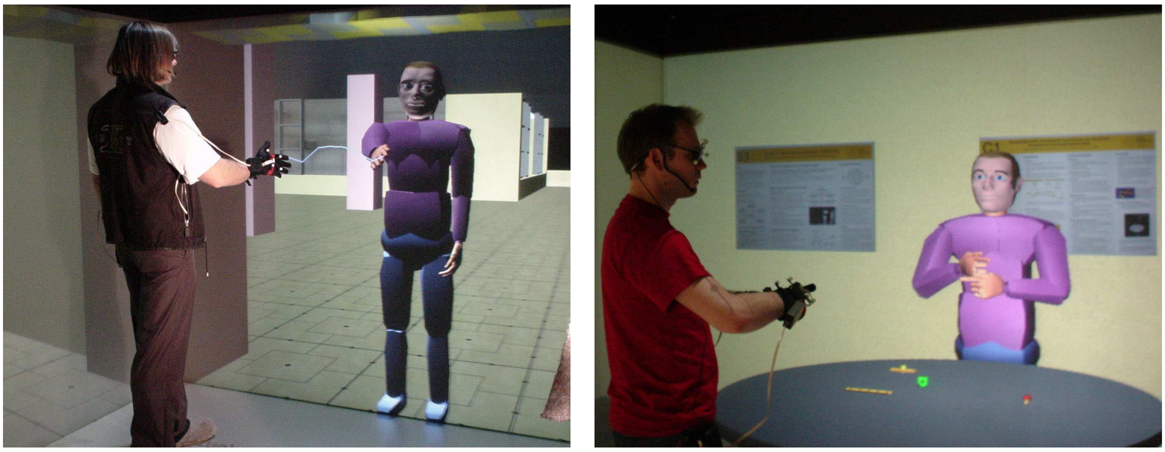
argue that “[i]n the presence of a human-like figure, it is natural to include means of verbal interaction, especially when gestural manipulation is impossible or unnatural [..].” (Wachsmuth et al. 1997, p. 517)

Through the Hamilton agent the otherwise omnipresent AI is personified and, thus, directly addressable. It is situated in the virtual environment. The “situatedness” of interface agents is of central interest to this thesis, because simulated affective states are visualized by means of an embodied agent. Hamilton’s expressive abilities, however, were quite limited until Kopp & Wachsmuth (2000) integrated a knowledge-based approach for lifelike gesture animation. Even after this improvement Hamilton was still incapable of producing lip-sync facial animation or any other kind of facial expression. Together with the increasing quality of real-time computer graphics the human interlocutors, however, expect an even more human-like virtual interface agent.

The Multimodal Assembly eXpert “MAX”

In the context of the Collaborative Research Center (SFB) 360, which was concerned with the design of “Situated Artificial Communicators” that “integrate multimodal conversational abilities for task-oriented dialogs in dynamic environments” (Kopp, Jung, Leßmann & Wachsmuth 2003, p. 11), the development of the embodied conversational agent “MAX”—the “Multimodal Assembly eXpert”— was started (Kopp & Wachsmuth 2002). Since then, our group’s three-sided large-screen projection system together with sophisticated video-based sensor technology and speech recognition enables us to interact most naturally in virtual reality (VR) as shown in Figure 1.2. With the positive experiences gained with “Hamilton” and based on the increased computing power the development of an ECA with extended expressiveness was the logical next step toward an even more natural interface.

In a student project the outer appearance of MAX was designed as to resemble an adult human man (cf. Figure 1.2(a)). In his PhD-thesis, Kopp (2003) presents an implementation of synchronous speech and gesture animation for MAX that is well-founded in the theoretical context and has proven to produce natural and believable results (Kopp & Wachsmuth 2004). MAX’s gestural expressivity enables him to imitate a human interlocutor’s gesture based on a high-level abstract description of the gesture’s content instead of applying direct motion



(a) MAX as a guide to our virtual lab “shaking hands” with the human interlocutor

(b) MAX in the SFB 360 scenario imitating a human interlocutor’s gesture

Figure 1.2: The Multimodal Assembly eXpert MAX in two different scenarios in our CAVE-like three-sided large-screen projection system

capture techniques (cf. Figure 1.2(b)).

In her diploma thesis Leßmann (2002) started to conceptualize a cognitive architecture, which is used for modeling the cognitive abilities of MAX (cf. Figure 1.3). It builds upon the aforementioned perceive-reason-act triade and enables MAX to combine fast, reactive behaviors with relatively slower, deliberative ones. The reasoning capabilities are realized by means of a “cognitive loop” (Leßmann, Kranstedt & Wachsmuth 2004, p. 60), which is based on the Belief-Desire-Intention (BDI) approach (cf. Bratman 1987; Rao & Georgeff 1991). The internal reasoning capabilities of this cognitive architecture are detailed in Chapter 6.

With respect to the integration of emotions into our agent’s cognitive architecture a separate emotion simulation system was devised in the author’s diploma thesis (Becker 2003). As detailed in Chapter 4 it has proven to provide a believable emotion dynamics in a conversational museum guide scenario (Becker, Kopp & Wachsmuth 2004). Over the last four years a number of extensions together with further empirical studies have been accomplished. The rationale for these extensions together with first evaluations of their effects on humans are the topic of this thesis.

1.3 Thesis scope and objectives

This thesis aims to provide a comprehensive, fully-implemented, and well-founded simulation of affect for virtual as well as robotic humanoid agents. The conceptualized architecture is called “Affect Simulation Architecture” or, alternatively, WASABI architecture. It builds upon the author’s existing implementation of emotion dynamics, which is integrated in the Affect Simulation Architecture as a highly interconnected, though concurrent module.

The motivation to propose such an architecture is twofold, because the WASABI architecture is (1) supposed to increase an agent’s believability in social interaction and (2) based on highly interdisciplinary research in the hope to help establishing ties between cognitive science, psychology, neurobiology, and computer science.

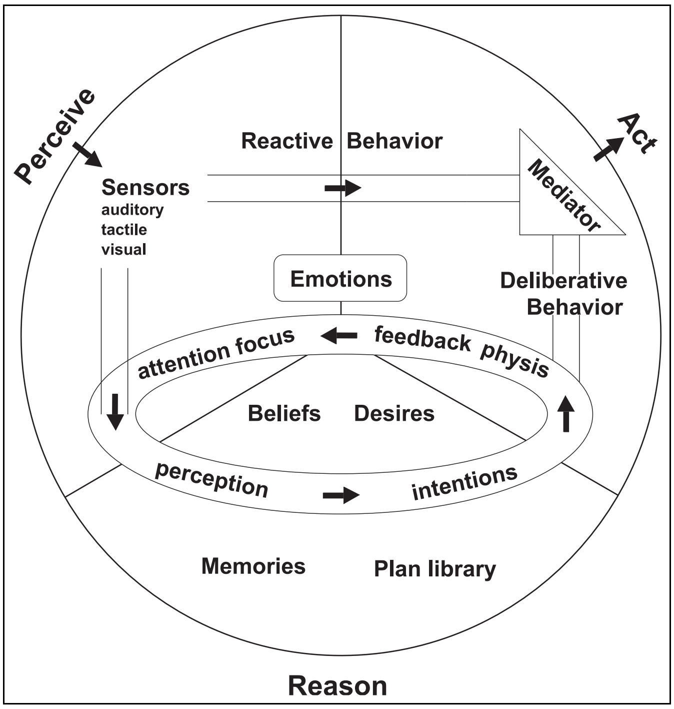


Figure 1.3: MAX's overall cognitive architecture as the perceive-reason-act triade (Leßmann et al. 2006)

1.3.1 Increasing believability

Humans confronted with virtual agents such as MAX naturally expect a high degree of sophistication with respect to the agent's interactive capabilities. With an agent having a humanlike face one expects facial expressions of a certain quality and style being expressed in accordance with the situational context. Being equipped with two arms and two legs one expects the agent to perform natural gestures in synchronization with verbal and non-verbal expressions.

Our group's virtual human MAX is devised as a testbed for evaluating different approaches to naturalize human-computer interaction. He is able to perform a variety of facial expressions, lip-sync facial animations in accord with any verbal utterance, tightly synchronized co-verbal gestures, and he perceives the human interlocutor by means of a multitude of sensors such as camera-based motion trackers, data-gloves, and microphones (cf. Figure 1.2). Furthermore, MAX plans his actions based on a variety of planners, which are incorporated into a domain-independent cognitive architecture and combined with reactive and proactive behaviors at runtime.

To this respect, MAX resembles an adult human capable of rational problem solving and problem-focused, multimodal interaction, but social interaction includes an understanding and appropriate expression of affective states and processes. The better such a sophisticated humanoid agent as MAX is able to take part in social interaction the more believable he will be. This assumption—motivating the development of the WASABI architecture—is not taken for granted in this thesis, but is verified within two empirical studies. Computer scientists who are interested in increasing the believability of their agents through the simulation of affective phenomena follow the *believable-agent motive* (Burghouts et al. 2003).

1.3.2 Interdisciplinary research

“What is an emotion?” has been asked many times within the last 150 years of research on emotions. This question is also the title of the influential article by James (1884).

In the aim to find answers to this question researchers proposed a variety of theories, which are based on introspection, intensive study of human or animal behavior, intercultural studies of facial expressions of emotions in humans and animals, investigations of structures derived from linguistic labels for emotions, neurobiological findings in humans and animals, among others (cf. Chapter 2).

Beginning in the 1980’s cognitive scientists as well as some psychologists gained interest in computer simulations of their theories (cf. Chapter 3). Computer scientists motivated by this idea follow the *experimental-theoretical motive* (Burghouts et al. 2003).

With the WASABI architecture the author presents his attempt to combine different findings and conceptions of emotion psychology, neurobiology, and developmental psychology in a fully-implemented, clearly arranged, and last but not least well-founded computational architecture that proved to provide a useful emotion simulation for a virtual human in two different human-computer interaction scenarios.

1.4 Thesis structure

In Chapter 2 the interdisciplinary background—including findings of psychology and neurobiology—is presented and discussed. The physical components and mental abilities necessary to capture emotions are central to this chapter’s interdisciplinary overview. After starting with the assumption that emotions result from the self-perception of bodily changes as first proposed by James (1884) and afterwards refined by Cannon (1927), a more introspective view of emotions is adapted in the discussion of dimensional theories of emotions. Taking a top-down perspective by investigating the cognitive structures and processes assumed to underly human emotions the broad field of cognitive emotion theories is examined next. We then take a look at the brain, because it is central to cognitive functions of different complexity. Reviewing and discussing recent findings of neurobiology yields evidence for a distinction of at least two classes of emotions, primary and secondary ones. Furthermore, research on emotional development supports this distinction of different classes of emotions that go hand in hand with the acquisition of increasingly complex cognitive abilities during ontogenesis. It is also shown that rational reasoning is influenced and may in turn be supported by emotions that themselves make use of body-maps representing the bodily state within the brain.

Computational architectures for modeling emotions (may they be applied to virtual or robotic humanoid agents) are reported on in Chapter 3. Not only are the conceptual considerations discussed in the light of the previous chapter, but also are the different types of physical and virtual agents reviewed.

Chapter 4 provides a suitable understanding of primary and secondary emotions together with an explanation of the central concept of emotion dynamics. In explaining this dynamics the distinction between mood and emotion is clarified and how personality-related aspects are reflected in high-level parameters of the Affect Simulation System. Subsequently, the simulation of primary emotion dynamics is described together with slight modifications and extensions that had to be applied to the initial conception of Becker (2003). The integration of three secondary emotions is detailed next, before the connection between MAX's cognitive component and his concurrently simulated emotion dynamics is explained. It is based on a distinction between conscious and non-conscious appraisal arguing for at least two different kinds of representations realizing appraisal mechanisms on different timescales and granularities. Based on findings from neurobiology a connection to the proposed "as-if body-loop" Damasio (1994) is drawn supporting the conceptual distinction of body and brain. In effect, it is reasonable to introduce the concept of conscious and non-conscious emotions as resulting from the body-brain interaction. Finally, it is outlined how the conscious emotions can become subject to reappraisal and how emotions in general can influence the cognitive processes at different levels.

Chapter 5 first describes the successful application of primary emotion simulation in a conversational agent scenario. With the positive experiences gained in this scenario the author applied the Affect Simulation Architecture to a more controllable, non-conversational interaction scenario. This competitive gaming scenario is outlined in the context of an empirical study, that was conducted to evaluate the effect of emotional and empathic agent behavior. For this study the integration of bio-metrical emotion recognition and empathic agent feedback is explained, before the results of the study are described in detail. In the summary of Chapter 5, an argument is given for the integration of secondary, more adult-like emotions as an extension to the simulation of primary emotions that has proven reasonable.

Chapter 6 concentrates on explaining the computational integration of secondary emotions for which a number of changes and extensions to the cognition as well as the emotion module of the cognitive architecture (outlined in Chapter 4) had to be applied. In result, the WASABI architecture is introduced as a fuller account of an Affect Simulation Architecture and exemplary utilized in the gaming scenario. The BDI-based cognitive reasoning capabilities are detailed and plans are presented that give MAX the ability to process expectations in the gaming scenario. Special purpose plans are then introduced by which the two-way connection between cognition and emotion is established, leading to the elicitation and expression of mood-congruent secondary emotions. The results of a final empirical study—comparing the pure simulation of only primary, child-like emotions with the combined simulation of primary and secondary emotions—are presented and discussed in the end of this chapter.

Chapter 7 concludes this thesis with a critical review of what has been achieved and how much further this architecture might be extended in the future.

Parts of the concepts and results developed in this thesis were already published in (Becker et al. 2004; Becker, Kopp & Wachsmuth 2007; Becker, Leßmann, Kopp & Wachsmuth 2006; Becker, Nakasone, Prendinger, Ishizuka & Wachsmuth 2005; Becker, Prendinger, Ishizuka & Wachsmuth 2005a; Becker et al. 2005b; Becker & Wachsmuth 2006a,b; Becker-Asano, Kopp, Pfeiffer-Leßmann & Wachsmuth 2008; Boukricha, Becker & Wachsmuth 2007; Kopp, Becker & Wachsmuth 2006; Prendinger, Becker & Ishizuka 2006).

2 Interdisciplinary background

In his discussion of the connection between “Communication and Affect”, Sloman (1992) distinguishes three kinds of theories for modelling affect[1]:

1. Design-based theories locate human mechanisms within a space of possible designs, covering both actual and possible organisms and also possible non-biological intelligent systems.

2. Semantics-based theories attempt to make sense of the structure of some portion of the lexicon of ordinary language.

3. Phenomena-based theories assume that some particular kind of phenomenon can be intuitively recognized (e.g. emotional states) and then investigate other phenomena that are correlated with it in some way, e.g. physiological causes, physiological effects, behavioral responses, cognitive processes.

In this chapter different concepts related to emotion are outlined, resulting from the different scientific disciplines together with their continuously changing methodologies. At first the psychological approaches are discussed. Their majority falls into the category of phenomena-based theories, some others into the semantics-based theories, and even fewer follow the design-based approach. As every psychological analysis of emotions aims to be as sound and complete as possible, none of them can be assigned to one of the above classes exclusively. Nevertheless, in an attempt to structure the theories their major conceptual approach is classified with the help of the three kinds of theories above whenever possible and useful.

2.1 Psychological background

According to Scherer (1984), the “psychological construct” labeled emotion can be broken up into the following components:

a) The component of cognitive appraisal or evaluation of stimuli and situations.

b) The physiological component of activation and arousal.

c) The component of motor expression.

d) The motivational component, including behavior intentions or behavioral readiness.

e) The component of subjective feeling state.

[1] These three kinds of theories are not incompatible and, thus, sometimes combined (Sloman 1992, p. 233).

As explained in Chapter 1 in the context of embodiment "mental states"—such as feeling happy—are assumed to result from some kind of dynamics between cognitive processes and bodily states. Early psychologists, who investigated this dynamics, concentrated on the physiological component of emotions taking into account aspects of emotion expression and subjective feeling state. Accordingly, their so-called "feedback theories" only provide little information about the appraisal processes necessary to evaluate an event or situation. They belong to the class of phenomena-based theories as they were—at least in the beginning—mainly based on an intuitive understanding of the processes involved in emotion elicitation. A comprehensive discussion follows in Section 2.1.1.

With an interest in the motivational component an overview of the so-called "basic emotion" theories is given in the beginning of Section 2.1.2. The dimensional theories (presented subsequently) are best suited to account for the physiological component, although they also contain aspects that represent an individual's "subjective feeling state".

The (cognitive) evaluation of stimuli is central to the class of emotion theories labeled "appraisal theories" discussed in Section 2.1.3.

2.1.1 Feedback theories

William James (1884) and Carl Lange (1885) almost at the same time brought up the theory that a brain alone would not suffice to generate emotions. In their opinion, bodily changes (e.g. in the viscera but also by means of facial expressions) are not the result but the necessary precursor of felt emotions. This is often summarized by: We don't cry because of feeling sad, but we feel sad because we cry. With a series of experiments this strict—and for most people contra-intuitive—sequence of body-cognition dynamics was criticised and refined, most prominently by Walter Cannon (1927). The original theory of James (1884) as well as an outline of the neo-jamesian theories are presented next.

The James-Lange-Theory of emotions

In the first part of his influential article, James (1884) explicitly limits his theory to so-called "standard emotions", which are characterized by "distinct bodily expressions" such as facial expression or quickening of pulse or breathing. He acknowledges the existence of emotions—the non-standard emotions one might say—that are assumed to be "bound up with mental operations, but having no obvious bodily expression for their consequence [..]." This important limitation is often disregarded in later discussions of the theory. The non-standard emotions are seen as the product of "processes in the ideational centres exclusively" that reside within the brain. For example, the "intellectual delight" or "torment" are assumed to occur after a problem is solved or has to be left unfinished. In Section 2.2 this distinction is reconsidered in the discussion of the two different classes of emotions, primary and secondary.

The standard emotions, namely "surprise, curiosity, rapture, fear, anger, lust, greed, and the like" (James 1884, p. 189), are proposed to purely result from the perception of bodily changes (cf. Figure 2.1). These changes directly follow the perception of the exciting fact in form of reflexes that are based on predispositions of the nervous system, so-called "nervous anticipations" (James 1884, p. 191). James supports his view of innate predispositions as the origin of bodily arousal with Darwin's studies on emotion expression (presumably Darwin

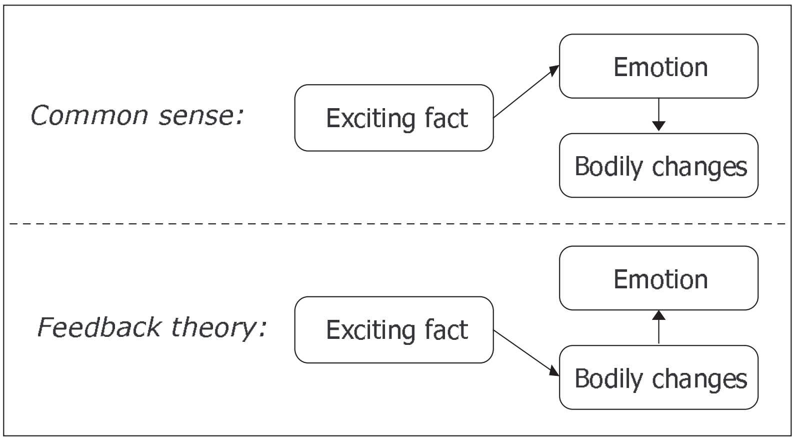


Figure 2.1: James's reversal of common sense and his "feedback theory" (adapted from (Parkinson et al. 2005, p. 5))

(1898)). Interestingly, he also mentions a "mental mood" as the outcome of even the slightest emotional reverberation:

> "[T]he various permutations and combinations of which these organic activities are susceptible, make it abstractly possible that no shade of emotion, however slight, should be without a bodily reverberation as unique, when taken in its totality, as is the mental mood itself." (James 1884, p. 192)

Although James did not work out the details of this differentiation between mood and emotions, it is important to note that the idea of mood being influenced by emotions appears already in such early psychological writings.

His claim of emotions as felt bodily changes contradicts common sense (cf. Figure 2.1) and James himself discussed the following possible objections:

1. If the emotion is nothing but the feeling of the reflex bodily effects of its "object" by means of connate "nervous anticipations" (see above), it can be objected that "most of the objects of civilized men's emotions are things to which it would be preposterous to suppose their nervous systems connately adapted." (James 1884, p. 194)

2. "Is there any evidence [..] for the assumption that particular perceptions *do* produce widespread bodily effects by sort of immediate physical influence, antecedent to arousal of an emotion or emotional idea?" (James 1884, p. 196)

3. "[A]ny voluntary arousal of the so-called manifestations of a special emotion ought to give us the emotion itself." (James 1884, p. 197)

4. "Since musical perceptions, since logical ideas, can immediately arouse a form of emotional feeling [..] is it not more natural to suppose that in the case of the so-called 'standard' emotions [..] the emotional feeling is equally immediate, and the bodily expression something that comes later and is added on?" (James 1884, p. 201)

Concerning the first objection James emphasizes the varying social environments in which humans are subject to development. He further proposes that during phylogenesis the emotion's eliciting conditions might have changed from direct perceptions of vitally important events, such as the offering of food or the threatening with a knife, to more abstract types of rewards and punishments, e.g., being awarded a honorary degree or getting cut in the street. He summarizes:

> "What the action itself may be is quite insignificant, so long as I can perceive in it intent or *animus*. *That* is the emotion-arousing perception; [..]" (James 1884, p. 196)

This notion of perceived intention can be interpreted as a kind of fast schematic appraisal and is also found in recent cognitive theories of emotions (e.g. Ortony, Norman & Revelle 2005; Scherer 1984) discussed in Section 2.1.3. These appraisal schema are understood as the product of phylo- and ontogenetical development and the inclusion of these social factors of emotional development is discussed in Section 2.2.

The other three objections were later also brought up in comparable terms by Cannon (1927)[2], who proposed an alternative theory. Notably, James himself was already proposing to conduct empirical studies in order to falsify his feedback theory and many researchers in the beginning of the 20th century followed his advice.

The neo-jamesian theories of emotion

Cannon's critic was widely accepted to speak against the James-Lange-Theory and, consequently, the idea of bodily feedback as a necessary and in James's opinion also sufficient condition for the elicitation of felt emotions was not further investigated. In the 1960s a new kind of feedback theory was worked out by different scientists and is today commonly labeled "facial feedback hypothesis" (McIntosh 1996). According to this hypothesis, facial expressions and not visceral changes are seen to be a necessary or at least possible factor in emotion elicitation. In his comprehensive discussion McIntosh (1996) states four questions, which refer to the four common general proposals related to facial feedback. Three of them are discussed next.

Does facial configuration *correspond* to emotions? Based on studies using facial electromyography (EMG) several researches provided evidence that facial expressions not only consistently change together with particular emotions, but also predict self-reported emotions. Most notably, the well-known studies of Ekman, Friesen & Ancoli (1980) led to the proposal of so-called "basic emotions", which have been frequently criticized and refined later on (Ekman 1992, 1994; Ortony & Turner 1990). Ekman (1999a) supports his theory of a set of distinguishable "basic emotions" with culture-invariant "distinctive universal signals". According to Ekman (1999b), distinct facial expressions were found for the six basic emotions happiness, anger, fear, sadness, disgust and surprise.

Ekman analysed the results of seven independent studies in which the members of 31 different groups in 21 countries were asked to select one emotion term from a short list of six to ten emotion terms translated to their own language to label static facial expressions. Due to

[2] A comprehensive discussion of Cannon's critic can be found in (Meyer, Reisenzein & Schützwohl 2001).

this procedure, Ekman's basic emotion theory has to be classified as a semantics-based theory, starting from language terms and not from a phenomenon as in the case of the James-Lange-Theory.

The proposed set of basic emotions, however, is in principle not limited to these six basic emotions. First, Ekman presents a list of eleven "characteristics which distinguish basic emotions from one another and from other affective phenomena" (Ekman 1999a, p. 56f) and then explicitly does not allow for "non-basic" emotions and clarifies that "all the emotions which share the characteristics I have described are basic." For the aim of this thesis only the six basic emotions explained above are important and the question of how many other emotions might exist and be classified is postponed to Section 2.1.2.

Does facial movement *modulate* emotions in the presence of other emotional stimuli? If an emotion is already stimulated one might ask whether the accompanying facial display also feeds back on the emotional experience itself. In contrast to the question above only this effect could be labeled "facial feedback". McIntosh (1996) notes that this feedback could manifest in two different ways: either the intensity of a prevailing emotion could be changed or the quality of the felt emotion itself. Most studies, however, concentrate on the intensity effects.

Strack, Martin & Stepper (1988), for example, asked their subjects to hold a pen in their mouths while reading a cartoon. One group was advised to hold the pen with their teeth resulting in a facial configuration similar to a smile while the members of other group had to use only their lips such that a facial expression is provoked that is similar to a sad face. Members of a control group had to hold the pen in the non-dominant hand during reading the cartoon.

The results not only show the postulated influence of facial configuration on felt emotions, but also that this facial feedback operates only on the affective and not on the cognitive component of humor response. This interpretation once again suggests to distinguish at least two components in emotion simulation: a bodily-grounded, affective component and a cognitive component.

Is facial action *necessary* for the presence of emotions? McIntosh gives a very good counter-example for this strong claim: "People experience emotions during times of facial paralysis, most commonly in REM dreaming when there is striate muscle paralysis." (McIntosh 1996, p. 131) But he also mentions the possibility that the central nervous system (CNS) representations of facial expressions could already be sufficient without the need to produce actual facial motion. This idea is supported by the work of Damasio (1994), which is discussed in Section 2.2.

Conclusion

The necessity or—in the extreme—sufficiency of bodily feedback in the elicitation process of emotions is not finally being agreed upon. The previously outlined ongoing discussion has first led to a much weaker position concerning bodily feedback, namely, that it is supportive for felt emotions. During the investigation of the processes, however, two aspects recurred that are of special interest to this thesis:

1. Often some longer lasting, diffuse aspect of emotional experience is mentioned and consistently labeled *mood*. It is always considered as an influencing factor in emotion elicitation and often associated with bodily states or processes such as general arousal level.

2. On the one hand, the initially introduced class of “standard emotions” (James 1884, p. 189) affords a principal distinction of at least two classes of emotions. On the other hand, Ekman’s “basic emotions” (developed in the context of the “facial feedback hypothesis”) are described as independent seed crystals that are the product of evolution.

The second aspect suggests to further investigate how else an emotion can be conceptualized, if it is not sufficient to rely on distinctive patterns of bodily feedback. One might be tempted to ask: Is there a set of emotions that are more “basic” than others? Are non-basic emotions—if existent—to be described as mixtures of basic ones? These questions provoked an ongoing debate over the last 25 years and some aspects of this debate are discussed next.

2.1.2 Basic emotions and dimensional theories

Ortony & Turner (1990) distinguish two conceptions underlying the assumption that emotions can be grouped into basic (or primary, fundamental) ones and non-basic (or secondary) ones: biological primitiveness based on the evolutionary origin of basic emotions and psychological primitiveness, that is, basic emotions as “irreducible constituents of other emotions.” (Ortony & Turner 1990, p. 317) The previously described theory of Ekman (1999a) is based on the first conception[3]. The second conception is also called the “palette theory of emotions” (Scherer 1984), because basic emotions are comparable to a set of basic colors out of which other secondary emotions/colors are mixed. Most proponents of basic emotions, however, do not subscribe themselves to only one of the two conceptions but rather argue that these conceptions support each other.

Starting with McDougall (1919) the conception of basic emotions being understood as psychologically primitive building blocks has found a number of proponents (Meyer, Schützwohl & Reisenzein 2003; Reisenzein 2000a). In the following Plutchik’s theory is examined as a representative of this class of emotion theories.

Primary emotion	**Basic behavioral pattern**
Acceptance	Incorporation
Fear	Protection
Surprise	Orientation
Sadness	Reintegration
Disgust	Rejection
Anger	Destruction
Anticipation	Exploration
Joy	Reproduction

Table 2.1: Eight primary emotions and their underlying prototype functional patterns of behavior

[3] A further discussion of the ontogenetical aspects of emotions is presented in Section 2.2.2.

A structural model of emotions

Plutchik (1980) first distinguishes “eight basic prototype functional patterns of behavior” (Plutchik 1980, p. 152) that are the product of evolution and give rise to eight basic—or in his own terms primary—emotions. He then proposes to use introspective language to name his eight primary emotions as given in Table 2.1.

To this respect he follows the first conception mentioned above assuming biological primitiveness. In his further explanations he compares his conception of emotions with color representation in three-dimensional space of hue, saturation and intensity/value (cf. Figure 2.2, right), which is a common concept in computer science (Schwarz, Cowan & Beatty 1987).

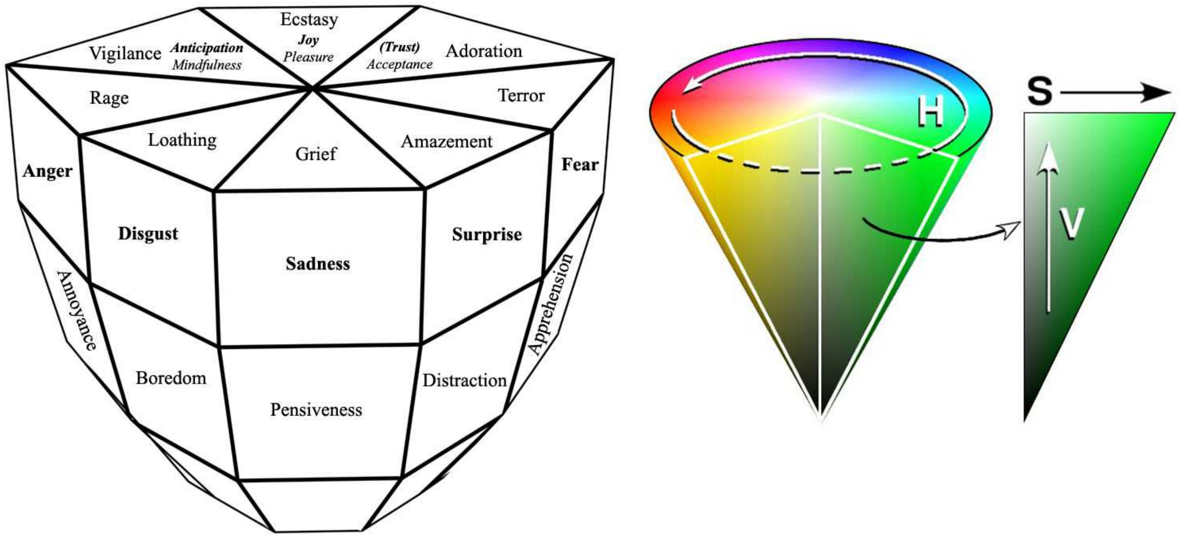


Figure 2.2: Plutchik’s three-dimensional structural model of emotions (left, after (Plutchik 1980, p. 157)) compared to the HSV color space (right, after Wikipedia (2008))

With respect to his idea of “primary emotions” Plutchik proposes the following analogy:

> “[I]t is necessary to conceive of the primary emotions as analogous to hues, which may vary in degree of intermixture (saturation) as well as intensity. The primary emotions vary in degree of similarity to one another, just as do colors. Emotions also have the property of bipolarity, or complementarity, as do colors.” (Plutchik 1980, p. 153)

These eight emotions correspond to the eight “primary emotion dimensions” that are arranged to each other on the basis of bipolarity and similarity. Plutchik (1980) refers to the underlying behavioral response tendencies to explain, for example, that “anger” and “fear” are bipolar, because anger leads to attack and fear to withdrawal. Consequently these two primary emotions lie on opposite sides of the emotion cone presented in Figure 2.2, left.

When following the intensity axis (labeled “V” for value in Figure 2.2, right) from bottom to top the diversity of emotions is assumed to increase together with higher intensity. Starting with low intensity and a hue of disgust, for example, results in boredom. Higher intensity given the same hue leads to loathing. The proposed effect of saturation, however, remains underspecified, because the closer one gets to the center of the cone the less distinct an emotion is. In case of zero saturation and maximum intensity (respectively value) all primary emotions are equally involved and this emotional state is called “conflict” as represented by “C” in Figure 2.3.

Emotion compounds How to decide for similarity of emotions depends on the types of measures used, which may be based on facial expressions or subjective feeling of perceived emotions. The concept of “dyads” is introduced (Plutchik 1980, p. 161) to refer to mixtures of any two types of primary emotions in a similar fashion as new colors can be mixed out of two basic colors. Primary dyads result from the mixture of two adjacent primary emotions, secondary dyads are built out of primary emotions that are once removed on the circle and tertiary dyads result if the primary emotions are twice removed. In Figure 2.3 all primary dyads are given outside of the circle.

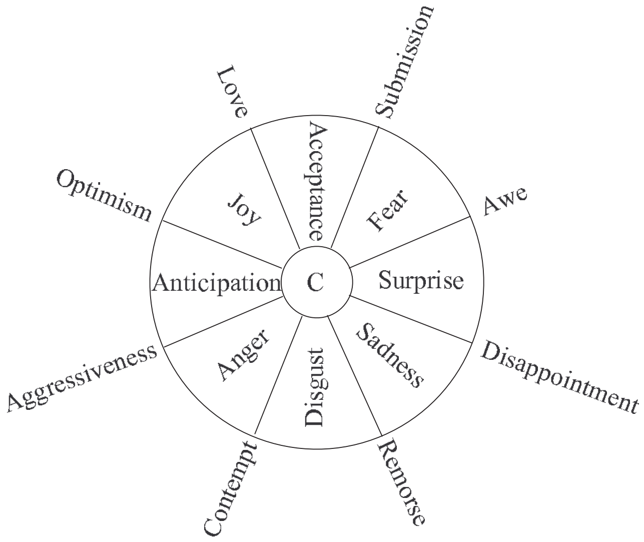


Figure 2.3: Primary dyads formed by the combinations of adjacent pairs of primary emotions (Plutchik 1980, p. 164)

At this point the problem of naming becomes even more obscure. The primary dyad that results from fear and acceptance is labeled submission (cf. Figure 2.3), a term that refers to personality related aspects in social psychology (see below). Plutchik is well aware of this difficulty and states:

> “Perhaps our language does not contain emotion words for certain combinations, although other languages might. Certain combinations may not occur at all in human experience, just as chemical compounds can be formed only in certain limited ways. [..]
>
> One other important point might be made about [..] a problem almost identical with that [of developing] a system for the numerical specification of what a color looks like to the ordinary man or woman. [..] The average data from a small number of selected observers provided an imaginary standard observer and all results [..] are adjusted so as to satisfy the requirements of this standard observer.

> [..] Perhaps a similar system may be developed for the psychology of emotions." (Plutchik 1980, p. 161)

With these statements Plutchik reveals his motivation for suggesting a finite set of basic emotions. He believes that his emotion model together with the proposed combinatorial method covers all aspects of emotional life. This, however, makes it necessary to fill some gaps with such non-emotional or at least questionable concepts as anticipation and surprise. A table with lists of basic emotions that were proposed by different psychologists during the last century (Ortony & Turner 1990, p. 316) shows how different the proposed sets of basic emotions are. Also the basis for inclusion differs considerably among the different theories. For James (1884) the basis for inclusion is listed as "bodily involvement" where as in case of Ekman et al. (1980) it is "universal facial expression" (see also Section 2.1.1). As Plutchik's list of eight fundamental emotions is based on the "relation to adaptive biological processes" the divergence to the other sets of basic emotions is explainable.

Summary Although Plutchik's model is debatable some of the underlying ideas are agreed upon and can, thus, be found in other theories as well. In Table 2.2 three aspects are described that are further elaborated.

Aspect	**Description**
Intensity	The intensity of an emotion is often disregarded or at least not as important as it should be in other models of emotion.
Mixed emotions	Although it is not agreed that secondary emotions can be described in terms of a mixture of primary ones, it is nevertheless agreed that two or more emotions, primary or secondary, may coexist at any given moment in time.
Basic dimensions	The idea of identifiable basic dimensions that are underlying emotions is followed in this thesis as well, but these dimensions do not correlate with some kind of fundamental emotions.
Bipolarity	It will be argued that the emotional dimensions introduced next are also bipolar.

Table 2.2: Positive aspects of Plutchik's structural model of emotions

What the structural model misses is the temporal development of emotions, in terms of actual development at a given moment in time as well as ontogenetical development. This emotion dynamics is investigated next in the context of another kind of dimensional theories.

Wundt's three-dimensional theory

Before William James (1884) brought up his idea of bodily feedback as discussed in Section 2.1.1, Wilhelm Wundt (1863) already argued that nothing would be more incorrect as to understand emotional life as the sum of essentially unchangeable elementary feelings. According to (Wundt 1863, p. 243), the qualitative richness of feelings results from mutual interaction of simultaneous as well as consecutive feelings and, therefore, is in principle inexhaustible.

The original idea Wundt (1863) used the psychological method of "introspection", which was disregarded in the beginning of the 20th century for being too much based on subjectivity to allow for scientifically valuable insights. This difference in methodology explains the discrepancy to newer emotion theories, for example, Ekman's theory of basic emotions detailed above. Ekman (1999a) is focusing on the inter-subjectively accessible behavior to derive his set of six basic emotions whereas Wundt concentrates on a feelings subjective quality that is experienced through introspection.

Axis	**Description**
1. *pleasure* ↔ *displeasure*	Quality or hedonic valence of emotional experience (Lust ↔ Unlust)
2. *excitement*↔ *inhibition*	Level of (physiological) arousal or (neurological) activation accompanying an emotional experience (Erregung ↔ Beruhigung)
3. *tension* ↔ *relaxation*	Temporal aspect of the emotion eliciting event (Spannung ↔ Lösung)

Table 2.3: Wundt's three principal axes together with their elementary feelings

In Wundt's words, nothing would be more misleading than describing emotional experience as "the sum of essentially invariable elementary feelings." (Wundt 1863, p. 243) To this respect he does not follow the distinction of basic and nonbasic emotions as discussed above. Wundt's theory belongs to the class of phenomena-based theories (as introduced in the beginning of this chapter on page 15), because his "distinction of elementary and nonelementary feelings is purely *phenomenological* in character [..]." (Reisenzein 1992, p. 144). He further analyzes emotional experience and postulates a subjective feeling state, which Russell (2003) labels "core affect".

To capture the subjective feeling state Wundt introduces the concept of a so-called "total feeling"[4] as the momentary mixture of potentially conflicting feeling states and considers it to consist of a certain quality and intensity (Wundt 1863, p. 239). Elementary feelings, in the contrary, are assumed to constitute the three principal axes described in Table 2.3[5], which form an orthogonal, three-dimensional emotion space presented in Figure 2.4.

A momentary emotion is represented within this three-dimensional emotion space by a single point. A concrete event, however, always results in a "certain, continuous course of feeling" and in principle describes a trajectory that "represents the feeling state in any given moment"[6] (Wundt 1863, p. 245). It most often starts and ends in the point of origin.

The exemplary course of feeling indicated in Figure 2.4 begins with an increase of excitement, displeasure and tension. Then a phase of decreasing excitement is accompanied by increasing pleasure, before relaxation (as indicated by the dotted part of the curve) leads back to the point of origin. Also assumed possible are courses of emotion that continue another course, which did not finish at the point of origin.

[4] German: "Totalgefühl"

[5] Translations taken from Reisenzein (1992)

[6] German: "Indem ein einzelner Punkt nur ein momentanes Gefühl bezeichnet, wird aber irgendein konkretes Geschehen immer in einem bestimmten, stetig zusammenhängenden Gefühlsverlauf bestehen und im allgemeinen durch eine Kurve dargestellt werden können, die für jeden Augenblick die Gefühlslage angibt."

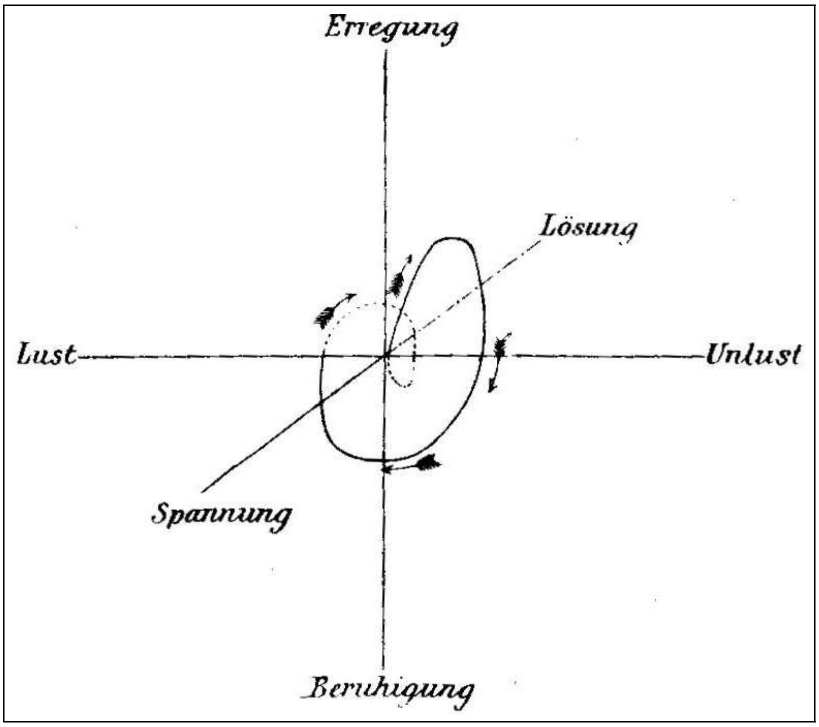


Figure 2.4: The three principal axes of orthogonal emotion space (Wundt 1863, p. 246)

Notably, Wundt does not explain where single emotions are to be located in this abstract emotion space and he also does not tell how an emotion's intensity might be determined given that some emotionally relevant event or object is perceived. In Reisenzein's structuralists reconstruction of Wundt's theory (Reisenzein 1992, 2000b) these missing features are discussed and a number of solutions presented. Reisenzein (1992) gives an informal description of Wundt's theory of emotion taking Wundt's later writing into account. During the discussion he explains that Wundt adopted a "dualistic view of the elements of consciousness" (Reisenzein 1992, p. 143) in that he proposed the existence of two kinds of psychic elements resulting from psychological analysis: sensory elements or sensations (e.g. touch, tone, heat or light) and affective elements or simple feelings (e.g. sensory pleasure or displeasure possibly accompanying simple sensations). Reisenzein further explains:

> "All nonelementary conscious experiences were viewed by Wundt as complexes or compounds of these kinds of psychic elements. Complexes of sensory elements were called ideas (*Vorstellungen*); complexes of feeling elements, emotions (*Gemütsbewegungen*). Three subtypes of emotions were distinguished: Compound feelings (*zusammengesetzte Gefühle*), affects (*Affekte*), and volitions (*Willensvorgänge*). Whereas compound feelings are products of a momentary state [see 'momentary emotion' introduced above], affects and volitions are mental processes, that is, characteristic, temporally extended sequences (*Verlaufsformen*) of (compound) feelings (see also Wundt 1863, p. 99)." (Reisenzein 1992, p. 143)

This idea of two independent psychic elements or components was taken up again many decades later. Zajonc (1980) refers to Wundt when he proposes separate and partially independent systems, which control affect and cognition and are influencing each other in a variety of ways. He summarizes Wundt's idea with the label "affective primacy idea" (Zajonc 1980, p. 152) highlighting the assumed precedence of affect before cognitions.

Furthermore, in Zajonc's interpretation, it is this independent and parallel process of affect generation outlined by Wundt that turns cold cognitions into hot ones. He explicitly concentrates on the "class of feelings" that are "involved in the general quality of behavior that underlies the approach-avoidance distinction." To this respect his approach is much less differentiated than Plutchik's structural model of emotions, which is based on eight basic behavioral patterns listed in Table 2.1 on page 20. Consequently, Zajonc admits ignoring "other emotions such as surprise, anger, guilt, or shame" and the like. In summary, he presents a considerable amount of empirical findings that let him argue against treating affect "as unalterably last and invariably post-cognitive." (Zajonc 1980, p. 172) To explain the automatism with which affective responses are generated, he refers to Freud's work on the unconscious. The distinction between conscious and unconscious processing—important for this thesis as well—is reconsidered in Section 2.1.3.

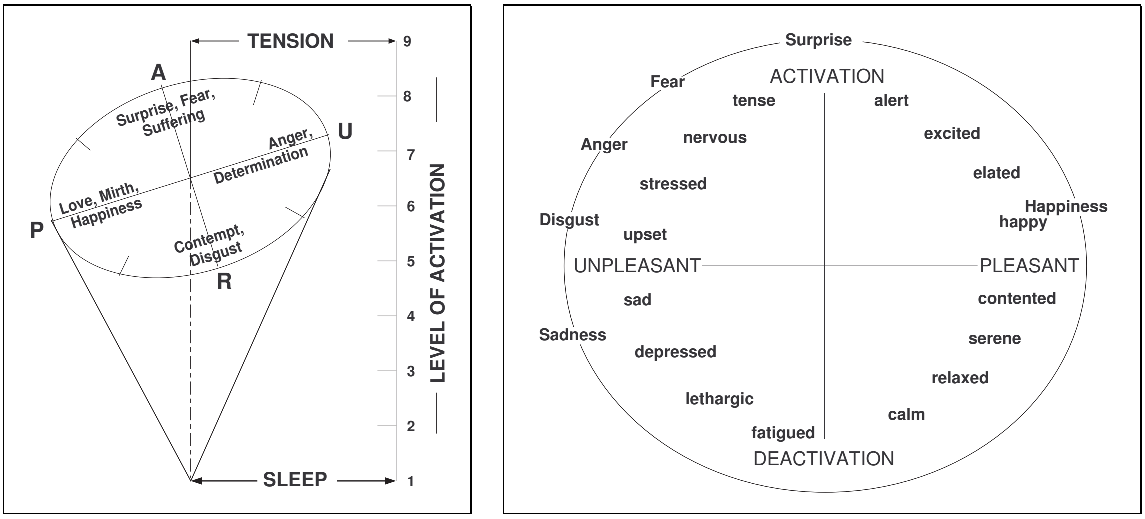
Other dimensional theories

Even before Zajonc's considerations of "affective primacy", Schlosberg (1954) closely examined the *activation* dimension (cp. second axis in Table 2.3, p. 24) in his proposal of "three dimensions of emotion". The three dimensions pleasantness–unpleasantness, level of activation, and attention–rejection form a three dimensional space as presented in Figure 2.5(a). Based on ratings of emotional pictures he finds that unpleasantness is correlated with higher arousal than pleasantness. Mirth is located at an intermediate level whereas contempt is believed to combine pleasantness with rejection and consists of rather low activation. The third axis "activation" is considered necessary to distinguish "some expressions that are not separated by the original two axes; for example, grief, pain, and suffering all have the same P-U and A-R values, but grief is considerably below the other two expressions in level of activation." In general, the activation dimension ranges from sleep at its low end, over alert attention at its middle, to strong emotions at its high end.

The similarity of Schlosberg's concept to Plutchik's structural theory of emotions (cp. Figure 2.2, p. 21) is apparent and also Schlosberg compares his activation dimension with the intensity dimension of color space. Schlosberg, however, does not propose a fundamental set of basic emotions, but rather arranges a theoretically derived set of dimensions—similar to Wundt—in such a fashion that a cone-shaped space of subjective feeling is formed.

Concerning the possibilities to detect activation and the problems to detect any other dimension, Schlosberg (1954) summarizes:

> "Neither skin conductance nor any other physiological measure [..] has yet given us much beyond the intensitive dimension. Further research may furnish such evidence, but for the present we may profitably turn to facial expression to find the qualitative dimensions along which emotion may vary. Here, we have good evidence that the whole range of expressions may be described rather well in terms of a roughly circular surface, whose axes are pleasantness-unpleasantness and attention-rejection. We have some idea how level of activation comes into this figure as a third dimension, but further research is needed here, too." (Schlosberg 1954, p. 87f.)



(a) “Three dimensional figure” (after (Schlosberg 1954, p. 87)) formed by the dimensions “pleasantness–unpleasantness”, “attention–rejection”, and “level of activation”

(b) The circumplex model of core affect (after (Russell & Feldmann Barrett 1999, p. 808)) formed by the dimensions of pleasant-unpleasant and activation-deactivation

Figure 2.5: Schlosberg’s three dimensional figure of emotional expression and Russel’s two dimensional circumplex model of core affect.

Further evidence for a circular, two-dimensional model of emotions was provided by Russell (1980). His circumplex model of core affect (cf. Figure2.5(b)) consists of the two dimensions pleasant-unpleasant and activation-deactivation. Notably, he argues against the necessity of a third dimension and claims that the second dimension is that of activation-deactivation and not that of attention-rejection as postulated by Schlosberg (1954). This interesting difference is backed up by Russell (1980) with a number of later studies, in which the two separate dimensions attention-rejection and activation were often statistically indistinguishable. Further support for the importance of the two dimensions pleasantness and activation comes from Osgood, Suci & Tannenbaum (1957), who conducted studies on the measurement of meaning in natural language. They found the dimensions evaluation, activity and power to be major components of meaning. The third dimension is subject to ongoing discussions in psychology, because different studies relying on different methods found different interpretations concerning the meaning of the third dimension.

One might now be tempted to ask why such a third dimension is needed at all?

Russell & Mehrabian (1974) examined the difference between “anger” and “anxiety” (the last of them being quite similar to the emotion “fear”) to argue for a third dimension labeled *dominance*. In their study their subjects at first had to read a description of a situation. After imagining themselves to “actually [being] there” and getting “into the mood of the situation” (Russell & Mehrabian 1974, p. 80), they had to rate their feelings on 21 adjective pairs measuring emotional states.[7] Russell & Mehrabian (1974) hypothesized that the difference between anger and anxiety could be found in reported level of dominance. By means of regression analysis they found that anger has a significantly positive amount of dominance (+.09) and

[7] Due to this method, the resulting three-dimensional emotion theory is most likely to be attributed to the class of semantics-based theories as introduced in the beginning of this chapter.

anxiety a significantly negative amount of dominance (-.11). Pleasure and arousal, however, were both equally signed; -.74 pleasure and +.36 arousal for anger, -.54 pleasure and +.49 arousal for anxiety. In the discussion of their results they state:

> "These data provide direct support for all aspects of the proposed hypotheses: both anger and anxiety contain high arousal and low pleasure. The distinction between anger and anxiety lies along the dominance dimension: anger involves high dominance, anxiety involves submissiveness. The smaller magnitudes of the coefficients for dominance [..] are partially due to our reliance on the physical qualities of situations to vary dominance-submissiveness feelings. Social situations contain greater variations along dominance-submissiveness and thus provide a better test of the hypothesized effects." (Russell & Mehrabian 1974, p. 81f.)

When comparing this argumentation to the locations of "Anger" and "Fear" in Russell's circumplex model (cf. Figure 2.5(b)), the difficulty in representing emotions in a space of only two dimensions is apparent. Fear and anger lie relatively close to each other in pleasure-activation space. In pleasure-attention space, however, (cf. Figure 2.5(a)) the same two "basic" emotions are much better distinguishable.

A further investigation of the three dimensions pleasure-displeasure, degree of arousal, and dominance-submissiveness, undertaken by Russell & Mehrabian (1977) yielded evidence that they "are both necessary and sufficient to adequately define emotional states." (Russell & Mehrabian 1977, p. 273) They report on the replication of their previous findings that "anger (hostility, aggression) involved a feeling of dominance, whereas anxiety (fear, tension) involved a feeling of submissiveness." (Russell & Mehrabian 1977, p. 282) Furthermore, they take these facts as "especially important in establishing the necessity for the dimension of dominance-submissiveness for a comprehensive description of emotional states [..]." This study was not limited to the two emotions "anger" and "anxiety" and accordingly they present a table of 151 terms denoting emotions. A selection of these terms is presented in Table 2.4.

The emotion terms written in italics in Table 2.4 are of special interest to this thesis, because they are quite similar to Ekman's proposed set of six basic emotions (cp. Section 2.1.1). The emotion "disgusted" (number 75 in Table 2.4) is considered less important for the Affect Simulation Architecture of a purely virtual embodied agent.

In Figure 2.6 the six emotions "Happy", "Anxious", "Surprised", "Angry", "Fearful", and "Sad" are located in the three-dimensional emotion space, which is spanned by the bipolar dimensions "pleasantness-unpleasantness" (labeled +P and -P), "arousal-sleepiness" (labeled +A and -A), and "dominance-submissiveness" (labeled +D and -D).[8] The dominance values of emotions represented by a circle in Figure 2.6 are marked in Table 2.4 with a star, because they do not differ significantly from zero. This is especially interesting for emotion number 50, "Anxious", which was argued to bear a significant degree of submissiveness in the study before (cf. Russell & Mehrabian (1974)). Number 52, "Surprise", is—once again—a questionable emotion term, because its dominance value does not differ significantly from zero, as indicated by the circle in Figure 2.6. All other four emotions were scaled up or down to the top or bottom of the dominance axis according to the sign given in Table 2.4 for their dominance values. Thus, "Happy" and "Angry" are the only emotions with positive dominance and "Fearful" as well as "Sad" have negative dominance values or, to state it in other terms, come along with a feeling of submissiveness.

[8] From now on the term "PAD space" will be used to refer to this emotion space.

	Pleasure		Arousal		Dominance	
Term	Mean	SD	Mean	SD	Mean	SD
20. Joyful	.76	.22	.48	.26	.35	.31
24. Friendly	.69	.23	.35	.28	.30	.27
31. *Happy*	.81	.21	.51	.26	.46	.38
41. Enjoyment	.77	.17	.44	.26	.42	.29
50. *Anxious*	.01*	.45	.59	.31	-.15*	.32
52. *Surprised*	.40	.30	.67	.27	-.13*	.38
59. Relaxed	.68	.30	-.46	.38	.06*	.49
75. Disgusted	-.60	.20	.35	.41	.11*	.34
82. *Angry*	-.51	.20	.59	.33	.25	.39
84. Enraged	-.44	.25	.72	.29	.32	.44
93. Cold anger	-.42	.29	.67	.27	.34	.44
96. Frustrated	-.64	.18	.52	.37	-.35	.30
97. Distressed	-.61	.17	.28	.46	-.36	.21
101. *Fearful*	-.64	.20	.60	.32	-.43	.30
120. Angry but detached	-.42	.22	.28	.41	-.03*	.33
121. Confused	-.53	.20	.27	.29	-.32	.28
126. Depressed	-.72	.21	-.29	.44	-.41	.28
132. Bored	-.65	.19	-.62	.24	-.33	.21
151. *Sad*	-.63	.23	-.27	.34	-.33	.22

Table 2.4: A selection of terms denoting emotions in terms of pleasure, arousal, and dominance (Russell & Mehrabian 1977, p. 286ff). Emotion terms in *italics* are further discussed in the text.
* The mean does not differ significantly $(p < .01)$ from 0.0.

Although this representation might appear convincing, the relatively high values of standard deviation (labeled SD in the respective columns of Table 2.4) are problematic. Gehm & Scherer (1988) critically observe "that any kind of factor analytic or multidimensional scaling technique depends almost exclusively on the kind of material that is put into the analysis for its outcome." (Gehm & Scherer 1988, p. 100) Consequently, in their study a "fairly comprehensive list" of 235 German emotion-describing adjectives was used. Furthermore, they highlight the importance of intra- and inter-individual differences in the nature of the semantic emotion space. Using clustering techniques similar to those applied by Russell (1980), they found that "the degree of inconsistency increases with age and that subgroups of participants with similar education tend to judge similarly." (Gehm & Scherer 1988, p. 105) With regard to the circumplex model of Russell (1980) they state:

> "Although Russell repeatedly found a rather systematic structure (a circumplex model) of the 28 items he investigated, we could in no case replicate his findings with our more comprehensive list of items: Neither the configuration of the total sample or the subsamples nor the adjectives used by Russell himself were ordered circulary in our study." (Gehm & Scherer 1988, p. 106)

The results of their multidimensional scaling yields evidence for two major dimensions as well, but Gehm & Scherer (1988) propose to label these dimensions with "hedonic valence"

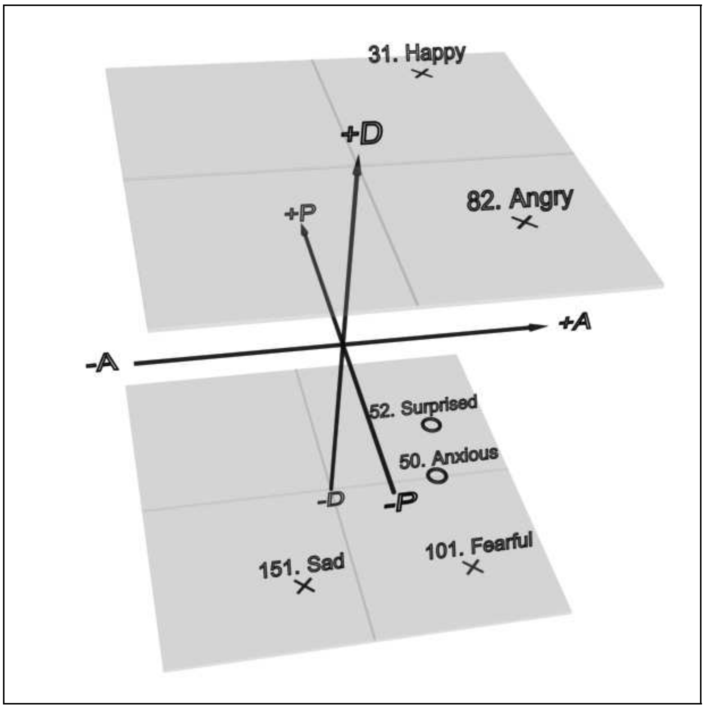


Figure 2.6: Three-dimensional emotion space (PAD space, in short) formed by the dimensions “pleasantness-unpleasantness” (+P, -P), “arousal-sleepiness” (+A, -A), and “dominance-submissiveness” (+D, -D) (as proposed by Russell & Mehrabian (1977)) together with six emotions of Table 2.4

and “power/control” (Gehm & Scherer 1988, p. 108). Their lack of identification of an independent activation dimension is assumed to result from an item selection criterion: Synonyms and adjectives expressing slight differences in intensity were deliberately excluded. It is therefore not surprising that Gehm & Scherer (1988) propose a tetrahedral model of emotion space (cf. Figure 2.7), in which the activation dimension is interpreted as the connection between the two other dimensions.

The first dimension is labeled “hedonic valence” and ranges from “predominantly unpleasant” (at point A) to “well being” (at point B). High level of control/power[9] leads to “happy excitement” (at point D) whereas low level of control/power is labeled “conflict” in Figure 2.7. The third dimension of “activation” is spanned between the orthogonal edges of the previous two dimensions with adjectives closer to the hedonic valence dimension containing lower and adjectives close to the control/power dimension higher activation. Gehm and Scherer’s tetrahedral model is, thus, more similar to Schlosberg’s emotion cone (cf. Figure 2.5(a), p. 27) than to Russell and Mehrabian’s three dimensional model (cp. Figure 2.6).

Recently, Scherer, Dan & Flykt (2006) have pointed out that most researchers proposing dimensional emotion representation did not discuss the processes that underly their subjects’ ability to rate the emotionally relevant adjectives or pictures. How does a human appraise a given object or situation? Are there different levels of processing accounting for different

[9] This dimension is also labeled “dominance” by Gehm & Scherer (1988) in accordance with Russell & Mehrabian (1977). In the context of Scherer’s Component Process Model of emotions (cf. Scherer (1984), Scherer (2001)) discussed below, however, control and power are interpreted as two independent components in the appraisal process.

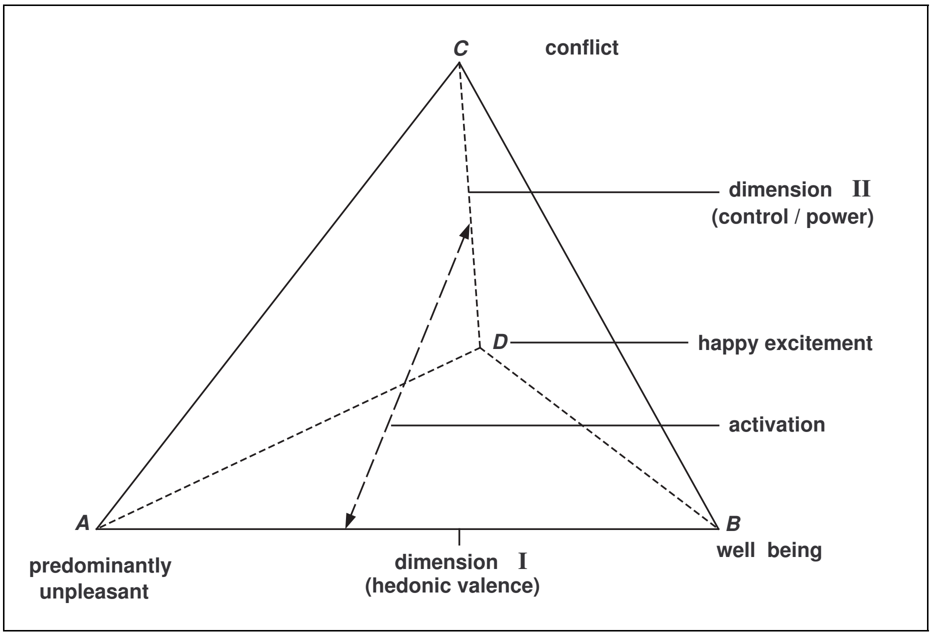


Figure 2.7: Tetrahedral model of subjective emotional space (Gehm & Scherer 1988, p. 112) formed by the dimensions of hedonic valence (from B to A), control/power (from D to C), and activation (connection of the two orthogonal edges)

types of emotions? What timescales can be distinguished and on which basis? After some concluding remarks on basic emotions and dimensional theories, these questions are discussed in the light of appraisal theories of emotions in the next section.

Summary and Conclusion

The previous discussion shows that the arguments in favor of a "palette theory of emotions" as proposed e.g. by Plutchik (1980) and explained in Section 2.1.2 are relatively weak. Although it seems plausible to assume some kind of biological primitiveness underlying emotional behavior, this behavioral basis must not be overemphasized. Often, for example, it is better not to run but to stay with the group when facing danger and experiencing fear. The idea of basic dimensions adequately capturing basic elements of felt emotion, however, could find a good number of proponents and critics over the last century.

Hue/Pleasure/Valence dimension The first and most important component is widely agreed upon to denote valence of emotion. An emotion is always either positive or negative. What exactly it is that lets a subject judge a given emotional term or a presented emotional picture as positive or negative is not so clear. Subjects are sometimes instructed to imagine themselves in the described situation (Russell & Mehrabian 1974), in other studies (Gehm & Scherer 1988; Russell 1980) they have to rate different selections of emotional terms or adjectives. The first approach focuses on the subject's subjective feeling state, in the latter case a more cognitive appraisal of a given term's emotional connotative meaning is acquired. These differences play an important role in the context of "appraisal theories" discussed in the next section.

Ever since Schlosberg (1954) it is still undecided how to interpret and label the further dimensions discussed next.

Intensity/Activation/Arousal/Excitement dimension The level of physiological arousal or neurological activation is mostly regarded as the second component of emotion space. For Russell & Feldmann Barrett (1999) no further dimension is needed to capture the constituents of subjective feeling. Compared to Plutchik (1980) we find an interesting similarity here. Although Plutchik's first dimension "hue" (cp. Figure 2.2, p. 21) is considered to be composed of eight "basic emotions", its assumed bipolarity yields a similarity to Russell's "pleasantness" dimension (cp. Figure 2.5(b), p. 27). Moreover, Plutchik's "intensity" dimension is similar to Russell's "activation" dimension. The third dimension "saturation" that Plutchik argues for is missing in Russell's circumplex model, because the aforementioned "hue" dimension is only considered one-dimensional in the circumplex model. In other words, one can only "vary in degree of intermixture (saturation)" (Plutchik 1980, p. 163) if assuming an at least two-dimensional basis of eight "primary emotions" (cp. emotion terms in bold in Figure 2.2, left). Russell's "pleasantness" dimension, however, is only one-dimensional, making a third dimension of "hue" incompatible.

Saturation/Attention/Dominance/Control/Power dimension According to both the work of Schlosberg (1954) and Scherer et al. (2006) this dimension is even more important than the activation dimension. Especially in the case of high activation Gehm & Scherer (1988) found that taking the level of control and social power of an individual into account is useful in distinguishing certain emotion-describing adjectives. This finding is supported by Russell & Mehrabian (1977), who could show that anger and fear both consist of similarly high displeasure and arousal values and can only be distinguished due to their different values on the dominance scale (cp. Figure 2.6, p. 30).

Implications for the thesis In the beginning of this section two possible underlying conceptions for the "basic emotions" approach were introduced: biological primitiveness and psychological primitiveness. A closer look at the psychological conception revealed the huge amount of seemingly similar technical terms that unfortunately most often do not denote a sufficiently similar scientific concept.

Plutchik's idea of "primary emotions" being similar to basic colors, which are to be mixed systematically to achieve more complex emotions, is not further followed in this thesis. What is referred to as "primary emotions" in this thesis can best be described as the set of ontogenetically earlier types of emotions that can be expressed by facial expressions in accordance with the six basic emotions of Ekman et al. (1980).

The three dimensional emotion model presented in Figure 2.6 on page 30 consisting of pleasure, arousal and dominance dimensions (PAD space) is adapted for the Affect Simulation Architecture. In this thesis the two dimensions pleasure/valence and arousal/activation are modeled to range from -100 to +100 on a continuous scale. The values for the dominance dimension, however, usually do not significantly differ from zero in Table 2.4 and the interpretation of this dimension's meaning is particularly controversial. Therefore, it is decided to abstain from modeling this dimension on a continuous scale in PAD space. Analogue to the example given in Figure 2.6 only high versus low dominance is distinguished in the Affect Simulation System proposed in this thesis.

After the discussion of general emotion representations and its effects on facial and bodily expressions, the cognitive processes underlying the elicitation of emotions are now investigated. Theories dealing with these aspects of emotions are commonly labeled “appraisal theories” (see Ellsworth & Scherer 2003, for a review) due to their focus on evaluation or appraisal processes that are believed to be necessary at the start of an emotion episode.

2.1.3 Appraisal theories

In common sense an emotion is a reaction to some event after its implication for the self has been assessed by an individual. The term “appraisal” refers to this evaluative process in emotion theory. The subjective significance of an event is believed to be evaluated by an individual against a number of variables. Some of these variables are related to an agent’s goal to protect itself from being harmed or to sustain or achieve pleasurable situations.

In the aim to simulate affect for virtual agents, they must be able to somehow appraise events with respect to their goals and desires in order to start an emotion process. As the previously developed simulation of emotion dynamics (Becker 2003) is limited to quite simple types of emotions, the underlying appraisal process does not need to be very complex (cf. Section 4.2). In case of the Affect Simulation Architecture presented here, however, this appraisal process has to be refined. Consequently, it is necessary to take a closer look at appraisal theory.

Scherer (1999) distinguishes four major strands of theoretical approaches to appraisal based on the nature of their underlying appraisal dimensions.

1. The classical approach is based on the idea that individuals use a fixed set of dimensions or *criteria* to evaluate the significance of events. It goes back to the work of Arnold and Lazarus and is explained here in the context of the work of Scherer (1984).

2. The second approach focuses on the nature of the causal *attribution* involved in emotion-antecedent appraisal. Weiner (1985) proposes such an attributional theory.

3. Taking an agent’s goals as a starting point for emotional appraisal, the goal-relatedness of an event is evaluated by applying specific patterns or *themes* (such as “separation anxiety” (Oatley & Johnson-Laird 1987, p. 41)) in this approach. Oatley & Johnson-Laird (1987) present a “cognitive theory”, which is a representative of this category.

4. As already mentioned in Section 2.1.2, semantics-based theories are mostly interested in analyzing the semantic field of emotion-denoting natural language. With respect to “appraisal theories” this idea formed the basis for the model of emotions proposed by Ortony, Clore & Collins (1988), which is detailed and discussed later in this section.

With a focus on the processes underlying appraisal in humans, Scherer (1984) proposes a detailed model of emotions known as the “Component Process Model”. With respect to the above distinction Scherer (1984) follows the classical approach, because in his analysis he establishes a set of concrete appraisal dimensions. Over the last two decades several empirical studies and theoretical extensions were applied to this model—a brief overview is given next.

A layered process model of emotions

As discussed in Section 2.1.2 above, defining emotions by means of semantic analysis of verbal labels is notoriously difficult. Therefore, Scherer (1984) proposes to focus on the functions that emotions could serve within an individual and in the context of social interaction. In relation to the five components listed in the beginning of Section 2.1 on page 15, Scherer (2001) postulates a relationship between the functions, components and organismic subsystems that he summarizes according to Table 2.5.

Emotion function	Emotion component	Organismic subsystem (and major substrata)
Evaluation of objects and events	Cognitive component	Information processing (CNS)
System regulation	Peripheral efference component	Support (CNS, NES, ANS)
Preparation and direction of action	Motivational component	Executive (CNS)
Communication of reaction and behavioral intention	Motor expression component	Action (SNS)
Monitoring of internal state and organism-environment interaction	Subjective feeling component	Monitor (CNS)

CNS: central nervous system; NES: neuro-endocrine system; ANS: autonomic nervous system; SNS: somatic nervous system. The organismic subsystems are theoretically postulated functional units or networks.

Table 2.5: Relationship between the functions and components of emotion and the organismic subsystems that subserve them (after Scherer 2001, p. 93)

Scherer (1984) further believes that with his functional perspective on the appraisal process much more consensus on the nature of emotion can be achieved than with a conceptual or structural approach[10]. Table 2.5 shows that Scherer takes a broad view on emotions including physiological and expressive aspects, although his theoretical model mostly elaborates on the first component of cognitive stimulus processing.

With respect to the level of consciousness involved in the realization of the different functions listed in Table 2.5, Scherer (2005) distinguishes three overlapping circles that represent different aspects of monitoring (cf. Figure 2.8). The bottommost circle (A) contains all but one of the components listed in Table 2.5, which are assumed to be based on unconscious processes of reflection and regulation. On this level somatosensory feedback together with "massive projections from both cortical and subcortical central nervous system (CNS)" are assumed to be processed (cp. Section 2.1.1). According to Scherer, "one might call the content of the circle [(A)] *integrated process representation*." (Scherer 2005, p. 321) When consciousness is taken into consideration, the second circle (B) becomes relevant representing the quality and intensity of subjective feeling state. Very cautiously Scherer relates the content of this circle to "what philosophers and psychologists have referred to as *qualia*." The content of the topmost circle (C) contains conscious processes enabling an individual to verbalize his or her emotional experience. This verbalization process heavily depends on "(1) the limited

[10] This assumption is reconsidered in the context of the work of Ortony, Clore & Collins (1988), who propose a structural theory of emotions.

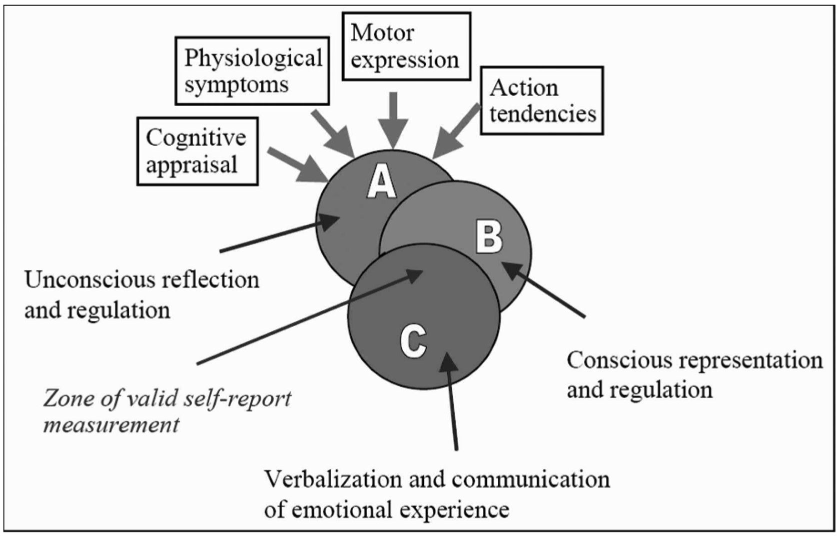

Motor expression
Physiological symptoms
Action tendencies
Cognitive appraisal
A
B
C
Unconscious reflection and regulation
Zone of valid self-report measurement
Conscious representation and regulation
Verbalization and communication of emotional experience

Figure 2.8: Scherer's three modes of representation of changes in emotion components: unconsciousness, consciousness, and verbalization (after Scherer 2005, p. 322)

availability of appropriate verbal categories [..], and (2) on the individual's intentions to control or hide some of his or her innermost feelings" (Scherer 2005, p. 322) and, accordingly, it overlaps only in part with circle (B) containing the conscious representation.

Interestingly, the "subjective feeling component" of Table 2.5 seems not to appear in Figure 2.8. One might argue that this component is realized in the overall function of "monitoring" such that the sum of all "aspects of monitoring" explicated in Figure 2.8 constitute the subjective feeling state itself. However, a closer examination of the relations between dimensional theories and Scherer's appraisal theory helps to clarify this uncertainty.

Relation to dimensional theories of emotions Recently, Scherer et al. (2006) empirically investigated possible connections between dimensional emotion theories and appraisal criteria. Using the International Affective Picture System (IAPS; Lang, Bradley & Cuthbert 1999) Scherer and colleagues conducted an empirical study to examine the question of "What factors determine the position of a feeling in affective space?" (Scherer et al. 2006, p. 93). With the term "affective space" Scherer et al. refer to the tetrahedral model by Gehm & Scherer (1988) presented in Figure 2.7 on page 31. To explain their understanding of the term "feeling" they refer to the work of Wundt (1863) and contrast it with the term "emotion" in the following way:

> "In [the component process] model, feeling is seen as a component of the emotion process, serving a monitoring function and constituting the basis for emotion regulation. Concretely, Scherer [..] has proposed that feelings integrate the central representation of appraisal-driven response organisation in emotion in the form of highly differentiated qualia, unique forms of subjective experience that reflect the configuration of component changes during the emotion episode for the individual. He has suggested that these qualia form the primitive organisation of feeling, which can then be mapped into language-specific semantic fields or into a dimensional affective space of the kind suggested by Wundt [..]." (Scherer et al. 2006, p. 93)

In the discussion of “unconscious processes in emotion” the integration across components over time is labeled “qualia” (Scherer 2005, p. 327) and directly compared to what dimension theorists such as Russel label “subjective experience” of emotion or “core affect”. This integration is assumed to be just another label for the process of component synchronization, which is believed to happen outside of awareness (see circle (A) in Figure 2.8). However, when a monitor system detects a qualitative change in the degree of coupling and synchronization that “surpasses the normal baseline fluctuations” (Scherer 2005, p. 327) the resulting feeling might enter consciousness (see circle (B) in Figure 2.8).

Definition of emotion In the attempt to define the term emotion, Scherer (2005) suggests a differentiation of seven types of affective states together with examples (printed in *italics*):

- **Preferences** as the evaluative judgements of stimuli in the sense of liking or disliking; *like, dislike, positive, negative*

- **Utilitarian emotions** as relatively brief episodes of synchronized response of all or most organismic subsystems (cp. third column of Table 2.5) to the evaluation of an external or internal event as being of major significance for personal goals and needs; *angry, sad, joyful, fearful, ashamed, proud, elated, desperate*

- **Aesthetic emotions** resulting from evaluations of auditory or visual stimuli in terms of intrinsic qualities of form or relationship of elements; *moved, awed, surprised, full of wonder, admiration, etc.*

- **Mood** as diffuse affect state, most pronounced as change in subjective feeling, of low intensity but relatively long duration, often without apparent cause; *cheerful, gloomy, irritable, listless, depressed, buoyant*

- **Attitudes** in terms of relatively enduring, affectively colored beliefs and predispositions toward objects or persons; *loving, hating, valuing, desiring*

- **Personal traits** as emotionally laden, stable personality dispositions and behavior tendencies, typical for a person; *nervous, anxious, reckless, morose, hostile, envious, jealous*

This classification of affective states is very helpful, although it naturally cannot provide an all-embracing, precise definition of the term emotion. With respect to the computational simulation of emotions proposed later in this thesis the distinctions between “preferences”, “utilitarian emotions”, and “mood” are of special interest. What Scherer refers to as “preferences”—namely valenced reactions of liking or disliking—leads to the notion of “emotional impulses” in this thesis. Understanding “utilitarian emotions” as brief episodes instead of static states lets one remember Wundt’s original idea of a “certain, continuous course of feeling” in three-dimensional affect space discussed in Section 2.1.2, p. 23. A mood, in contrast, is introduced here as a more diffuse, i.e. less object-centered, affective state with a longer duration. The question of whether to accept such strong and longer lasting affective states as “love” and “hate” as emotions reappears again, but for the computational realization proposed here, the above discrimination is followed and these difficult and very complex affective states are not included in the simulation (for a further discussion on love see (Sloman 2000)).

Directly compared to the work of Russell & Mehrabian (1977), discussed in Section 2.1.2 (p. 23), an interesting difference with respect to the emotion term “anxious” is evident. Scherer (2005) uses this label as an example of a personality trait, whereas for Russell & Mehrabian (1974) this term at first clearly denotes an emotion and later (cf. (Russell & Mehrabian 1977) and Figure 2.6, p. 30) causes trouble with respect to its dominance value. Scherer’s interpretation of “anxiety” as a personality trait solves this problem of ascribing a particular dominance value, because Russell & Mehrabian (1977) used verbal descriptions of situations in their study assuming that interpersonal differences in personality of their subjects could be ignored. Consequently, little agreement was found with respect to such an intra- and interpersonal judgement of dominance in the case of a personality trait as described by the term anxiety.

In order to differentiate among emotions, Scherer (2001) describes his idea of a “sequential check theory” that is based on a set of so-called “stimulus evaluation checks” (SECs). These checks are considered to capture the minimal set of criteria “necessary to account for the differentiation of the major families of emotional states.” (Scherer 2001, p. 94)

Stimulus evaluation checks (SECs) Scherer (2001) distinguishes four “appraisal objectives” to which each one of the 13 stimulus evaluation checks is ascribed (cf. Table 2.6). Every appraisal objective can be characterized by a typical question presented in Table 2.6 in the top rows of each of the four appraisal objectives.

Appraisal objective	**Major question & Stimulus evaluation checks**
1. Relevance Detection	How relevant is the event for me? Does it directly affect me or my social reference group?
→ 3 SECs	Novelty check; Intrinsic pleasantness check; Goal relevance check
2. Implication Assessment	What are the implications or consequences of this even and how do these affect my well-being and my immediate or long-term goals?
→ 5 SECs	Causal attribution check; Outcome probability check; Discrepancy from expectation check; Goal/need conduciveness check; Urgency check
3. Coping potential determination	How well can I cope with or adjust to these consequences?
→ 3 SECs	Control check; Power check; Adjustment check
4. Normative Significance Evaluation	What is the significance of this event with respect to my self-concept and to social norms and values?
→ 2 SECs	Internal standards check; External standards check

Table 2.6: Appraisal objectives and stimulus evaluation checks (after Scherer 2001, p. 94ff.)

1. Relevance detection is the first step in the postulated sequence of appraisal and three SECs are believed to realize this process (Scherer 2001, p. 95). Sudden stimuli (i.e. with abrupt onset and high intensity) are registered by primitive level processes to be *novel* and to deserve attention and an evaluation of familiarity, probability and predictability is believed to follow on a higher level.

This evaluation of *intrinsic pleasantness* is proposed to be part of relevance detection. Notably, Scherer (2001) believes intrinsic pleasantness to be “orthogonal to goal conduciveness”, because e.g. a piece of chocolate cake, in spite of being intrinsically pleasant for a given individual in general, can be evaluated negatively, after the individual was already forced to eat two or more pieces before. That another piece of chocolate cake is evaluated negatively can be explained by taking the internal state of the individual into account, such as his *goals and needs*. Although the goal/need conduciveness is to be evaluated later in the context of implication assessment, Scherer (2001) proposes a *goal relevance check* to take place in advance. He assumes that a stimulus is to be judged more relevant to an individual, if it has the potential of inflicting damage on one or more goals he or she currently pursues.

2. Implication assessment is suggested to be the next appraisal objective in the sequence of SECs. Scherer (2001) first clarifies the use of the term *goal/need* to capture the ill defined meanings of “motivational constructs” such as *drives, needs, instincts, motives, goals* and *concerns*. For him implication assessment forms the central objective of the appraisal process, because it directly deals with the questions of how supportive or destructive a given stimulus and its possible consequences are for an individual’s well-being. The first check with respect to this appraisal objective is concerned with the *cause* of the stimulus or event. In Scherer’s opinion, this check includes the assessment of another agent’s motive or intention, if such other agent can be made responsible and the event did not happen by chance or was caused by nature. He gives the example of a student, who received a failing grade and finds the cause for failing the exam either in a transcription error or in the professor’s intent to punish him for not attending to the course.

The next check, labeled *outcome probability*, is future-related in that it deals with the probability of an event to lead to a desired or undesired outcome. The parent’s reactions, for example, to their son’s failing (see the example above) can only be anticipated with a certain likelihood. A previous expectation, however, also influences the evaluation of an event and this aspect is covered by the *discrepancy from expectation check*. If the above student, for example, expects his parents to get angry about him failing the exam, but then comes to believe that the parents are rather happy about him, would lead to a high degree of expectation discrepancy.

Whether the event is *conducive or obstructive to an individual’s goals or needs* is to be checked next. Assuming that our student wanted to (i.e. had the goal to) be successful in the exam, failing to pass it can be interpreted as obstructive to that goal. The level of obstruction depends, however, on the relative importance of the exam. At last the *urgency of the response* to an event is to be evaluated for this appraisal objective. Urgency is assumed to increase together with the priority of the goals/needs that are threatened by the event.

3. Coping Potential Determination is an important appraisal objective for an individual, because the better one can cope with a stimulus event the better one can accept the inevitable consequences and, in effect, the “concern with the eliciting event disappears” (Scherer 2001, p. 97). The first check of *control* is not to be confused with that of *power*, although these to labels both denote the third dimension of the dimensional theories discussed in Section 2.1.2 (p. 31ff.). The term *Control* is used for the controllability of an event in general. For example, the event of someone shooting with a water pistol is considered more controllable in principle than that of getting wet by the rain, even if both events might have the same effect.

The *power* check, however, is concerned with an individual’s power to control an event, which has previously found to be controllable. Interestingly, Scherer (2001) mentions the possibility to distinguish anger and fear based on the estimation of relative levels of power between rivals. For example, if the other person shooting the water pistol is believed to be stronger than the appraising individual, the lack of power makes fear more likely as an appraisal outcome than anger.

The ability of an individual to adjust to, adapt to or live with an outcome of an event is evaluated by the *adjustment check*. If the student, for example, already thought himself to be better in another field of research as that in which he just failed to pass the exam, he or she might not get too emotional about that event after all.

4. Normative Significance Evaluation can be seen as the high-class of appraisal objectives, because it presupposes the existence of norms and values, which naturally develop in humans only after a significant time of socialization. The *internal standards* are a matter of personal moral codes and related to some sort of self-ideal in Scherer's opinion. The *external standards* are formed by the individual in relating him- or herself to social groups, which "implies shared values and rules (norms) concerning status hierarchies, prerogatives, desirable outcomes, and acceptable and unacceptable behaviors." (Scherer 2001, p. 98)

Summary Although the above SECs are proposed to be executed in a sequence, Scherer clarifies that the evaluation of an emotion in general results in "a continuous and constantly changing process." (Scherer 2005, p. 318) In consequence, the first appraisal objective "relevance detection" functions as a kind of "selective filter" responsible of filtering stimuli or events that do not exceed a certain threshold on *novelty* or *intrinsic pleasantness/unpleasantness* or *goal/need relevance* (Scherer 2001, p. 99). Goals and needs, however, can be expected to change quite frequently for an individual and past experiences also build up over time. As extrapolations into the future, that are based on these changing experiences, are used to form expectations, which in turn are again another building block in the appraisal mechanism, the long-term development of emotions is highly dynamic.

Notably, Scherer sees a direct correlation between the three dimensions of valence, activation and power/control on the one hand and appraisal criteria on the other hand in that "(1) the valence dimension reflects appraisal on intrinsic pleasantness and goal conduciveness, (2) activation reflects pertinence and urgency, and (3) power/control reflects coping potential [..]." (Scherer 2005, p. 329) The evaluation of normative significance is not related to dimensional theories, presumably because social norms and values lie outside the scope of dimensional theories. Furthermore, the proposed SECs working on this appraisal objective can be interpreted as a process of post-hoc reappraisal, which is neglected in dimensional theories (at least in the ones that are in tradition of Wundt's original proposal, which Scherer refers to in his explanations), because dimensional theorists are mostly concerned with core affect or subjective feeling state. Accordingly, dimensional emotion theorists get into serious trouble, when being asked to represent complex social emotions such as "shame" or "pride" within their two or three dimensions of affect space exclusively. As shown later these social emotions are labeled "tertiary" or "Machiavellian" emotions by some theorists (e.g. Griffiths 2002; Sloman 2000).

Before further explanations of different classes or groups of emotions are given in Section 2.2, it is helpful to take a look at the OCC model of emotions, because many computational models of emotions are based on this theory (cf. Chapter 3). It is best assigned to the fourth strand of theoretical approaches, which were distinguished in the beginning of this section (p. 33), to the semantics-based emotion theories.

The cognitive structure of emotions

Ortony, Clore & Collins started to collaborate on the topic emotion in 1980 from the perspective of cognitive psychology. They aim "to characterize the range of 'psycho-logical' possibilities for emotions rather than to describe the emotions and emotion-related processes local to any specific time or cultural group." (Ortony et al. 1988, p. x) In contrast to the broad view of Scherer's "Component Process Model" (cf. Section 2.1.3, p. 34) Ortony et al. (1988) explicitly limit their investigation on "the contribution that cognition makes to emotion." (Ortony et al. 1988, p. 1) To this respect they knowingly leave aside all but the first component of Table 2.5 on page 34 and one might be tempted to think that they merely deal with the semantics of emotion words as represented by the topmost circle (C) in Figure 2.8, p. 35. They explicitly state, however, that their theory is "decidedly *not* a theory about emotion *words*." (Ortony et al. 2005, p. 1) The terms they use are intentionally as independent of emotion words as possible assuming that (1) the structure of any lexicon of emotion words does not reflect the structure of emotions themselves and (2) "a theory about emotions has to be a theory about the kinds of things to which emotion words refer, not about the words themselves." Despite this explicit statement Scherer (1999) assigns their approach to the major strand of semantics-based appraisal theories (p. 33f.).

With regard to computational modeling of emotions Ortony et al. (1988) assume that their theory could in principle enable Artificial Intelligence systems to *reason about* emotions. The difference between *reasoning about* and *having* emotions is often neglected by computer scientists that base their implementation of emotions on the OCC model.

Interestingly, Ortony et al. (1988) also refer to James (1884), who introduced the term "standard emotions" as a label for a class of emotions that come along with "a wave of bodily disturbance". As discussed in Section 2.1.1 on page 16, James (1884) acknowledges the existence of emotions that include more cognitive elaboration and Ortony et al. (1988) take this as evidence for the necessity of their cognitive approach. Their theory, however, is not limited to model some subclass of emotions but aims to capture the whole phenomenon of emotion.

In discussing the usefulness of "linguistic evidence" for investigating of the structure of emotions Ortony et al. split the "infinitude of phenomenally possible emotions" into "manageable proportions", which they also label "representative groups or clusters." (Ortony et al. 1988, p. 15) The resulting six groups can be found as boxes in Figure 2.9, p. 41. Emotions in the OCC-theory are valenced reactions to one of three types of stimuli (cf. Figure 2.9, top of tree), which are discussed in following.

Consequences of events If the stimulus is an *event* one can in general be pleased or displeased about it. The elicitation of further emotions is conditioned by the possible consequences that the event might have for oneself or for another agent.

In the case of *consequences for other*, the resulting emotion depends on the stance that is taken toward the other agent. If the consequence of the event is *desirable for other* than one might either be *happy-for* the other agent or feel *resentment* for him or her, i.e. be *jealous* of the other agent. A different twin of emotions is the outcome of consequences of an event that is *undesirable for other*. *Gloating* is the representative of types of emotions that is elicited, if one does not like the other agent and is happy about the fact that something bad happens to him or her. One *pities* the other agent, in contrast, if one feels sorry for his or her misfortune. These four emotions are representatives of the FORTUNES-OF-OTHERS cluster of emotions.

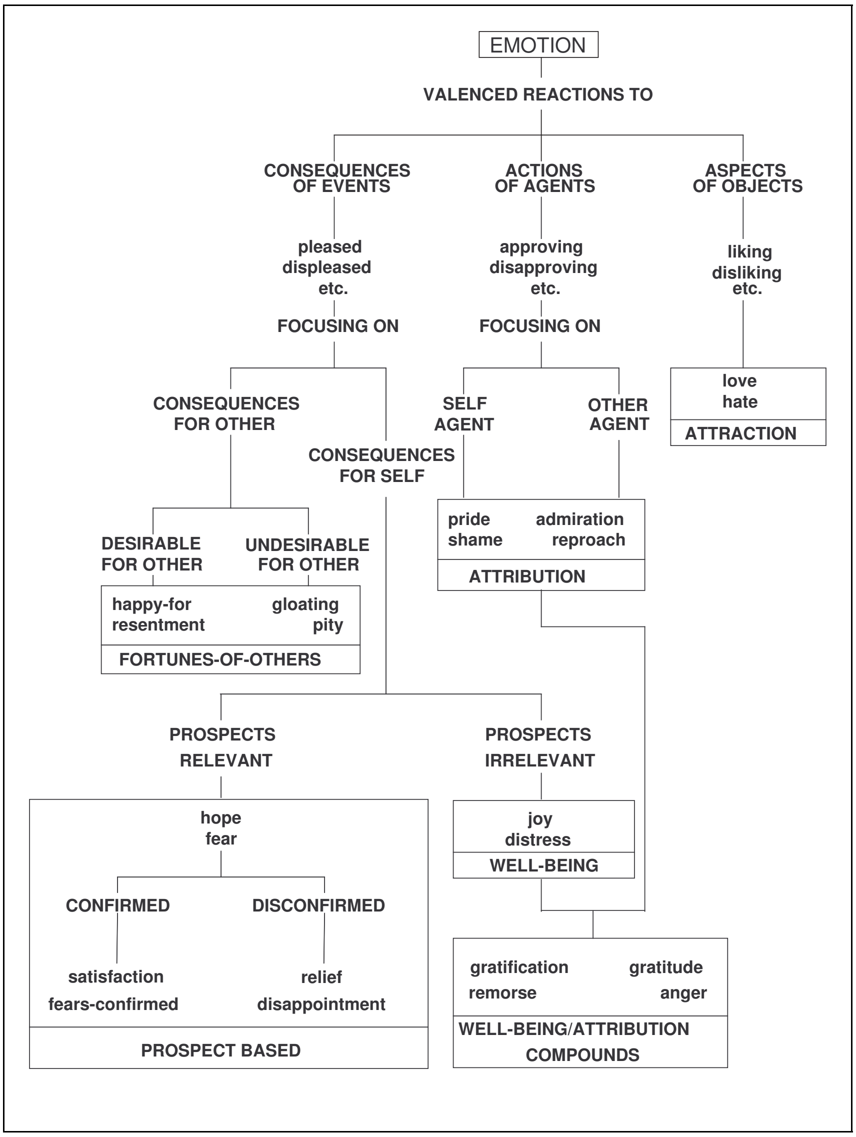


Figure 2.9: The OCC-model of emotions (after (Ortony et al. 1988, p. 19))

If the event has *consequences for self* it has to be decided, if *prospects* are *relevant* or *irrelevant*. In the case of relevant prospects *hope* or *fear* might be elicited first, e.g. one might hope that the opponent in a card game is playing a certain card that helps oneself to proceed in that game or one might fear that the opponent plays a card the hinders one’s own progress. Furthermore, it must be evaluated, if the prospect of a desirable or undesirable event is finally *confirmed* or *disconfirmed*. If a desirable event is *confirmed*, *satisfaction* might be elicited, but if an undesirable event is finally *confirmed*, one’s *fears* are *confirmed* resulting in the metaphorical emotion labeled *fears-confirmed*. In the case of the *disconfirmation* of an undesirable event, however, one might be *relieved*, but if the *disconfirmed* event was desirable instead, *disappointment* is elicited. These six emotions build the group of PROSPECT-BASED emotions. If prospects are irrelevant either *joy* or *distress* might be elicited with respect to the consequences of events. These two prototypical emotions form the WELL-BEING cluster.

Actions of agents If another agent is to be made responsible for an event, one can in general *approve* or *disapprove* the action. The further distinction between the members of the ATTRIBUTION cluster once again depends on the focus. If the focus lies on the *self*, one might feel *proud* or *ashamed*, but if the focus lies on the *other agent*, it is cognitively reasonable to experience *admiration* or *reproach*. The combination of this cluster with the WELL-BEING cluster forms the WELL-BEING/ATTRIBUTION COMPOUND cluster in which the four prototypical emotions *gratification*, *gratitude*, *remorse* and *anger* reside. *Gratification* is assumed to combine the appraisal variables of *pride* with those of *joy* and *remorse* is felt when *shame* and *distress* come together. In the case of another agent's actions the combination of *admiration* and *joy* is believed to result in *gratitude*, whereas combining *reproach* with *distress* leads to the emotion compound labeled *anger*. Ortony et al., however, explicitly point out that the compositionality of compound emotions does not imply any temporal relation of their constituents, nor is a compound emotion to be understood as the simple co-occurrence of its underlying emotions. In their view, *anger* is elicited, if one focuses at the same time on the eliciting conditions of *reproach* as well as those of *distress*. Furthermore, they argue that any compound emotion "is likely to be more intense than their constituent emotions", which might or might not be felt at the same time. (Ortony et al. 1988, p. 147)

Aspects of objects Although the evaluation of this stimulus type results only in two prototypical emotions, Ortony et al. emphasize the inherent complexity that is involved in judging the *attractiveness* of objects or aspects of objects. In such a judgement the "appealingness" of an object is important, which itself depends on one's attitudes including tastes. In their opinion, a general stance toward an object can be taken that is best described by *liking* or *disliking* the object. In their discussion of a value system that might possibly underly these attribution emotions Ortony et al. mention tastes to be especially difficult to explain. It is difficult to analyze the reasons for liking an object, if this liking is purely based on one's personal taste. If someone is asked, for example, why he or she "liked the music of Rossini" the answer might be that "its vibrant, excited, and optimistic quality" (Ortony et al. 1988, p. 158) was found most appealing. One is then forced to asked why those qualities themselves were positively evaluated but the answer to this question would not "reveal much more." To this respect the evaluation of an object in the OCC-model is quite similar to the concept of "intrinsic pleasantness" in the context of Scherer's "Stimulus Evaluation Checks" (cp. Table 2.6, p. 37).

During their discussion of context-effects Ortony et al. also mention the influence of "affective state or mood" (Ortony et al. 1988, p. 162) on the evaluative results of "momentary liking". They cite a number of studies providing evidence for an effect of mood on liking or disliking and they explain this effect as a consequence of the tendency to causally relate a given affective state with any stimulus that happens to co-occur with it. Despite mentioning this fundamental effect they do not integrate any influence of mood into their model but give the following illusive argument:

> "While [..] the cited research show[s] how irrelevant affect can bias liking, these effects presumably occur only because affective reactions ordinarily provide accurate and useful feedback from one's appraisal processes [..] [T]he feelings encountered when focusing on a particular stimulus are usually genuine reactions to that stimulus, and they [..] provide important information for subsequent judgement and decision making." (Ortony et al. 1988, p. 163)

The object of emotions in the ATTRACTION cluster can also be non-physical as Ortony et al. (1988) point out. With respect to the prototype emotion *love* they note the complexity of its meaning and emphasize that most often one loves not an object but another human being or at least an animate being. Accordingly, they remind the reader that the type specifications (e.g. linking or disliking an object) are not intended as definitions of emotion words (such as *love* or *hate* in this case).

Intensity variables In the OCC-model the intensity of any of the 22 emotion types depends on a number of "intensity variables" (Ortony et al. 1988, p. 59ff). The authors distinguish three classes of intensity variables that are summarized in Table 2.7.

Variable class	**Description**
1. Global variables	Influencing the intensity of all three classes of emotions: sense-of-reality; proximity; unexpectedness; arousal
2. Central variables	Each one is uniquely associated with a class of emotions: desirability (Event-based emotions); praiseworthiness (Attribution emotions); appealingness (Attraction emotions)
3. Local variables	Having only local effects on some emotions but not others: likelihood, effort, realization (Prospect-based emotions); desirability-for-other, liking, deservingness (Fortune-of-others emotions); strength-of-cognitive-unit, expectation-deviation (Attribution emotions); familiarity (Attraction emotions)

Table 2.7: Three classes of intensity variables of the OCC-model (Ortony et al. 1988, p. 59ff)

Interestingly, one of the "global variables" listed in Table 2.7 is labeled "arousal". Some dimensional theorists use the same term to label one of their three dimensions of emotion space (cp. Figure 2.6, p. 30). In line with the dimensional theorists Ortony et al. refer to a central aspect of one's physiology with the term "arousal". Comparable to their discussion of mood effects on appraisal, they again mention long term effects of this slow response bodily feedback loop that might influence emotional feeling. In their opinion, physiological arousal can also have non-emotional causes and it has a relatively slow rate of decay. Consequently, they assume that "it can carry forward in time from its cause and be mistakenly experienced as part of one's reaction to a subsequent event." (Ortony et al. 1988, p. 66)

Later, they explain an interesting sequence of emotions, which might be explained by such effects as preexisting arousal. The prototypical emotion "anger" is explained as the "(disapprovement of) someone else's blameworthy action and (being displeased about) the related desirable event" (Ortony et al. 1988, p. 148), whereas "frustration" belongs to the class of "Disappointment" emotions that is defined by being "(displeased about) the disconfirmation of the prospect of a desirable event." (Ortony et al. 1988, p. 122) The previously mentioned sequence consists of first getting frustrated and then becoming angry about the same situation or event. As explained above, in general the intensity of compound emotions (such as angry) is assumed to be higher than that of every other non-compound emotion, because the intensities of the constituting emotions are assumed to add up. According to (Ortony et al. 1988, p. 67), once the initial frustration fades away the blameworthiness of someone else's action alone might not generate enough physiological arousal to support the high level of angriness one just expressed toward the other agent. Accordingly, one is “likely to feel sheepish, embarrassed, and apologetic.” (Ortony et al. 1988, p. 68)

This example, however, is only comprehensible on the basis of a rather complex definition of “anger” as given above. The other emotion theories previously discussed use the same term to refer to more basic emotional or behavioral concepts such as “Destruction” in Table 2.1 on page 20. This difference shows once again the difficulties in comparing different emotion theories.

Summary The OCC-model of emotions has the advantage of being comprehensible and precise enough to form the basis of computational implementations of emotions. Its 22 emotion types are explicated in such a detail and the limitations of the theory are discussed so thoroughly that many computer scientist felt comfortable to base their implementations on this theory (see Chapter 3). This tendency to use the OCC-model of emotions in computational implementations might as well be due to the author’s discussion of “computational tractability” in the end of their book (Ortony et al. 1988, p. 181ff), where explicit rules for some emotion types are given in pseudo code. Furthermore, the notion of emotions as “valenced reactions” purely derivable on the basis of cognitive processes that are themselves to be captured in a handful of conditional rules is naturally very tempting for Artificial Intelligence researchers.

As further discussed in Chapter 3, attempts to implement OCC theory revealed a number of drawbacks. For the Affect Simulation architecture proposed here, a distinction of conscious and non-conscious emotions and processes is of general interest. Ortony et al. shortly discuss the possible correlation between “emotion experiences and unconscious emotions” (Ortony et al. 1988, p. 176ff.). With reference to Freud they state that “the experience is the sine qua non of emotions” and they further elaborate that the “beliefs or cognitions on which emotions are based can be unconscious [..] but the emotions themselves cannot be unconscious.” They pinpoint their argumentation by the example of someone encountering a bear in the woods and explaining: “One does not run away from a bear in the woods, one runs away because one is *afraid* of the bear in the woods.” (Ortony et al. 1988, p. 177ff.)

Interestingly, in their last paragraph they mention “latent emotions” as a possible candidate for unconscious emotions. These latent emotions result from situations in which the eliciting conditions of an emotion are indeed satisfied, but the intensity of the emotions does not suffice to exceed a necessary threshold. This kind of background emotion is labeled “emotion potential” by Ortony et al. (1988) and they believe that in subsequent appraisals the intensity variables might change in such away as to “allow the emotion to surface, so that if one does view an emotion potential as a kind of unconscious emotion, it is one that can potentially manifest itself as a normal emotional experience with a change in conditions.” (Ortony et al. 1988, p. 178) It is exactly this dynamic interplay of conscious and unconscious emotions that forms a central idea of the Affect Simulation Architecture proposed in this thesis.

Conclusion

Central to appraisal theories is their focus on mental processes that are based on cognitive evaluation of stimuli. Compared to the initially presented conceptions of James (1884) and Lange (1885) (cf. Section 2.1.1) it is evident that appraisal theorists are more likely to believe in the “common sense” route of emotion elicitation. In their view, cognitive processing of stimulus information, first, gives rise to an emotion independent of any bodily changes that might or might not occur later on, as depicted in the top of Figure 2.1 on page 17. It has to be noted, however, that these bodily aspects of felt emotion together with different levels of consciousness are not neglected by proponents of appraisal theories. They are not central to their theories, however, and often understood as one of many other factors influencing the otherwise rationally describable process of emotion elicitation.

Similar to Scherer's considerations of "unconscious processes in emotions" and "qualia" and to the three levels of processing proposed by Leventhal & Scherer (1987), also Ortony, Norman & Revelle (2005) recently discuss different levels of processing in "effective functioning" in more detail and introduce a distinction between "emotions" and "feelings". They understand feelings as "readouts of the brain's registration of bodily conditions and changes" whereas "emotions are interpreted feelings." (Ortony et al. 2005, p. 174) Their further considerations of three different levels of information processing (cf. Figure 2.10) are compatible with Scherer's three modes of representation given by the three circles in Figure 2.8 (p. 35).

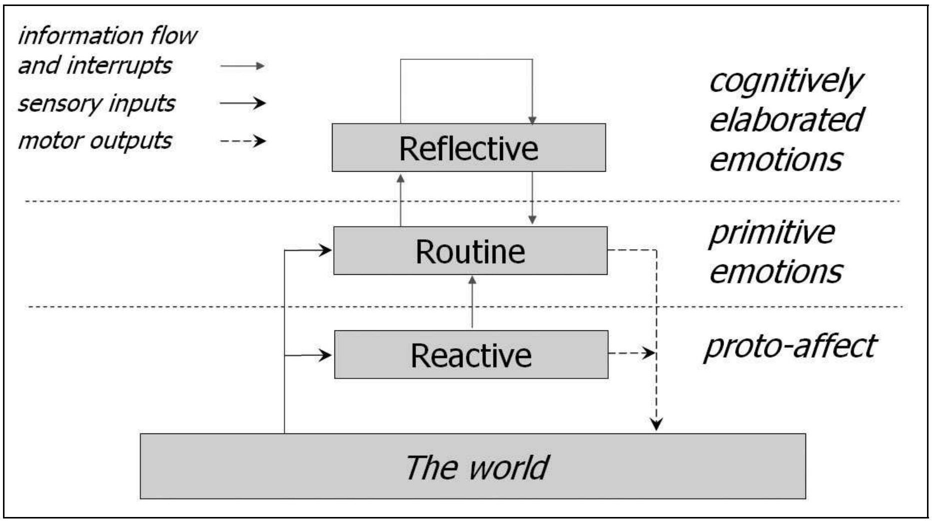


Figure 2.10: The three processing levels together with their principle interconnections (Ortony et al. 2005)

The first level is labeled "reactive" and considered to be the locus of "hard-wired releasers of fixed action patterns" giving rise to approach and avoiding behaviors. The kind of affective states that are triggered by this level is labeled "proto-affect" in Figure 2.10. Primitive and unconscious emotions are assumed to reside on the next higher "routine level" on which "well-learned automatized activity" is supposed to work on "unconscious, uninterpreted expectations". Only on the "reflective level" is higher-order cognitive processing including meta-cognition assumed to take place leading to the emergence of so-called "cognitively elaborated emotions". In their view, it is only these high-level emotions that are consciously experienced and, thus, they are the only ones that appraisal theorists are concerned about. With respect to the interaction between "full-fledged" emotions and feelings Ortony et al. (2005) note:

> "Thus, we propose that the best examples of emotions, which we often refer to as 'full-fledged emotions,' are interpretations of lower-level feelings and occur only at the reflective level, influenced by a combination of contributions from behavioral, motivational, and cognitive domains. At the middle, routine, level,

> we propose basic feelings, ‘primitive emotions,’ which have minimal cognitive content [..]. All that is possible at the reactive level is an assignment of value to stimuli, which we call ‘proto-affect.’ This in turn can be interpreted in a wide range of ways at higher levels from a vague feeling that something is right or wrong (routine level) to a specific, cognitively elaborated, full-fledged emotion (reflective level).” (Ortony et al. 2005, p. 177)

In their conclusion Ortony et al. emphasize the important contributions of lower-levels in experiencing “hot” emotion. “Cold, rational anger” could be solely the product of the cognitive component “without the concomitant feeling components from lower levels.” (Ortony et al. 2005, p. 197) A purely primitive feeling of fear, on the contrary, lacks the necessary cognitive elaboration to become a full-blown emotion. In their opinion, “it is only a feeling (albeit unpleasant) waiting to be ‘made sense of’ by reflective-level processes.”

Implications for the thesis Some process of appraisal has to be integrated into the Affect Simulation Architecture proposed in this thesis, because the concept of an “emotional impulse” requires some kind of evaluation to determine its valence dimension. Especially with respect to simulating secondary emotions one has to be able to generate expectations and to check current events against these previous expectations. A possible way to achieve these abilities in a computational architecture is by making use of standard techniques for the design of rational agents such as explicitly modeling the beliefs, desires and intentions (BDI) of an agent.

Despite their profoundly different starting points the previously discussed appraisal theories show the following similarities:

1. Social aspects of emotions are important to both theories, although the respective roles of other agents take influence on different levels. In the Component Process Model the appraisal objective “Coping potential” consists of three SECs that evaluate an agent’s social rank. In the OCC model the second distinction taken for the cognitive structure of emotions is that of distinguishing oneself from the other agent.

2. Different levels of processing are postulated by both groups of researchers lately. Scherer (2005) distinguishes three modes of representation (cf. Figure 2.8) and Ortony et al. (2005) three processing levels (cf. Figure 2.10).

These similarities lead the way in designing a computational architecture of affect that aims to simulate “hot, felt” rather than “cold, purely cognitively” emotions. In their argumentation for three processing levels Ortony et al. (2005) refer to the findings of neurobiology that were acquired during the last 15 years by means of neuro-imaging techniques. In the following a short overview of this interesting field together with its findings relevant to emotion research is given.

2.2 Neurobiological and ontogenetical background

> “So can robots ‘have’ emotions? If you ask a patient who has been implanted with a mechanical device that pump his blood in the center of his chest if he

> has a heart, his answer will most certainly be 'Yes, I have an artificial heart!' Similarly, it will come a time when you will be able to ask your computer if it has emotions, and its answer will undoubtedly be 'Yes, I have computer-emotions!' In the meantime, how do we even begin to think about how to implement emotions? Why not use the brain as a source of inspiration?" (Fellous 2004)

With the example quoted above Fellous (2004) tries to enlighten his argument that one day robots might really "have" rather than only "show" emotions. According to Fellous (2004), the notion of "computer-emotions", that are most likely different from human emotions, is supported by neurobiological findings. Furthermore he notices the misleading oversimplification inherent in the term "emotional 'state', because emotions may be intrinsically dynamical phenomena of widely different time constants (from a few seconds for perceptual fear, to hours or days for moods, to month or years for depression or love)." This emotion dynamics is central to the conceptualization of a computational Affect Simulation Architecture in this thesis.

To explain the concept of "computer-emotions" Fellous (2004) rises the question whether one can reasonably ascribe the same kind of emotional experience to animals as to humans. To investigate theoretically possible differences between human and animal emotions he suggests looking at how their brain differs from ours. Many supportive arguments for the use of animals in studying the role of the brain in emotional processes are given by Joseph LeDoux (cf. (LeDoux 1995), (LeDoux 1996)) and an overview of his findings and conclusions is given next.

2.2.1 The Emotional Brain

> "Contrary to the primary supposition of cognitive appraisal theories, the core of an emotion is not an introspectively accessible conscious representation. Feelings do involve conscious content, but we don't necessarily have conscious access to the processes that produce the content. And even when we do have introspective access, the conscious content is not likely to be what triggered the emotional responses in the first place." (LeDoux 1996, p. 299)

LeDoux (1996) mainly concentrates on the investigation of the emotion "fear" as to him it is not reasonable to assume that one single brain region is responsible for all emotions in humans. First, LeDoux (1996) discusses the work of James (1884) ("feedback theory", cf. Section 2.1.1), Cannon (1927), and MacLean (1949) ("limbic system theory", MacLean 1970) and clarifies that MacLean's limbic system theory must have been convincingly enough to prevent neuroscientists from further investigating the connection between neuroanatomical processes and emotions for several decades.

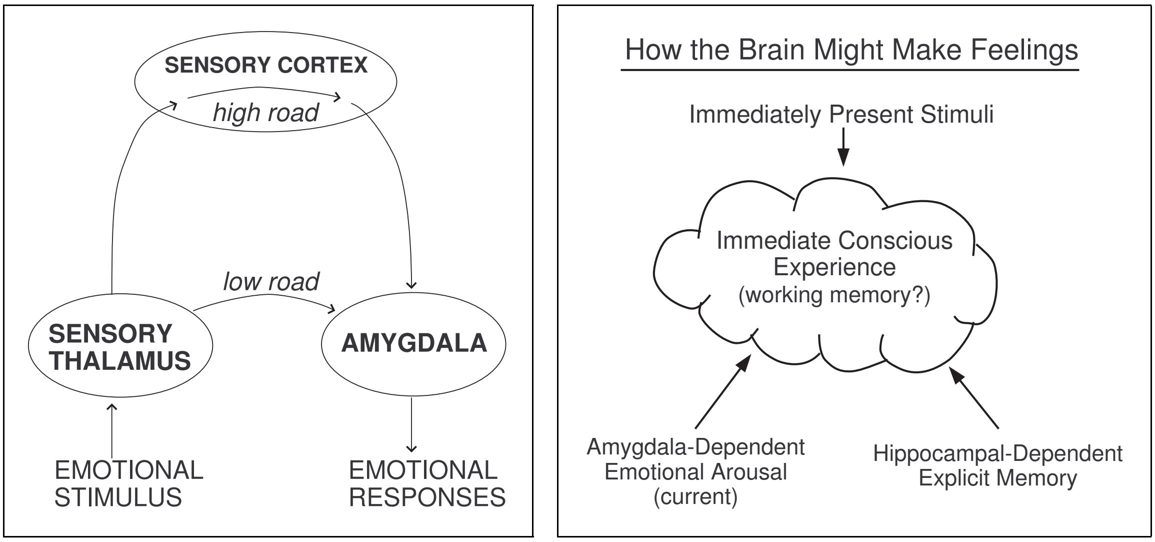
According to LeDoux (1996), however, it has never been sufficiently clarified, which parts of the brain are constituting the limbic system, and the region to which the limbic system traditionally was ascribed has been found to be active in non-emotional processes as well. Furthermore, LeDoux criticizes the idea that "the limbic system theory of the emotional brain was meant to apply equally to all emotions." (LeDoux 1996, p. 102) He believes this view to be in principle possible but also states that little evidence exists speaking in favor of it. Based on the idea to take the evolution of the brain as key to understanding emotions, LeDoux points to different survival functions of different emotions (cp. the discussion of primitiveness for the idea of “basic emotions” in Section 2.1.2). As for each of these functions different brain systems may have evolved, he argues for the possibility of more than one emotional system in the brain.

Fear conditioning and the amygdala

With his experiments on fear conditioning in animals LeDoux (1996) shows the importance of the amygdala in brain processes that result in behaviors commonly interpreted to accompany the experience of fear.

Imagine a rat being placed in a box with a loudspeaker in one corner. The base of the box is equipped with a fine net of electric cables through which a small amount of electricity can be transmitted to induce a relatively mild shock in the rat. When for the first time a sound is played the rat will orient toward the sound, but after several occurrences, the sound is ignored. Next, the sound is accompanied by a brief electric shock letting the rat orient itself toward the sound again. This way the sound by association with the shock has become a learned trigger of fear response, because the next time a sound alone is played the rat will show the same pattern of fear response as if an electric shock were present as well.

After carefully investigating the processes in the rat’s brain and comparing the results with several other neurobiological findings LeDoux (1996) distinguishes a low and a high road of fear elicitation in the brain (cf. Figure 2.11(a)). The processes responsible for emotional learning (as it occurs in the case of fear conditioning) can bypass the area of thinking, reasoning and consciousness (namely the neocortex) and directly exert influence on the amygdala.


SENSORY CORTEX
high road
low road
SENSORY THALAMUS
AMYGDALA
EMOTIONAL STIMULUS
EMOTIONAL RESPONSES
How the Brain Might Make Feelings
Immediately Present Stimuli
Immediate Conscious Experience (working memory?)
Amygdala-Dependent Emotional Arousal (current)
Hippocampal-Dependent Explicit Memory

(a) The low and the high roads to the amygdala; redrawn after (LeDoux 1996, p. 164)

(b) The generation of conscious experience of emotions; redrawn after (LeDoux 2000, p. 176)

Figure 2.11: LeDoux’s conception of the emotional brain and his possible explanation for conscious experience of emotions

The amygdala, in turn, influences the sensory areas of the cortex to an even greater extent than these areas influence the amygdala (LeDoux 1996, p. 268). Accordingly, the fast, low-level responses to a stimulus generated by the amygdala are believed to change perception and, furthermore, the cognitive processing of the emotional brain (LeDoux 1996, p. 284ff.).

From Conscious Appraisal to Emotions

Concerning subjective feelings LeDoux (1996) highlights the possible contribution of working memory in generating conscious experience. As presented in Figure 2.11(b) “immediately present stimuli” need to be accompanied by “amygdala-dependent emotional arousal” and “hippocampal-dependent explicit memory” to generate “immediate conscious experience”. The contribution of the amygdala is only considered relevant in case of fearful experiences and LeDoux (1996) states clearly that the output of other systems might be important as well. This mechanism of concurrent representation of symbolic derivatives from different subsystems in working memory is assumed to also underly other conscious feelings.

To explain the difference between purely cognitive appraisals and “full-blown emotional experience” (LeDoux 1996, p. 283) LeDoux presents the example of only generating “conscious representations” of the perception of a rabbit and a snake while walking through a forest. Imagine yourself walking through a forest and suddenly you see a rabbit. The visual perception is transformed into a representation activating relevant long-term memories that are integrated with the content of working memory allowing you to be consciously aware that the object, you are looking at, is a rabbit. A few moments later you encounter a snake. A similar process as before results in a conscious representation of the snake in working memory; this time, however, the contents of long-term memory also inform you that a snake is a potentially dangerous animal. These processes, so far, can be sufficiently explained by appraisal theories that were presented in Section 2.1.3. The emotion “fear” will most likely be the outcome of these appraisal processes.

According to LeDoux (1996), there is something else needed “to turn cognitive appraisals into emotions, to turn experiences into emotional experiences.” (LeDoux 1996, p. 284) A cognitive representation of “fear” is only turned into an emotional feeling, if it is accompanied by an activation of the amygdala, as LeDoux’s empirical findings suggest. The output of the amygdala is described in terms of three “basic ingredients” that together with selected content of long-term memory and short-term sensor representations create the conscious experience of subjective feeling in working memory.

Ingredient 1: Direct amygdala influences on the cortex Some areas of the cortex are responsible for the processing of all kinds of stimuli. As LeDoux (1996) points out, the amygdala has more connections back to the cortex than it gets inputs from the cortex. He highlights the significant influence that amygdala activation might have on the areas in the cortex processing visual stimuli. Thereby the amygdala might be responsible for directing attention to emotionally relevant stimuli. The amygdala is also believed to influence long-term memory networks such that the recall of relevant emotional implications of the present stimuli is facilitated. Furthermore, by way of connections to the orbital cortex the amygdala plays a role in rewards and punishments.

All these influences, however, only provide working memory with information about the goodness or badness of a given stimulus, but they cannot account for the emergence of a subjective feeling in LeDoux’s opinion.

Ingredient 2: Amygdala-triggered arousal The amygdala also exerts indirect influence on the cortex by means of different channels. According to LeDoux (1996), an "extremely important set of such connections involves the arousal systems of the brain." (LeDoux 1996, p. 285) Low arousal (as in case of drowsiness or sleep) is signified by a slow and rhythmic electroencephalogram (EEG), whereas high arousal (when being alert or paying attention) results in a fast and desynchronized EEG. LeDoux (1996) refers to dimensional theories (cp. dimensional theories, Section 2.1.2, and "arousal" as one of the global intensity variables of the OCC-model, cf. Table 2.7, p. 43) of emotions in explaining the possible connection of high levels of arousal with the inability to concentrate on other things than the emotion eliciting stimulus. Notably, according to LeDoux (1996), not only the amygdala activates arousal systems in the brain, but "the way they are turned on by a dangerous stimulus is through the activity of the amygdala." (LeDoux 1996, p. 290) In general, arousal is triggered by novel stimuli, but only if these stimuli are emotionally relevant, such an activation lasts for a longer time. In the case of emotional stimuli the initially triggered arousal is amygdala-independent, but the concurrent contribution of amygdala induced arousal is assumed to "add impetus to keep the arousal going." (LeDoux 1996, p. 290) An inherent circularity of the amygdala and the arousal systems is proposed to result in "self-perpetuating, vicious cycles of emotional reactivity."

Together with the above consideration of amgdala's influence on the cortex a nearly complete picture is obtained comparable to that of a two-dimensional emotion space. The valence detection is achieved by the first ingredient and the necessary arousal is triggered by the second. In LeDoux's opinion, however, one more ingredient is necessary—bodily feedback, as it was introduced in the beginning of this Chapter.

Ingredient 3: Bodily feedback With reference to Cannon's work (cp. Section 2.1.2) LeDoux (1996) claims that the autonomic nervous system (ANS), which controls the viscera, "has the ability to respond selectively, so that visceral organs can be activated in different ways in different situations." (LeDoux 1996, p. 292) Different emotions (anger, fear, disgust, sadness, happiness, surprise) are "to some extend" distinguishable "on the basis of different autonomic nervous system responses (like skin temperature and heart rate)." Concerning the relatively slow action of visceral responses, which he acknowledges, LeDoux (1996) points to the inherent dynamics of emotional states. Fear, for example, "can turn into anger or disgust or relief as an emotional episode unfolds" and LeDoux further suspects "that visceral feedback contributes to these emotional changes over time." (LeDoux 1996, p. 293)

LeDoux (1996) refers to the work of Antonio Damasio (1994) in explaining the importance of somatic responses that are assumed to be fast and differentiated enough to play a more direct role in emotion elicitation. Especially, Damasio's concept of "as-if loops" for bodily feedback is very important for this thesis and is therefore explained in the context of his influential "somatic marker hypothesis" in the following.

2.2.2 The Somatic Marker Hypothesis

In Damasio's opinion, the brain and the body are inseparably connected in the process of reasoning. Furthermore, the "high-level" and "low-level" regions of the brain always "cooperate in making reason" (Damasio 1994, p. xxiii). Because these low-level regions are in charge of regulating not only “virtually every bodily organ” but also the processing of emotions, Damasio (1994) concludes that “[e]motion, feeling, and biological regulation all play a role in human reason.”

The neurobiological findings of Damasio suggest that only humans with impairments in certain brain regions show a problem solving behavior, that is best described as based on purely rational, logics-based reasoning. This “high-reason” (Damasio 1994, p. 171) is considered the “rationalists conception” of human problem solving in which emotions and passions are judged as misleading and confusing and best kept out of the process. Damasio, however, uses the terms “brain” and “mind” interchangeably (Damasio 1994, p. 155) and in combination with the above explanations the mind cannot be seen as independent from the body.

The “somatic marker hypothesis” is derived from an extensive amount of different neurobiological and psychological findings. It is Damasio’s proposal of a mechanism that is believed to underly the dynamic interaction of brain and body finally resulting in conscious feelings. Before the somatic marker hypothesis is explained, however, Damasio’s distinction of primary and secondary emotions is introduced.

Primary and secondary emotions

Damasio (1994) begins his discussion of emotions with reciting William James’ feedback theory (cf. Section 2.1.1, p. 16) emphasizing its “preorganized mechanism” (Damasio 1994, p. 131). He highlights three important criticisms of James’ theory:

1. James (1884) completely neglected the cognitive processes involved in emotion elicitation. According to Damasio (1994), “[h]is account works well for the first emotions one experiences in life, but it does not do justice to what Othello goes through in his mind before he develops jealousy and anger [..].”

2. James (1884) does not allow for any alternative mechanism of feeling than bodily feedback. Without a body[11] there would be no feeling possible in James’ view.

3. All the diverse effects of emotions on cognition and behavior are not included in James’ theory despite their importance in the process of dynamic interactions between brain and body (see also Section 2.2.1).

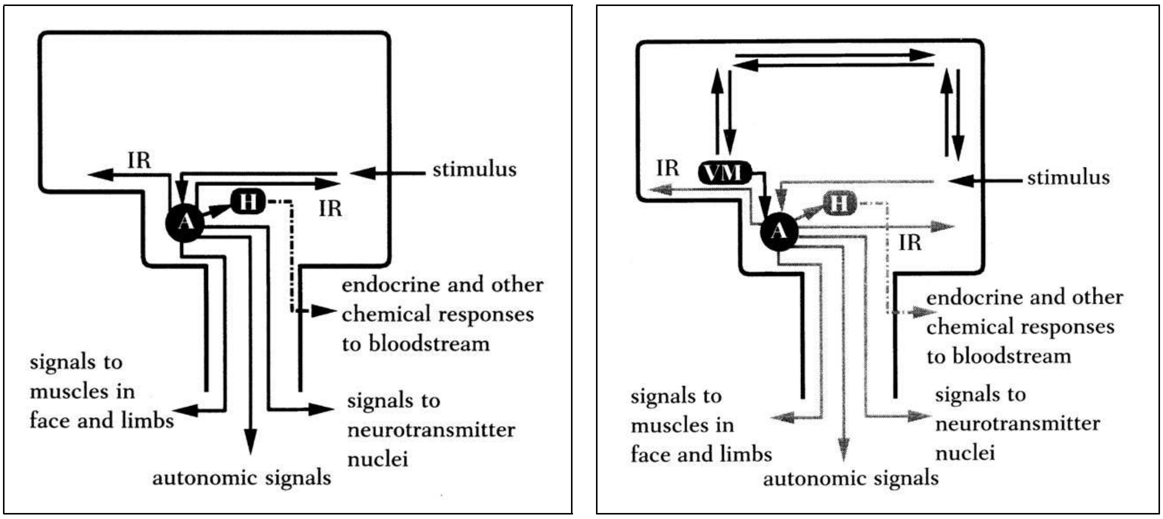
Based on these criticisms Damasio concludes as follows:

> “I begin with the perspective of personal history, and clarify the differences between the emotions we experience early in life, for which a Jamesian ‘preorganized mechanism’ would suffice, and the emotions we experience as adults, whose scaffolding has been built gradually on the foundation of those ‘early’ emotions. I propose calling ‘early’ emotions primary, and ‘adult’ emotions secondary.” (Damasio 1994, p. 131)

Interestingly, Damasio (1994) does not refer to the term “standard emotions” coined by James (1884), although his notion of primary emotions seems to be quite similar.

[11] Damasio defines the body as “the organism minus the neural tissue (the central and peripheral components of the nervous system) [..]” (Damasio 1994, p. 86) and abbreviates the term “nervous system” with the term “brain”.

Primary emotions With this term Damasio (1994) refers to a class of emotions that are supposed to be “wired in at birth”, i.e. inate. They are supposed to depend on “limbic system circuitry, the amygdala and anterior cingulate being the prime players.” (Damasio 1994, p. 133) Damasio also refers to the work of LeDoux (cf. Section 2.2.1) that supports the importance of the amygdala in emotional processes. The perceptional triggers of primary emotions are described as “certain features of stimuli in the world or in our bodies” that are “processed and then detected by a component of the brain’s limbic system, say, the amygdala” which gives rise to a bodily-state “characteristic of the emotion fear.” (Damasio 1994, p. 131) To this extent the processes described by Damasio (1994) are very similar to LeDoux’s considerations presented before. A graphical representation of these unconscious processes together with the brain regions involved is given in Figure 2.12(a).



(a) The amygdala (A) and the hippocampus (H) are supposed to be the brain regions involved in the elicitation process of primary emotions. “After an appropriate stimulus activates the amygdala (A), a number of responses ensue: internal responses (IR); muscular responses; visceral responses (autonomic signals); and responses to neurotransmitter nuclei and hypothalamus (H). The hypothalamus gives rise to endocrine and other chemical responses which use a bloodstream route. [..]” cited from (Damasio 1994, p. 132) Other brain structures are also involved in the process but deliberatively left out by Damasio.

(b) In case of secondary emotions the stimulus additionally gets “analyzed in the thought process, and may activate frontal cortices (VM)” as Damasio (1994) proposes. “VM acts via the amygdala. In other words, secondary emotions utilize the machinery of Primary Emotions. [..] Note how the VM depends on A to express its activity [..].” Damasio (1994) points out that he is once again “deliberately oversimplifying”. Cited from (Damasio 1994, p. 137)

Figure 2.12: Damasio’s conception of primary (a) and secondary (b) emotions together with the respective brain regions (Damasio 1994). The black perimeter in both pictures represents the brain and brain stem.

These primary emotions developed during phylogeny to support fast and reactive response behavior in case of immediate danger (see the discussion of “basic emotions” in Section 2.1.2, p. 20). In humans, however, the perception of the changed bodily state is combined with the object that initiated it resulting in a “feeling of the emotion” with respect to that particular object (Damasio 1994, p. 132). Being conscious of one’s own primary emotions offers us “flexibility of response based on the particular history of [our] interactions with the environment." (Damasio 1994, p. 133) In his later writing Damasio (2003) understands primary emotions as the class of prototypical, simple emotion types which can already be ascribed to one year old children.

There are even more powerful emotional mechanisms in our brain that develop in every normal, human individual during ontogenesis. Damasio (1994) explains as follows:

> "[..] I believe that in terms of an individual's development [the basic mechanisms] are followed by mechanisms of *secondary emotions*, which occur once we begin experiencing feelings and forming *systematic connections between categories of objects and situations, on the one hand, and primary emotions, on the other*. Structures in the limbic system are not sufficient to support the process of secondary emotions. The network must be broadened, and it requires the agency of prefrontal and of somatosensory cortices." (Damasio 1994, p. 134), italics in the original

Secondary emotions If bodily feedback were necessary for every instance of emotional experience then it could hardly be explained why and how "being told of the unexpected death of a person who worked close to you" (Damasio 1994, p. 134) could give rise to emotional experience. Similarly, admiring a sophisticated piece of art—be it an opera or a painting—does probably not involve any appraisal of the likelihood of a life-threatening outcome.

In explaining the rationale for secondary emotions Damasio (1994) points to the important role of one's individual experience. In the introductory example the elicitation of a secondary emotion is based on imagining a hypothetical situation—the unexpected death of a close collaborator. Damasio (1994) describes the process as follows (cf. Figure 2.12(b)):

A. Conscious, deliberate consideration: The idea of "mental images" is central to this cognitive processing step. By means of mental images a person is believed to reflect on the other person's current situation, the possible consequences for him- or herself and the other person, "in sum, a cognitive evaluation of the contents of the event [..]." (Damasio 1994, p. 136) In general, these mental images form representations that are "constructed under the guidance of dispositional representations held in distributed manner over a large number of higher-order association cortices."[12] (Damasio 1994, p. 136)

B. Non-conscious response of prefrontal cortex: The same "dispositional representations" as above are believed to hold knowledge of one's individual experience in terms of pairings of "certain types of situations" and "certain emotional responses". This knowledge is used by the prefrontal cortex to respond "automatically and involuntarily [..] to signals arising from the above images." The non-conscious learning of this kind of "acquired dispositional representations" is influenced by the earlier type of "innate dispositional respresentations". "To summarize: The prefrontal, acquired dispositional representations needed for secondary emotions are a separate lot from the innate dispositional representations needed for primary emotions." (Damasio 1994, p. 137)

C. Nonconscious response of amygdala and anterior cingulate: The response of the above prefrontal dispositional representations is signaled to the amygdala and the anterior cingulate and four kinds of responses ensue: (a) signals to the body via peripheral nerves

[12] In the caption of Figure 2.12(b) this process is described as analyzing a stimulus "in the thought process".

> resulting in changes of the state of the viscera; (b) signals to the motor system resulting in changes of body posture and facial expression; (c) activation of the endocrine and peptide systems resulting in chemical actions changing the body and brain states; and finally, (d) particular patterns activate nonspecific neurotransmitter nuclei in the brain stem and basal forebrain resulting in "chemical messages in varied regions of the telencephalon (e.g. basal ganglia and cerebral cortex)." (Damasio 1994, p. 138)

This outline of the process leading to the elicitation of secondary emotions might be judged as unsatisfactory by a computer scientist, because many questions arise such as how to make these major processing steps explicit enough for a computational implementation. Of course, Damasio does not aim to provide such a detailed and explicit description of the complex brain processes[13]. For the aim of this thesis, however, the following assumptions are derived from the above description:

1. In contrast to primary emotions, the process resulting in secondary emotions starts with conscious, cognitive evaluation. (A)

2. The deliberation process uses and modifies aspects of the past (memories, experiences) and the future (expectations). (A)

3. Some kind of higher-order, dispositional representation forms the basis of so-called "mental images" which can be pictorial or linguistical. (A)

4. The past experiences are crystallized in pairings of situations and (primary) emotions. Nonconscious processes work on these experiences to derive appropriate second-order dispositional representations that are needed for secondary emotions. (B)

5. The bodily responses (a), (b), and (c) cause an "emotional body state" (Damasio 1994, p. 138) that is subsequently analyzed in the thought process after having been signaled back "to the limbic *and* somatosensory systems." (italics in the original) (C)

6. In parallel, the cognitive state itself (i.e. the brain) is directly modulated during the process. (C)

After the reconceptualization of Damasio's description his proposal is comparable to Scherer's relationship between functions and components of emotion and the organismic subsystems (Scherer 2001) presented in Table 2.5, p. 34.

Concerning the bodily responses (see 5) one might still wonder how Damasio can account for those kinds of emotional experience that seem not to involve any bodily feedback, e.g. the introductory example of admiration. At this point Damasio (1994) introduces his "somatic marker hypothesis" together with an "as-if loop" of bodily feedback.

Somatic markers and the "as-if" loop of bodily feedback

Damasio (1994) summarizes his idea as follows:

[13] And he states the impossibility of such an endeavor because of the lack of further details.

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

The acquisition of these somatic markers is described as resulting from inherently social and developmental processes (Damasio 1994, p. 177). They are, thus, believed to be acquired "under the control of an internal preference system and under the influence of an external set of circumstances which include [..] also social conventions and ethical rules." (Damasio 1994, p. 179) This differentiation of internal and external control reminds one again of Scherer's "appraisal objective" labeled "Normative Significance Evaluation" in Table 2.6, p. 37.

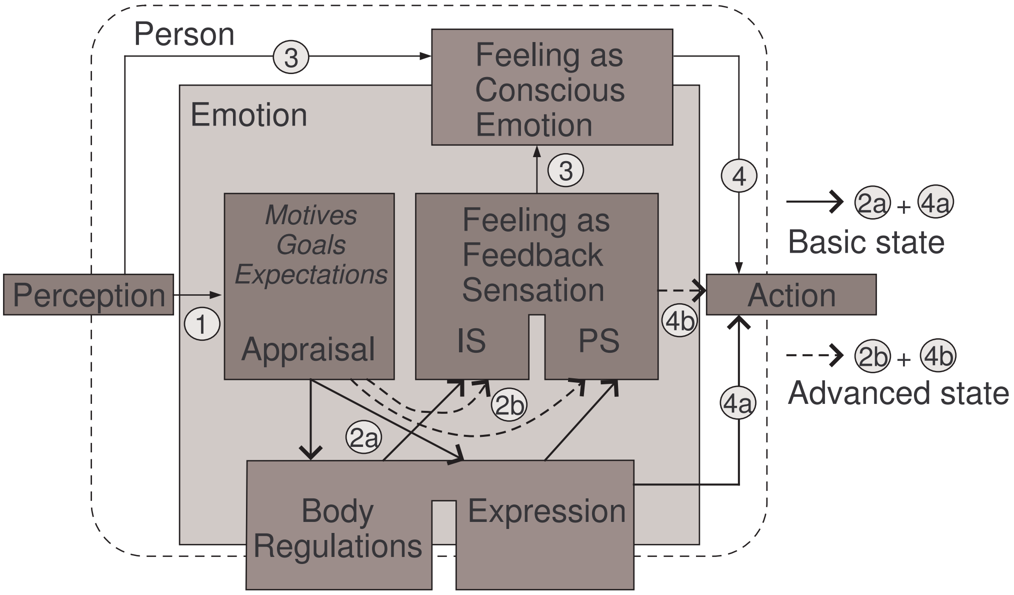


Figure 2.13: The internalization model of emotional development (after (Holodynski & Friedlmeier 2005, p. 68+70)). (1) A stimulus is perceived and appraised on the basis of current motives, goals and expectations. (2a) Basic state: Triggering body and expressive reactions. (2b) Advanced state: Body and expressive reactions can be bridged by mental representations of interoceptive (IS) and proprioceptive (PS) sensations. (3) Simultaneous representation of the cause of the emotion and the body and expressive reactions as conscious feeling. (4) Body and expressive (4a) *reactions* (basic state) or (4b) *sensations* (advanced state) plus conscious feeling trigger motive serving actions

Concerning emotional experience and expression Holodynski & Friedlmeier (2005) also believe that the variety of emotions increases during ontogenesis due to the availability of higher cognitive functions. They present an "internalization model of emotional development" in its "basic" and "advanced" state (cf. Figure 2.13).

In their discussion Holodynski & Friedlmeier (2005) also refer to Damasio's idea of "somatic markers" by which otherwise unemotionally perceived causes of events become "marked and coloured” (Holodynski & Friedlmeier 2005, p. 68) by expressive and bodily sensations. The difference between the basic and the advanced state of their model consists of an adults ability to internalize his or her bodily and expressive feedback preventing a directly observable expression of his or her emotional state. A child, on the other hand, is almost unable to bypass the body loop and to refer only to somatic sensations without corresponding expressions and body reactions.

After somatic markers have built up during ontogenesis, they are believed to reside in the somatosensory system of the brain. If the above process of secondary emotion elicitation makes use of these learned bodily experiences instead of the real, less responsive body, the “as-if” loop of bodily feedback is established. Damasio states clearly that “[t]he processing in the ‘as-if’ loop bypasses the body entirely.” (Damasio 1994, p. 156)

In the ninth chapter of his book Damasio (1994) presents first results of empirical tests of his somatic marker hypothesis (SMH). The “Iowa gambling task” (IGT, see Bechara, Damasio, Tranel & Damasio (2005) for a description) is mostly used to falsify the prediction that emotional impairments influence rational decision making as derivable from Damasio’s work. Based on the IGT, Bechara et al. (2005) provided additional support for the general reasonability of the SMH, but the interpretability of the acquired data is still a highly debated topic (cf. Maia & McClelland (2004), Dunn, Dalgleish & Lawrence (2006)).

Other classes of emotions

Before summarizing this Chapter two other classes of emotions are outlined, which have been introduced by Damasio (2003): Background and Social emotions.

Background emotions They are considered to be different from moods (e.g. as defined by Scherer (2001)) but they bear some resemblance with Scherer’s definition of preferences (p. 36). According to Damasio, when spontaneously being asked how one feels one is likely to answer in terms of a background emotion. Accordingly, background emotions “are composite expressions of [..] regulatory actions [(e.g., basic homeostatic processes, pain and pleasure behaviors, and appetites)] as they unfold and intersect moment by moment in our lives.” (Damasio 2003, p. 44) Damasio admits, however, that this concept still needs to be clarified by further investigation.

Social emotions Damasio calls the previously introduced secondary emotions now “social emotions” and presents “sympathy, embarrassment, shame, guilt, pride, jealousy, envy, gratitude, admiration, indignation, and contempt” (Damasio 2003, p. 44) as examplary members. Sloman (2000) and Griffiths (2002) introduced this class of emotions before already naming its members “tertiary emotions” (Sloman) and “machiavellian emotions” (Griffiths).

Zinck & Newen (2007) further split up social emotions into primary and secondary cognitive emotions. The first subclass refers to, e.g., such types of joy “in which a minimal set of cognitive content is present in the emotional pattern”, for example, when listening to “the clear composition of the triumphant conclusion to a Beethoven symphony.” (Zinck & Newen 2007, p. 13) Contrary, secondary cognitive emotions are labeled ”high-level cognitive emotions” (Zinck & Newen 2007, p. 14) and are based on cognitive evaluation of situations including social norms, expectations and the like.

2.2.3 Conclusion

Taking into account that the term “emotion” still refers to many different affect-related concepts it is not suprising that neurobiologists interpretations of their findings are less clear-cut as hoped for. Nevertheless, the following classes of emotions are derived from the above findings concerning neural machinery of emotions and their ontogenetical development:

1. Background emotions remain mainly unconscious and resemble our “state of being” on a scale between good or bad. Basic approach and avoidance behaviors result from these background emotions and the predisposition to experience primary emotions is changed as well.

2. Primary emotions are the class of prototypical, simple emotion types which can already be ascribed to one year old children. These emotions are accompanied by distinct facial expressions that are clearly identifiable across cultures and even across species. Examples include fear, anger, disgust, sadness and happiness.

3. Secondary or social emotions are the product of complex, cognitive processing based on social norms and values as well as experiences and expectations. A secondary emotion such as pride or embarrassment is often accompanied by a primary emotion’s facial expression.

Furthermore, the idea of an “as-if” loop for bodily feedback solves some problems with the original feedback theory proposed by James (1884) and Lange (1885) (cf. Section2.1.1).

It has to be pointed out, however, that neither LeDoux nor Damasio consider it possible for a robotic system to ever really “have” emotions.

For LeDoux (1996) the study of “how the brain processes emotional information” only helps to “understand how it creates emotional experience” but not to “program computers to have these experiences.” (LeDoux 1996, p. 37) He proposes instead to “use information processing ideas as the conceptual apparatus for understanding conscious experience.” (LeDoux 1996, p. 38) His further argumentation in the context of feelings, however, remains unsatisfying.

For Damasio (1994) the inability of a computer system to experience rather than only simulate emotions and feelings results from the cognitive scientist’s belief in the “mind as a software program” running on a separable hardware. Consequently, Damasio believes that Descartes made the following error:

> “This is Descartes’ error: the abyssal separation between body and mind, between the sizable, dimensioned, mechanically operated, infinitely divisible body stuff, on the one hand, and the unsizable, undimensioned, un-pushpullable, nondivisible mind stuff; the suggestion that reasoning, and moral judgement, and the suffering that comes from physical pain or emotional upheaval might exist separately from the body.” (Damasio 1994, p. 249f)

2.3 Summary

This chapter started with an introduction to feedback-theories in which bodily feedback was historically not only considered necessary but also sufficient for the experience of so-called “standard emotion” (James 1884). This assumption was subsequently refined several times and resulted in the so-called neo-jamesian theories. The different assumptions of the facial feedback theory were discussed and in particular Ekman’s studies on unversial expressions of emotions were detailed in its context. The resulting idea of “basic emotions” led to the investigation of one member of so-called “palette theories” of emotions, namely Plutchik’s three-dimensional structural model of emotions. After the questionable points of this theory were highlighted and its positive aspects detailed, the general class of dimensional theories was presented.

In explaining Wundt’s early idea of a “continuous course of feeling” (Wundt 1863) in three-dimensional, orthogonal emotion space the aspect of subjective feeling state became central. The “affective primacy idea” (Zajonc 1980) was elaborated, according to which cold cognitions are turned into hot emotions.

A detailed investigation of a number of other dimensional emotion theories led to the conclusion that three dimensions are necessary and sufficient to capture the main elements of an emotion’s connotative meaning—at least in case of simpler emotions such as primary or basic ones. The three dimensions chosen for emotion representation in this thesis are labeled *Pleasure*, *Arousal*, and *Dominance* (Russell & Mehrabian 1977) spanning an orthogonal space, which is labeled PAD space. Five of Ekman’s six basic emotions were located in PAD space according to Russell & Mehrabian (1977).

With a focus on the processes that form the basis of emotional episodes two appraisal theories were discussed next. At first, the classical approach was examplified with a discussion of Scherer’s Component Process Model (Scherer 1984). In this context Scherer’s considerations of the difference between conscious and unconscious processes in appraisal (Scherer 2005) were introduced and contrasted with dimensional theories, for which a definition of affective states was given. Afterwards, the thirteen stimulus evaluation checks (SEC) were detailed together with their respective appraisal objectives.

The OCC-theory of Ortony et al. (1988) was finally outlined as a second, important example of an appraisal theory. Thereby, the connection to the other emotion theories—esp. to the ideas of James (1884)—was drawn whenever possible to show that this theory is not only trying to explain the semantic field of emotion words, although it has to be grouped into the class of semantics-based emotion theories.

In concluding the appraisal theories the recent ideas of Ortony et al. (2005) were introduced and compared to the ideas of Scherer (2005). In summary, also appraisal theorists recently consider some kind of bodily feedback resulting from lower-level, presumably unconscious processing important for realizing “hot” emotions out of “cold” cognitions.

In Section 2.2 the neural machinery of the brain was examined. LeDoux’s work on fear conditioning provided the idea that the co-representation in working memory of immediately present stimuli, amygdala dependent arousal, and hippocampal-dependent explicit memory—if it is accompanied by bodily feedback—might explain conscious experience of fear.

In the context of Damasio’s influential work the principal distinction between (prototypical, inborn) primary emotions and (learned, adult) secondary emotions was introduced. A connection to developmental psychology was drawn substantiating these two classes of emotions and, finally, further classes of emotions were presented.

In the WASABI architecture the distinction of three classes of affective states—mood, primary emotions, and secondary emotions—is followed together with distinguishing an agent’s cognitive abilities and its dynamics of bodily feeling as detailed in Chapter 4.

3 Related work in Affective Computing

In her book “Affective Computing” Rosalind Picard (1997) argues for the development of so-called “affective computers” that might serve the following purpose:

> “It is my hope that affective computers, as tools to help us, will not just be more intelligent machines, but will also be companions in our endeavors to better understand how we are made, and so enhance our own humanity.” (Picard 1997, p. xi)

Picard defines the term “affective computing” as “computing that relates to, arises from, or deliberately influences emotions” and emphasizes that this “includes implementing emotions, and therefore can aid the development and testing of new and old emotion theories.” (Picard 1997, p. 3) She compares the traditional AI approach of rule-based expert systems with rational laws and emotions with “songs” of a society and points out that “laws and rules are not sufficient for understanding or predicting human behavior and intelligence.” (Picard 1997, p. 5) This assumption is supported by psychological and neurobiological findings (cf. Chapter 2) that are extensively discussed by Picard (1997).

The term “Affective Computing” in itself, however, is questionable as Hollnagel (2003) believes. He gives two reasons for his statement that “affective computing” can be qualified as a “brainless phrase” (Hollnagel 2003, p. 65). In his opinion “computing by its very nature cannot be affective” and using the term to refer to a specific *type* of computing is misleading, because it can only refer to a specific *use* of computing. To support his arguments Hollnagel—with reference to Descartes—divides emotions into three aspects: “(1) the behavioral aspect, (2) the physiological aspect and (3) the subjective aspect (also called the introspective or phenomenological aspect).” (Hollnagel 2003, p. 66) He then pinpoints the computer’s lack of “anything similar to an autonomic nervous system” that is generally agreed on by emotion theorists to be a “sine qua non” for affect and emotion in humans. Because computers are purely based on logical information processing, “there is no way which they can be emotional or affective in the normal meaning of the words.” (Hollnagel 2003, p. 68)

Hollnagel contrasts the illusive term “Affective Computing” with the concept of “Effective Computing” (Freeman 1995) and suggests to use emotions to “improve the effectiveness of communication.” In his understanding, however, expressing the affective modality “by different means such as grammatical structure [..], the choice of words, or the tone of voice [..]” cannot be labeled “affective computing as such.” (Hollnagel 2003, p. 69) He summarizes this idea as follows:

> “Instead the style of computing—or rather, the style of communication or interaction—is *effectual*. It does not try to transmit emotions as such but rather settles for

> adjusting the style of communication to achieve maximum effectiveness.” (Hollnagel 2003, p. 69)

Freeman (1995) defines the aim of “effective computing science” as “consisting of a community of scholars with a strong *intellectual core of computer science,* coupled with emphasis areas that focus on interactions with other disciplines.” Computer science is understood as “an effective element of some larger, often real-world context.” (Freeman 1995, p. 28) Accordingly, “affective” computing can be described as one subfield of a much larger area of “effective” computing, whereby the term affective highlights the special interest in the influence of emotions and related concepts on the interaction between humans and computers.

In response to Hollnagel’s critical assessment of the term “Affective Computing” Hudlicka states more precisely the aims of researchers in Affective Computing:

> “One of the aims of the field is to answer precisely this question: When is affect helpful in human-machine communication? When should the machine recognize and respond to the user’s affect? And how? To answer these questions we must first construct machines capable of recognizing and ‘simulating’ affect. And that is precisely one of the aims of affective computing and affective HCI. [..] And one of the roles of affective computing is to better understand the capabilities (and limitations) of our affective-cognitive system, and thereby (hopefully) provide improved computer tools to assist us.” (Hudlicka 2003a, p. 74)

This discussion mainly arises due to the slippery nature of the underlying concepts “affect” and “emotion” that have not been defined precisely enough in scientific literature (cf. Chapter 2). This indetermination is especially problematic as one starts to program computers to recognize and simulate affect. With growing interest in more natural interaction with computers in the form of Embodied Conversational Agents (cf. Sections 1.2 and 3.2) or Social Robots (cf. Section 3.3), however, researchers have begun to integrate a certain level of affective competence into their agents’ architectures.

This Chapter gives an overview of related work in the field of “Affective Computing” which is still in the fledgling stages. The overview is split into general emotion architectures (cf. Section 3.1), architectures for virtual humans (cf. Section 3.2; cf. Vinayagamoorthy, Gillies, Steed, Tanguy, Pan, Loscos & Slater (2006) for an overview), and architectures for social robots (cf. Section 3.3; cf. Dautenhahn, Nourbakhsh & Fong (2003) for an overview), because different platforms bring about different affordances for simulating affect.

3.1 General Emotion Architectures

> “The need to cope with a changing and partly unpredictable world makes it very likely that any intelligent system with multiple motives and limited powers will have emotions.” Sloman & Croucher (1981)

When personal computers became affordable for everyone in the 1980s, the scientific community intensified the use of computers to evaluated their theoretical models of emotions. As reported in Section 2.1.3, Ortony et al. explicitly label their OCC-model a “computationally tractable model of emotion” (Ortony et al. 1988, p. 181) and accordingly propose conditional rules suitable for implementation.

Early computational models such as the "Affective Reasoner" by Elliott (1992) or the "Em" emotion module within the "Tok" architecture[1] are based on the OCC-model and were accompanied by an ongoing discussion about "cognitive-emotional interactions" (cf. Hudlicka 2003b; LeDoux 1995; Zajonc 1980), see also Section 2.1.2 (p. 23). Ortony et al. deliberately left out "other important aspects of emotion, such as the physiological, behavioral or expressive components" (Ortony et al. 1988, p. 2), which are necessary seed crystals of emotional episodes for researchers like James (1884) and Lange (1885) and recent neurobiological findings support their view (cf. Sections 2.1.1 and 2.2).

Staller & Petta (1998) present a comprehensive overview of the "Affective Reasoner" together with a critical discussion and Elliott (1994) himself discusses the problems of shallow architectures of emotions that are following the expert systems approach. Nevertheless, Elliott, Rickel & Lester (1997) used the Affective Reasoner as basis for emotion simulation in STEVE—one of the first three-dimensional virtual agents (Johnson & Rickel 1997). Johnson, Rickel & Lester (2000) give an excellent review of these early developments of so-called "Animated Pedagogical Agents".

The following section focuses on emotion architectures, that are not explicitly focusing on some kind of virtual agent. They rather present more general computational models of human emotions.

3.1.1 The H-CogAff architecture

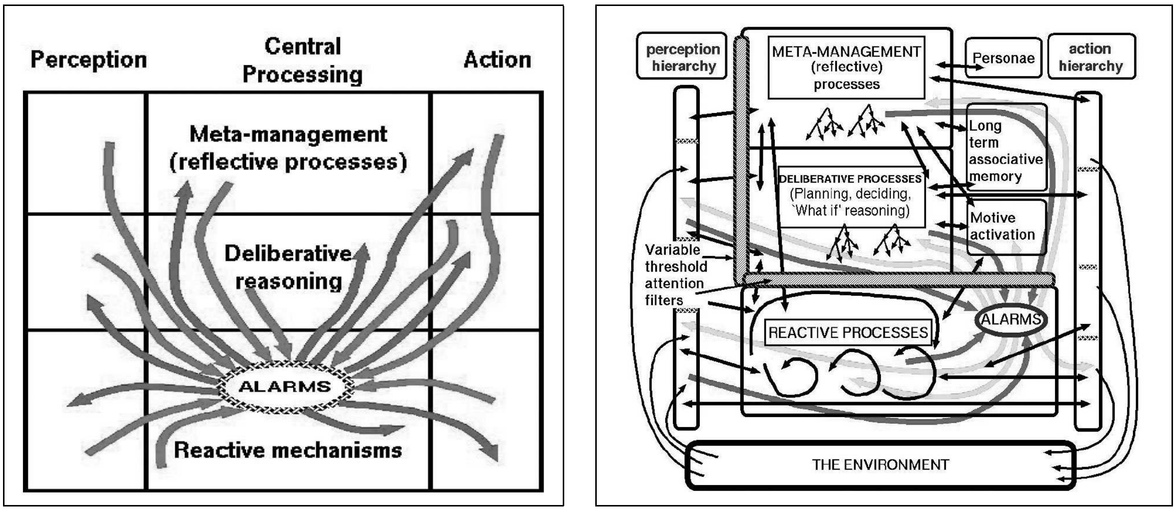
Based on his own distinction of three kinds of theories for modelling affect (presented in the beginning of Chapter 2 (p. 15)) Sloman (1992) is a proponent of design-based theories arguing in the following way:

> "I believe a proper analysis of the concept of an 'affective' state or process must be based on a more general theory of the coarse-grained architecture of mind. Such a theory, should describe the main sub-mechanisms, showing how they are related and how their causal roles within the total system differ. Various functions for mechanisms and states can be distinguished, but only relative to the whole architecture." (Sloman 1992, p. 233)

Therefore, his "H-CogAff architecture" (Sloman 1998, 2000; Sloman et al. 2005) (cf. Figure 3.1(b)) is derived as a special case from the more general "CogAff schema" (cf. Figure 3.1(a)) to "cover the main features of the virtual information-processing architecture of normal (adult) humans." (Sloman et al. 2005, p. 22) Sloman (1998) follows Damasio in distinguishing "primary" and "secondary" emotions (see also Section 2.2.2) but adds the class of "tertiary" emotions defined as "typically human emotional states involving partial loss of control of thought processes (perturbance), e.g. states of feeling humiliated, infatuated, guilty, or full of excited anticipation [..]." (Sloman 2000, p. 13)

Primary emotions might be elicited by an "alarm system" (cf. Figure 3.1(a)), which is believed to be activated by *reactive mechanisms* in case of emergency. According to Sloman et al. (2005), the general "perturbances" resulting from the alarm system's activation cause an interrupt in an agent's normal processing and this "actual or potential disturbance" is proposed

[1] The Tok architecture was developed in the context of the OZ project (Bates & Reilly 1992; Reilly 1996) as discussed in (Becker 2003).


Perception
Central Processing
Action
Meta-management (reflective processes)
Deliberative reasoning
ALARMS
Reactive mechanisms
perception hierarchy
META-MANAGEMENT (reflective) processes
Personae
action hierarchy
Long term associative memory
DELIBERATIVE PROCESSES (Planning, deciding, 'What if' reasoning)
Motive activation
Variable threshold attention filters
ALARMS
REACTIVE PROCESSES
THE ENVIRONMENT

(a) The general CogAff schema, cited from (Sloman et al. 2005, p. 20)

(b) The H-CogAff architecture, cited from (Sloman et al. 2005, p. 23)

Figure 3.1: Sloman's conception of an adult human's cognitive architecture

as a "very general definition of emotion." (Sloman et al. 2005, p. 25) The existence of primary emotions, in this view, only depends on the type of information processing that a given architecture supports.

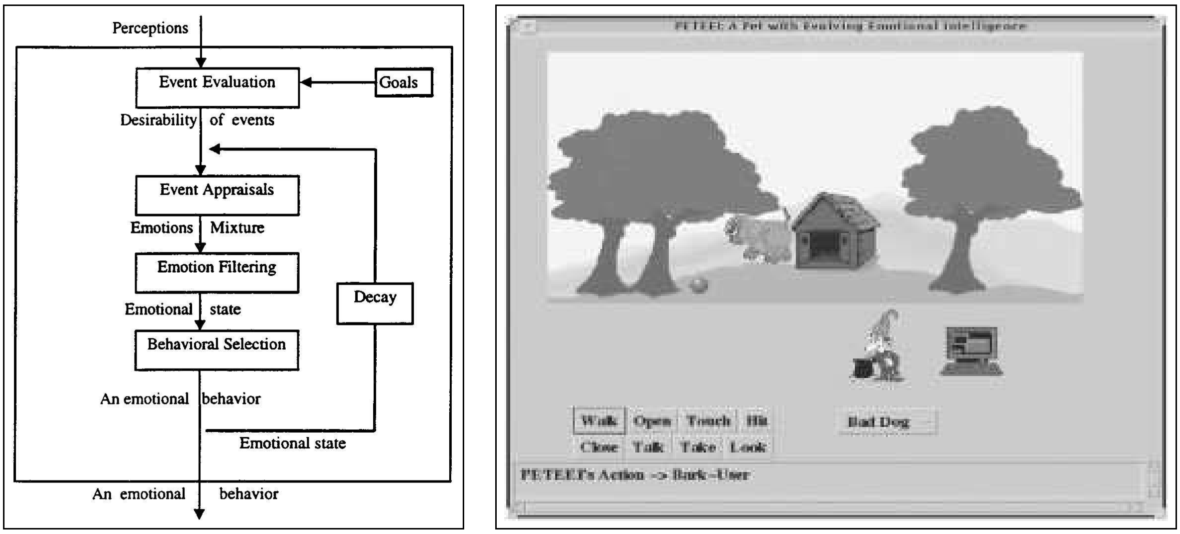
Consequently, as soon as *deliberative reasoning* can take place in an agent's architecture the elicitation of secondary emotions is assumed possible by Sloman et al. (2005). This layer (cf. Figure 3.1(b)) enables "Planning", "deciding", and "What if reasoning" and is, thus, comparable to Damasio's conception of secondary emotions (cp. Figure 2.12(b), p. 52). That the H-CogAff architecture is explicitly designed for an *adult* human goes in line with the discussion of ontogenetical development of emotions in Section 2.2.2, because young children still have to acquire the necessary ability to generate expectations based on prior experiences.

With the realization of *reflective processes* in an agent's architecture meta-management can give rise to "tertiary emotions", because they involve "actual or dispositional disruption of attention-control processes in the meta-management (reflective) system." (Sloman et al. 2005, p. 26) To this respect Sloman et al. go further than Damasio in proposing a third class of emotions, but they do not include Damasio's conception of background emotions (see Section 2.2.2).

In essence, cognitive appraisal in the CogAff architecture is realized along the lines of Ortony et al. (1988) and for Sloman et al. an agent's architecture must support "the ontological distinction between agents and objects." (Sloman et al. 2005, p. 31) Otherwise agent-based emotions such as being jealous cannot be represented.

3.1.2 FLAME: Fuzzy Logic Adaptive Model of Emotions

El-Nasr, Yen & Ioerger (2000) present a formalization of the dynamics of 14 emotions based on fuzzy logic rules (cf. Figure 3.2). The "emotional process component" starts with "Event Evaluation" (cf. Figure 3.2(a)). This process not only evaluates the importance of the goals that are affected by an event but also to what degree the event affects these goals. Fuzzy rules are applied to these values to calculate the event's "Desirability", which is then passed to the



PETEEI's Action -> Bark-User

(a) The "emotional process component" of FLAME, cited from (El-Nasr et al. 2000, p. 228)

(b) The user interface of PETEEI: A "PET with Evolving Emotional Intelligence", cited from (El-Nasr et al. 2000, p. 244)

Figure 3.2: The emotion process component of FLAME (El-Nasr et al. 2000) and a screenshot of the user interface of PETEEI (El-Nasr et al. 1999)

OCC-based "appraisal" process to determine the change in the emotional state. An emotion filter is applied next and an appropriate behavior is selected. Notably, the emotional state is "eventually decayed and fed back to the system for the next iteration" (El-Nasr et al. 2000, p. 227) letting this computational model also take into account the possible influence of previous emotional states.

Furthermore, a mood value is continuously calculated as the average of all emotion intensities. By introducing mood El-Nasr et al. provide a solution to the problem of conflicting emotions being activated at the same time. If, for example, an agent is in a negative mood and a positive emotion like *joy* has an intensity of 0.25 together with a negative emotion like *anger* having an intensity of 0.20, "the negative emotion inhibits the positive emotion" even though "the positive emotion was triggered with a higher intensity, because the agent is in a negative mood." (El-Nasr et al. 2000, p. 235)

FLAME further includes inductive algorithms for learning, enabling an agent to generate expectations based on rewards and punishments. It was successfully integrated into PETEEI (El-Nasr et al. 1999), an interactive simulation of a pet. Figure 3.2(b) presents a screenshot of PETEEI's graphical user interface by which the user can interact with the pet analogue to simple role-playing games. In the summary of their questionnaire-based evaluation of PETEEI's performance, however, El-Nasr et al. (2000) have to admit that the use of fuzzy logic was only useful to ease the integration of emotions, but did not contribute significantly to the perceived level of intelligence. They point to several future extensions including the use of FLAME for emotion modeling in virtual characters, in which case the additional integration of a personality model is argued for. They admit, however, that including personality in FLAME "would be a difficult but important task" (El-Nasr et al. 2000, p. 253) and some parameters of their model already account for personality related aspects of human behavior.

3.1.3 Émile and EMA: A computational model of appraisal dynamics

Gratch & Marsella (2004) present a domain-independent framework for modeling emotions that combines insights from emotion psychology with the methodologies of cognitive science in a promising way. Taking the "symbolic artificial perspective" Gratch & Marsella present a BDI-based approach[2] to integrate appraisal and coping processes in an agent's architecture that are central to emotion elicitation and social interaction.

Central to their idea are "appraisal frames and variables" by which the emotional value of external and internal processes and events are captured in concrete data structures. By making use of the agent's BDI-based reasoning power based on concepts such as likelihood and desirability, individual instances of emotion are first generated and then aggregated into a current emotional state and overall mood. An overall mood is seen to be beneficial, because it has been shown to impact "a range of cognitive, perceptual and behavioral processes, such as memory recall (mood-congruent recall), learning, psychological disorders (depression) and decision-making" (Gratch & Marsella 2004, p. 18). This mood value is also used as an addendum in the calculation of otherwise equally activated emotional states (such as fear and hope at the same time) following the idea of mood-congruent emotions.

The appraisal component is based on the work of Gratch (1999), who adapted Elliott's "Affective Reasoner" (Elliott 1992), which itself is based on the OCC-model of emotions (cf. Section 2.1.3). With Émile, a model of emotional reasoning, Gratch (2000) provides the first version of an emotion model to which the idea of "plan-based appraisal" is central. Gratch explains this idea in the following way:

> "Rather than appraising events directly, Émile appraises the state of plans in memory. [..] The relationship between events and an agent's disposition is derived more generally by a general purpose planning algorithm. [..] Émile replaces a large number of domain-specific construal frames needed by construal theory [as proposed by Elliott] with a small number of domain-independent rules."

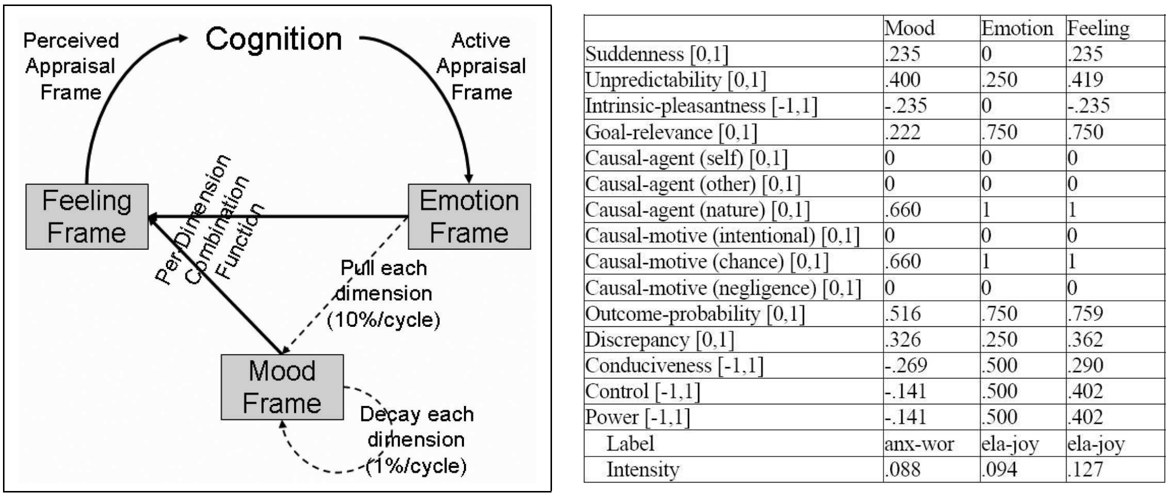
This idea has proven valuable and, thus, also underlies "EMA" (Marsella & Gratch 2006), in which dynamic aspects of appraisal are emphasized even more. Remarkably, their framework for modeling emotions is the first fully implemented, domain-independent architecture for emotional conversational agents.

3.1.4 Soar-Emote: Mood and Feeling from Emotion

With their "computational framework for emotions and feelings" Marinier & Laird (2004) aim to combine the work of Gratch & Marsella (2004) (cp. Section 3.1.3) with the findings of Damasio (1994) (cp. Section 2.2.2). In later publications (Marinier & Laird 2006, 2007), however, Damasio's work becomes less central and the authors follow the ideas of Scherer (2001) (cf. Section 2.1.3). The central idea of "appraisal frames" is based on the EMA model (see above) and Marinier & Laird (2007) explain in-depth how they model eleven of Scherer's

[2] Details of the Belief-Desire-Intention approach to modeling rational agents are given in Section 6.1.1.

sixteen appraisal dimensions[3] for integration in the Soar cognitive architecture, which also underlies the implementation of Gratch & Marsella (2004).


Cognition
Perceived Appraisal Frame
Active Appraisal Frame
Feeling Frame
Emotion Frame
Per Dimension Combination Function
Pull each dimension (10%/cycle)
Mood Frame
Decay each dimension (1%/cycle)

	Mood	Emotion	Feeling
Suddenness [0,1]	.235	0	.235
Unpredictability [0,1]	.400	.250	.419
Intrinsic-pleasantness [-1,1]	-.235	0	-.235
Goal-relevance [0,1]	.222	.750	.750
Causal-agent (self) [0,1]	0	0	0
Causal-agent (other) [0,1]	0	0	0
Causal-agent (nature) [0,1]	.660	1	1
Causal-motive (intentional) [0,1]	0	0	0
Causal-motive (chance) [0,1]	.660	1	1
Causal-motive (negligence) [0,1]	0	0	0
Outcome-probability [0,1]	.516	.750	.759
Discrepancy [0,1]	.326	.250	.362
Conduciveness [-1,1]	-.269	.500	.290
Control [-1,1]	-.141	.500	.402
Power [-1,1]	-.141	.500	.402
Label	anx-wor	ela-joy	ela-joy
Intensity	.088	.094	.127

(a) A "Feeling Frame" results from the combination of a "Mood Frame" with an "Emotion Frame" (Marinier & Laird 2007, p. 462)

(b) Exemplary combination of the intensities of a mood and an emotion frame resulting in a feeling frame (Marinier & Laird 2007, p. 466)

Figure 3.3: The Soar-Emote model: Computational Modeling of Mood and Feeling from Emotion (Marinier & Laird 2007)

Interestingly, Marinier & Laird claim to follow Damasio's distinction between emotion and feeling—that is a feeling as "the agent's perception of [an] emotion." (Marinier & Laird 2007, p. 461) With reference to James (1884) they mention the idea of feelings as conscious experience of bodily feedback (cp. Sections 2.1.1 and 2.2.2) and argue that this idea is captured by their concept of mood as a kind of "memory of recent emotions", which in combination with "the agent's appraisal of the current situation (emotion)" gives rise to feelings (cf. Figure 3.3(a)). They provide detailed functions for the calculation of a feeling's intensity based on given appraisal frames for emotion and mood.

In consequence, an "Active Appraisal Frame" (cf. Figure 3.3(a)), which is the result of a momentary appraisal of a given event, can be different from the "Perceived Appraisal Frame", which in turn results from the combination of the actual mood and emotion frames. In Figure 3.3(b) an example combination of the two appraisal frames "Mood" and "Emotion" is given resulting in a third appraisal frame labeled "Feeling". Thus, the intensity of a combined feeling can be higher than the maximum of each component in this model as given in the last line ("Intensity") of the Table presented in Figure 3.3(b).

In summary, Marinier & Laird (2007) present a promising alternative approach to computational modeling of emotions, even if their theoretical underpinning could be more elaborated. They complain much too often about the lack of empirical findings in support of their design decisions, even if such results can be found as will be detailed in later chapters of this thesis. Surprisingly, their approach resembles similar ideas as developed independently by Becker & Wachsmuth (2006a).

[3] In Table 2.6 (page 37) only thirteen SECs appear, because the "Novelty check" consists of three sub-checks and the "Causal attribution check" differentiates between "Cause: agent" and "Cause: Motive" (Scherer 2001, p. 114).

3.1.5 Summary and conclusion

Before going on with a discussion of virtual agents as embodied interaction partners it seems reasonable to reflect the previously presented models in the light of the Affect Simulation Architecture conceptualized in this thesis.

The H-CogAff architecture is a truly remarkable contribution to the field of “Affective Computing”. The principle distinction of three emotion classes arising from three different but highly interconnected architectural layers is supported by research in cognitive science (cf. Section 2.1.3) as well as neuroscience (cf. Section 2.2.2). Furthermore, it bears resemblance to the three processing levels of Ortony et al. (2005) presented in Figure 2.10, p. 45. Accordingly, the conceptual distinction of reactive mechanisms and deliberative reasoning is followed in this thesis not only with respect to our agent’s cognitive architecture in general (cp. Figure 1.3, p. 11), but also in the implementation of primary and secondary emotions.

FLAME provides an interesting concept of mood as an additional factor in the appraisal as well as the disambiguation process. The idea of expectation generation by means of learning based on user feedback is remarkable and the parameters of emotion dynamics in the Affect Simulation Architecture (cf. Chapter 4) can account for personality related factors similarly to the parameters of the FLAME architecture.

Émile and EMA have proven to be successful in a series of applications and are very good examples of OCC-based computational emotion models. However, David Traum (personal communication) had to admit that the high number of rules implemented in Soar are very difficult to administer and make it even more difficult to extend the system. As mentioned in the beginning of this section, Staller & Petta (1998) already criticized the brittleness of purely rule-based approaches to emotion modeling. Achieving a domain-independent architecture such as EMA is an important goal of recent research in Affective Computing and with the Affect Simulation Architecture the author aims to achieve a similar level of domain-independency.

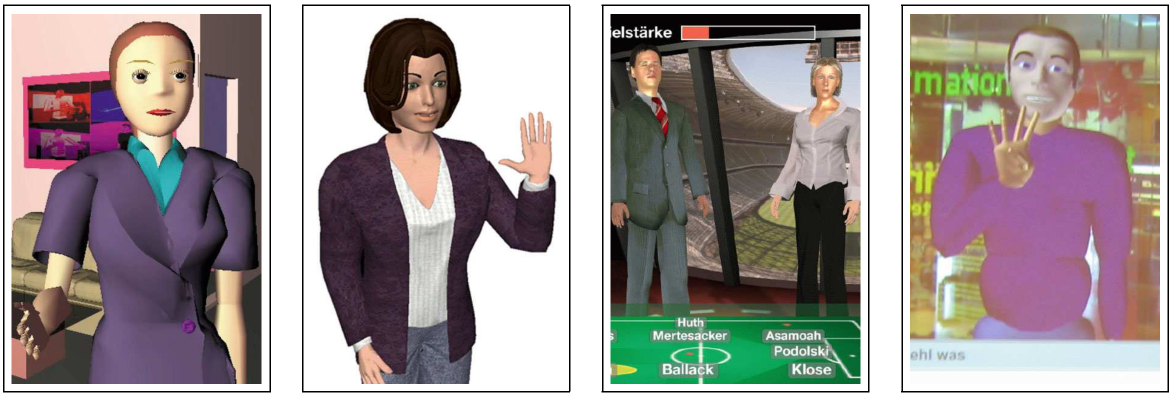
Soar-Emote could not yet keep its promise to provide a combination of mood and emotion resulting in feeling. Using the same data structure for all three affect-related concepts (emotion, mood and feeling) seems inappropriate, because the only aspect mood has in common with emotion is a valence component. Especially the appraisal dimensions *cause-agent* and *cause-motive* contradict the common definition of mood as a less object-centered affective concept. The underlying ideas of feelings as perceived emotions and mutual influence of emotion and mood, however, are taken up for the Affect Simulation proposed in this thesis.

3.2 Simulating Virtual Humans

Gratch, Rickel, André, Cassell, Petajan & Badler (2002) motivate the development of “virtual humans” in contrast to humanoid robots in their excellent review with the following words:

> “With the untidy problems of sensing and acting in the physical world thus dispensed, the focus of virtual human research is on capturing the richness and dynamics of human behavior.” (Gratch et al. 2002, p. 54)

They emphasize the importance of “psychology and communication theory to appropriately convey nonverbal behavior, emotion, and personality”, because of the high expectations people



(a) The Embodied Conversational Agent REA: A virtual "Real Estate Agent" (Cassell 2000a, p. 71)

(b) Greta: A 3D Embodied Conversational Agent (Pelachaud et al. 2008)

(c) Two "virtual characters" of the VirtualHuman project (Reithinger et al. 2006, p. 52)

(d) The virtual human MAX: A presentation agent in a computer museum (Becker et al. 2004, p. 161)

Figure 3.4: Four different approaches to the simulation of virtual humans: (a) The Real Estate Agent REA (cf. Section 3.2.1), (b) the ECA Greta (cf. Section 3.2.2), (c) two virtual characters of the VirtualHuman demonstrator system (cf. Section 3.2.3), and (d) the virtual human MAX as a guide to a museum (cf. Section 3.2.4)

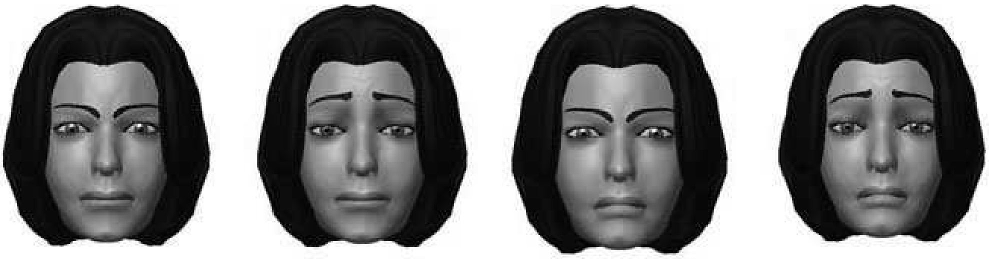
have when being confronted with humanoid agents. This section gives an overview of those virtual humans that are employed as "socially competent", multimodal interface agents.

3.2.1 REA: The Real Estate Agent

With respect to embodied conversational agents such as the "Real Estate Agent" REA (Cassell 2000a; Cassell, Bickmore, Billinghurst, Campbell, Chang, Vilhjailmsson & Yan 1999) (cf. Figure 3.4(a)), Bickmore & Cassell (2005) argue for the integration of non-verbal cues, whenever such agents are to take part in social dialogs.

In discussing the relationship between social dialog and trust they follow the multi-dimensional model of interpersonal relationship of Svennevig (1999) (cf. Svennevig & Xie 2002, for a review). This model distinguishes three dimensions namely *familiarity*, *solidarity* and *affect*, the last of which can be understood as representing "the degree of liking the interactants have for each other". (Bickmore & Cassell 2005, p. 30) In their implementation Bickmore & Cassell couple this dynamic parameter with the social ability of coordination, which in turn is seen as an outcome of fluent and natural small talk interaction. Coordination is understood as a means to synchronize short units of talk and nonverbal acknowledgement leading to increased liking and positive affect.

It is notable that the additional usage of non-verbal communicative means is sufficient to generate this kind of undifferentiated positive affect in the human user. In other words, Bickmore & Cassell (2005) believe that there is no need to simulate an embodied agent's internal emotional state to affect a human interlocutor positively.



(a) anger (b) superposition of sadness and anger (c) sadness masked by anger (d) sadness

Figure 3.5: Four emotional facial expressions of Greta: b) and c) are blendings of the two emotions anger and sadness (Ochs et al. 2005, p. 713)

3.2.2 Greta: A believable agent

With their development of the ECA "Greta" (cf. Figure 3.4(b)) Pelachaud & Bilvi (2003) are mainly concerned with believability of conversational interface agents. Consequently, their agent's facial expressivity (cf. de Rosis, Pelachaud, Poggi, Carofiglio & de Carolis 2003, for details) was extensively evaluated in the context of the European project "MagiCster" (de Rosis, Matheson, Pelachaud & Rist 2003)[4]. Based on the common hypothesis that the additional presentation of an embodied agent supports the human user's task performance, de Rosis et al. (2003) added video clips of Greta's face and synthetic voice to a dialog system. Several versions of this system were then compared to text only and human video settings in the healthy eating domain (Berry, Butler & de Rosis 2005). Even though the results of their study do not fully support the initial hypothesis, Berry, Butler, de Rosis, Laaksolahti, Pelachaud & Steedman (2004) see their definition of a methodology for evaluating the effects of animated characters as a positive result in itself. Furthermore, the consistency of facial expressions and message content was found to be most important.

To guarantee that Greta's facial expressions are always consistent with the situational context, de Rosis et al. (2003) model Greta's "mind" based on the BDI-approach by Rao & Georgeff (1991) (cp. Section 1.2.3). In their opinion, consistency is achieved as soon as Greta acts consistently "with her goal, her state of mind and her personality." (de Rosis et al. 2003, p. 86) In addition, her BDI-based "mental state includes a representation of the beliefs and goals that drive the feeling of emotions and the decision of whether to display or to hide them." (de Rosis et al. 2003, p. 87)

Greta's emotion model consists of a "Dynamic Belief Network (DBN)" (de Rosis et al. 2003, p. 95) and includes the event-based emotions of the OCC model (cp. Figure 2.9, p. 41). With dynamic belief networks the time dimension is integrated in the representation of uncertainty of beliefs. To this end, time is divided into *time slices* that resemble a state in the belief network. As soon as Greta's belief about the achievability of a goal changes or any goal is threatened by an event, the DBN is used to calculate the emotion on the basis of, first, uncertainty of beliefs and, second, utilities assigned to the achievement of goals. The domain-independency resulting from combining BDI and DBN is a clear advantage of this approach, but de Rosis et al. have to admit that "calibrating the prior and conditional probability tables

[4] According to (de Rosis et al. 2003, p. 111f), Greta was also tested in "a few toy dialogs" and in one other medical domain.

so as to avoid small, ‘spurious’ variations in the probability of monitored goals was rather difficult.” (de Rosis et al. 2003, p. 111)

Ochs, Niewiadomski, Pelachaud & Sadek (2005) present another BDI-based approach to implement OCC-based appraisal for Greta taking into account the socio-cultural context and integrating a computational model of emotion blending for facial expressions (cf. Figure 3.5). Recently, Ochs, Devooght, Sadek & Pelachaud (2006) extended their BDI-based emotion simulation to include the emotions “shame” and “pride” (cp. Figure 2.9, p. 41). They do not, however, provide facial expressions for these emotions. Greta’s abilities to mask her emotions are also explored in a gaming scenario by Rehm & André (2005). Their evaluation reveals a number of difficulties hindering the human player to notice Greta’s variations of communicative facial displays.

3.2.3 The VirtualHuman project

André, Klesen, Gebhard, Allen & Rist (1999) concentrate on designing believable “lifelike characters” by integrating models of personality and emotions. In their personality modeling they follow the Five Factor Model that consists of the five dimensions *Extraversion, Agreeableness, Conscientiousness, Neuroticism* and *Openness*[5]. Their computational model of emotions is based on the OCC model proposed by Ortony et al. (1988) (cf. Section 2.1.3).

Interestingly, André et al. (1999) distinguish primary and secondary emotions and discuss Sloman’s idea of tertiary emotions in the following way:

> “Primary emotions (i.e. being startled, frozen with terror, or sexually stimulated) are generated using simple reactive heuristics, whereas Secondary emotions are generated by the deliberative Affective Reasoning Engine according to the OCC model – Sloman introduces the additional class of Tertiary emotions as secondary emotions which reduce self control, but these will not be implemented in our initial prototype.” (André et al. 1999, p. 140)

In later publications (Gebhard 2005; Gebhard & Kipp 2006; Gebhard, Klesen & Rist 2004), however, this important distinction does not reappear. Gebhard et al. (2004) aim to improve “the quality of simulated conversations among virtual characters” letting the simulated affective states influence their character’s dialog contributions, way of articulation, and nonverbal expressions. With their OCC-based emotion simulation they concentrate on the *Well-being*, *Prospect-based*, *Attribution*, and *Attraction* clusters leaving aside OCC’s *Compound* and *Fortunes-of-others* emotions (cf. Figure 2.9).

Concerning the integration of a character’s personality Gebhard et al. (2004) opt for the Five Factor Model mentioned above. Every dimension of the personality model has a deterministic influence on the emotion’s intensity and decay, e.g. “an extravert character’s baseline intensity for joy is 0.15, whereas on introvert character’s baseline [..] would be 0.0.” (Gebhard et al. 2004, p. 132) They provide a graphical user interface for online manipulation of these complex interrelationships between five personality dimensions and 14 emotions. Furthermore, they integrate so-called “appraisal tags” and “dialog act tags” into a preexisting scripting language for dialog simulation.

[5] Taking the first letters of every dimension this model is also called the OCEAN model (cf. McCrae & John (1992) for an introduction).

In their discussion Gebhard et al. (2004) mention the simplifications they had to make to the OCC-model and that a better integration of the user in the dialog situation would be desirable. Consequently, Gebhard (2005) presents not only an extension to his emotion model, but also its integration into a new 3D environment that is developed in the context of the VirtualHuman project (cf. Figure 3.4(c)). In this environment the human user can participate in dialog by giving multiple choice answers. A layered model of affect (ALMA) is introduced, by which the intermediate affective quality "mood" is integrated into Gebhard's emotion model. The calculation of "mood" is based on a representation of the 14 OCC-emotions introduced above in Pleasure-Arousal-Dominance space of emotional meaning (cf. Section 2.1.2).

According to Gebhard (2005), eight types of mood can be distinguished and identified with the eight octants of PAD space as listed in Table 3.1[6]:

+P+A+D **Exuberant**	-P-A-D **Bored**
+P+A-D **Dependent**	-P-A+D **Disdainful**
+P-A+D **Relaxed**	-P+A-D **Anxious**
+P-A-D **Docile**	-P+A+D **Hostile**

Table 3.1: The eight mood octants in PAD space (Gebhard 2005, p. 31)

Gebhard (2005) implements the dynamics of mood as follows: If the current "active emotion" (i.e. its coordinates in PAD space) lies in another mood octant than the "current mood" (i.e. its coordinates in PAD space) then the mood is pulled in the direction of the emotion. If, however, the "active emotion" lies in the same mood octant as the "current mood" then the mood is intensified by pushing it away from the origin.

Gebhard & Kipp (2006) present an extension of the ALMA model to simulate 24 emotions together with a first evaluation based on textual interaction and a questionnaire. They conclude that the emotions and moods generated by ALMA are plausible, even if the eight moods were rated less distinguishable than the 24 emotions.

3.2.4 The virtual human MAX as a presentation agent

MAX is employed as a presentation agent in the Heinz-Nixdorf MuseumsForum (HNF; Paderborn, Germany). In this environment, the agent's task is to conduct multimodal smalltalk dialogs with visitors as well as to give explanations about the exhibition he is part of (Kopp, Gesellensetter, Krämer & Wachsmuth 2005). As the agent should be able to conduct natural language interactions, constraints on linguistic content (in understanding as well as in producing utterances) should be as weak as possible. Thus, a keyboard is used as input device, avoiding problems that arise from speech recognition in noisy environments. MAX responds to this input using synthetic speech, gesture, and facial expression.

The system's overall architecture (cf. Figure 3.6) is similar to those commonly applied in embodied conversational agents. It exhibits a two-level structure of concurrent reactive and deliberative processing, the latter being responsible for the agent's conversational capabilities. The emotion module—resulting from the author's diploma thesis (Becker 2003)—has been

[6] Unfortunately, they fail to support their argument against a simpler one-dimensional representation of mood with any scientific evidence, but only state that they "are convinced that mood is a complex affect type, like emotions are." (Gebhard & Kipp 2006, p. 344)

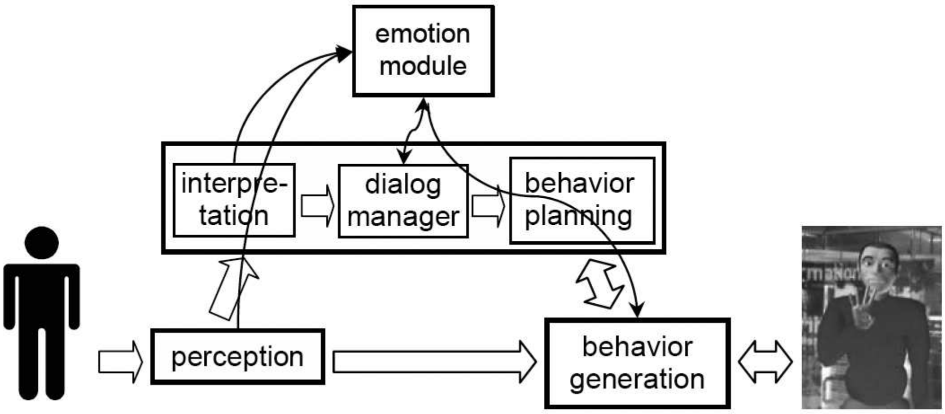


Figure 3.6: Integration of the emotion module into the agent's conversational architecture, cited from Becker et al. (2004)

added to this architecture as a separate module that incessantly receives input from and sends data to several other components as indicated by the arrows in Figure 3.6. Further details about this scenario and a dialog example with the corresponding trace of the agent's emotion dynamics are presented in Chapter 5.

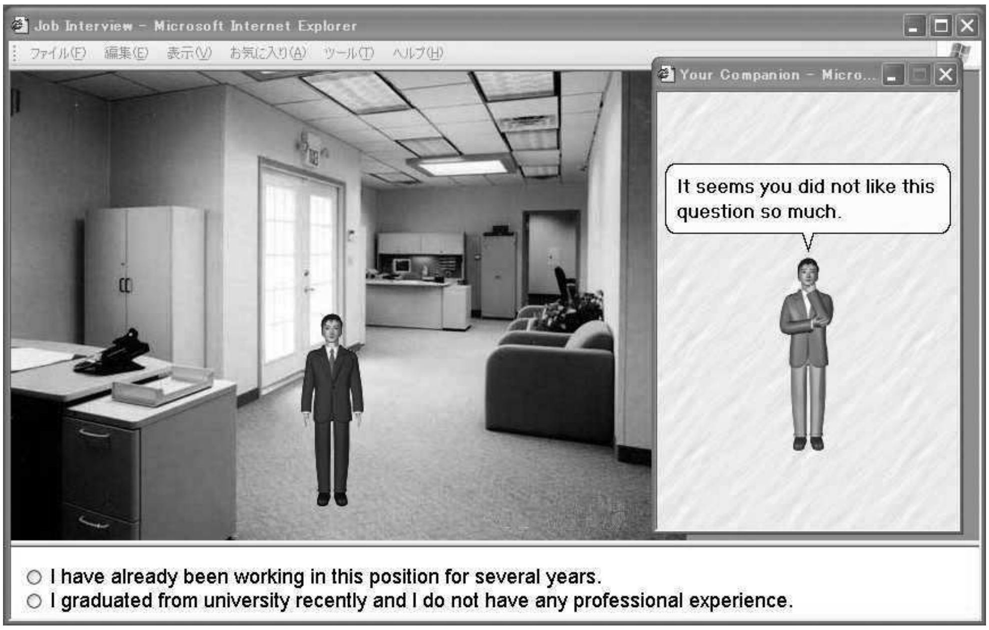
3.2.5 Life-like characters as empathic companions

Building on their experiences with the design and implementation of "socially intelligent agents" (Prendinger & Ishizuka 2001a) Prendinger & Ishizuka (2002) developed a scripting tool called SCREAM, which is based on sociological and psychological research. It enables an author to script the "mind" of an animated agent such as those presented in Figure 3.7. It combines a Prolog interpreter with a Java framework to implement the OCC model of emotions. Together with the Multimodal Presentation Markup Language (Prendinger, Saeyor & Ishizuka 2003) it was used to script a "Casino Scenario", in which three animated agent's (a dealer and two players) are playing "Black Jack" against the human, who is assisted by "Genie", an animated agent driven by the SCREAM engine. Accordingly, the human player can either follow or disregard Genie's advice letting Genie express a variety of emotions. Notably, the impact of different personality profiles (encoded according to the Five Factor Model as explained in Section 3.2.3) on the emotional reactions is also taken into account.

In their discussion, Prendinger, Descamps & Ishizuka (2002) mention the problem that "a rich repertoire of 'canned' affective verbal responses" have to be provided what is seen as a general problem of rather shallow, top-down approaches to emotion and personality simulation. As a possible solution they propose to abstract reactions to "good mood" and "bad mood" responses and to capture intensity levels by fuzzy labels like "neutral", "low intensity", and "high intensity". The second idea reminds one of FLAME described in Section 3.1.2.

Prendinger & Ishizuka (2001a) developed SCREAM with a special interest in modeling "social role awareness" (Prendinger & Ishizuka 2001b) and, thus, they include mechanisms for analysing and reacting to the human user's affective state with the development of "Empathic embodied interfaces" (Prendinger et al. 2004; Prendinger & Ishizuka 2005).

Prendinger & Ishizuka (2005) program animated agents to react on the user's affective state, which is derived from physiological activity in real time. Galvanic skin response and elec-


Job Interview – Microsoft Internet Explorer
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)
Your Companion – Micro...

It seems you did not like this question so much.

○ I have already been working in this position for several years.
○ I graduated from university recently and I do not have any professional experience.

Figure 3.7: The “Empathic Companion” in a job interview scenario (Prendinger et al. 2004, p. 57)

tromyography are tracked and analyzed by means of a Bayesian network that maps into two-dimensional emotion space (cf. Lang (1995); also Section 2.1.2) to derive one of the emotion categories fear, frustrated, sad, excited, joyful, or relaxed (Prendinger & Ishizuka 2005, p. 275), see also Chapter 5 for details.

Prendinger & Ishizuka (2005) applied this model to a “job interview scenario”, in which the human user is supposed to answer questions of an animated agent in the role of an interviewer (cf. Figure 3.7, left agent). During the study the human subject’s physiological activity was captured in form of galvanic skin response and heart rate. Prendinger & Ishizuka expected a positive effect of the empathic companion’s (cf. Figure 3.7, right agent) positively empathic remarks in case of a subject’s (assumed) frustration. They hypothesize that “[a]veraged over the entire interaction period, the presence of a (supportive) Empathic Companion will have users with lower levels of arousal and less negatively valenced affective states.” (Prendinger & Ishizuka 2005, p. 278) This hypothesis, however, could not be confirmed and Prendinger & Ishizuka assume that a more direct interaction between the companion and the human user would yield better results.

In Chapter 5 it is explained, how the simulation and expression of primary emotions with the virtual human MAX was combined with this emotion recognition system and evaluated in a competitive gaming scenario. As in this scenario the human player is directly playing against an “empathic” virtual human MAX, who is more expressive than the animated agents of Prendinger & Ishizuka (2005), the aforementioned problems are avoided.

3.2.6 Further models and architectures

The DER architecture: Dynamic Emotion Representation

Tanguy, Willis & Bryson (2003) present the “DER” (Dynamic Emotion Representation) architecture for the representation of “time-courses of internal states which underly complex, human-like emotional responses.” (Tanguy et al. 2003, p. 101) Inspired by Sloman’s H-CogAff architecture (cf. Section 3.1.1) they aim to provide a general-purpose architecture to support the communication-driven (such as REA, cf. Section 3.2.1) as well as the simulation-driven approaches (such as Greta, cf. Section 3.2.2) to modeling virtual agents.

Tanguy, Willis & Bryson (2006) concentrate on coherent emotional expressivity of facial animations and they claim to distinguish primary, secondary, and tertiary emotions (cf. Damasio 1994; Sloman et al. 2005). This distinction, however, gets somehow lost in the final architecture, because primary as well as secondary emotions are labeled similarly to each other. Tanguy (2006) labels one of his six primary emotions with “Sad” and a (probably corresponding) secondary emotion with “Sadness”. Furthermore, a connection is drawn between the concept of mood and Sloman’s meta-management layer (cf. Figure 3.1(b), p. 62), from which tertiary emotions are assumed to arise. Tanguy et al. (2006) introduce mood with reference to Thayer (1996) consisting of the two dimensions “calm/tense” and “energy/tiredness” (Tanguy et al. 2006, p. 297). Notably, they also discuss a possible mapping of these dimension into the Pleasure-Arousal space (cf. Section 2.1.2) in that *pleasure* equals *energy-calm*, *displeasure* equals *tired-tense*, *sleep* equals *tired-calm*, and *arousal* equals *energy-tense*.

Generic Personality and Emotion Simulation

Egges, Kshirsagar & Magnenat-Thalmann (2003, 2004) propose a generic model for the integration of personality, mood and emotion into virtual humans. By presenting detailed update functions that operate with high-dimensional vectors representing emotions and moods Egges et al. (2004) do not limit their framework’s applicability to certain emotion theories. Their own approach consists of a combination of the OCC model of emotions and the Five Factor Model of personality (cp. Section 3.2.3). An intermediate concept “mood” is introduced as an (in principle multi-dimensional) affective quality of longer duration than emotion but being less persistent than personality related aspects of a virtual human. They have to admit that they do not model the influence of a prevailing mood on the elicitation of emotions, but only derive mood from emotions and personality factors. A stable personality, for example, has the effect of smaller mood changes than an unstable personality in their simulation. The simulated mood effects the virtual human only indirectly by modulating the agent’s emotions.

Notably, Egges (2006) argues against the use of the OCC model when aiming at “emotional motion synthesis”, because their 22 emotions are assumed to be “too detailed with respect to what people can actually perceive.” (Egges 2006, p. 60) He opts for the two-dimensional activation-evaluation space with reference to Schlosberg (cp. Section 2.1.2[7]).

3.2.7 Summary and conclusion

Approaches to simulating affect for virtual humans are traditionally based on the OCC model of emotions, but recent developments start to include or at least acknowledged the existence of dimensional theories. Many researchers integrate medium (mood) and long-term (personality) affect-related concepts in their implementations—mostly concentrating on conversational systems. Accordingly, a number of high-level scripting languages exist that support the annota-

[7] Schlosberg’s emotion cone presented in Figure 2.5(a), p. 27, is not referenced by (Egges 2006, p. 25).

tion of affect. Some researchers successfully base their virtual human's cognitive architectures on the BDI-approach and exploit its high-level concepts to support domain-independency of appraisal mechanisms.

The diversity of the presented approaches shows that the problem of endowing virtual humans with emotions is still far from being solved. For certain scenarios, however, a significant progress could be achieved—not only with respect to virtual humans but also in the field of social robotics discussed next.

3.3 Social robots

The term "Social Robots" refers to robotic systems that act autonomously in our social environment and are able to "communicate, coordinate and engage in complex social behavior" (Duffy, Dragone & O'Hare 2005, p.18). As pointed out by Duffy et al. (2005), social robot research can be divided into the bottom-up and the top-down approach. Where the bottom-up approach tries to enhance given robotic systems with abilities to participate in social interaction (or, at least, appear to behave socially competent), most researchers following the top-down approach explicitly design their robots with anthropomorphic qualities such as humanlike faces and bodies (Duffy 2003).

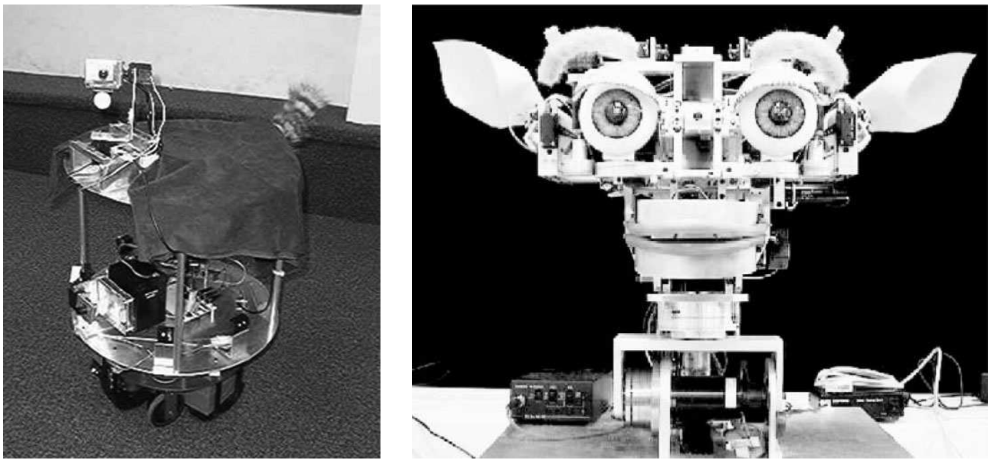
More precisely, the term "Social Robots" also includes "collective robots" that behave socially only among themselves but not toward a human. Explicitly excluding this class of robots, Dautenhahn, Nourbakhsh & Fong (2003) introduce the class of "Socially Interactive Robots" for which the following properties of human-human interaction competencies are assumed to be relevant (Dautenhahn et al. 2003, p.146):

- express and/or perceive emotions
- communicate with high-level dialog
- learn/recognize models of other agents
- establish/maintain social relationships
- use natural cues (gaze, gestures, etc.)
- exhibit distinctive personality and character
- may learn/develop social competencies

In the following, an overview of socially interactive robots is given with a special interest in those robotic systems, with which researchers follow the top-down approach, because the anthropomorphic features of these robots make their socio-emotional behavior better comparable to the virtual human MAX presented in this thesis.

3.3.1 Cathexis: Yuppy and Kismet

In this subsection the emotion architecture "Cathexis" is explained first, because it underlies the emotional capabilities of two sociable robots "Yuppy" and "Kismet" explained thereafter.



(a) The emotional pet robot “Yuppy” (Velásquez 1998, p. 74)

(b) The sociable robot “Kismet” (Breazeal 2003, p. 123)

Figure 3.8: The emotional pet robot Yuppy and the sociable robot Kismet

Cathexis

Velásquez & Maes (1997) introduce a computational model of basic and complex emotions with a special focus on time-dependency as well as the influence of emotions on behavior and motivation. In their “Cathexis Architecture” emotions, moods, and temperament are distinguished and modeled as a network of “special emotional systems” that each represent a “specific emotion family”, i.e. one of the “basic or primary” emotions *Anger*, *Fear*, *Distress/Sadness*, *Enjoyment/Happiness*, *Disgust*, and *Surprise* (cf. Section 2.1.1, p. 18).

Velásquez distinguishes emotions from moods in terms of arousal levels understanding moods as affective phenomena with a lower arousal than emotions. Thereby, he accounts for the predisposition to experience mood-congruent emotions as later supported by empirical studies of Neumann, Seibt & Strack (2001). Velásquez’ concept of temperament is quite similar to that of personality reported in Section 3.2.3. It is modeled by different activation thresholds of emotions.

Furthermore, every emotion can have an inhibitory or excitatory effect on each other emotion in the network and by means of an integrated learning algorithm the agent can generate secondary emotions by associating primary emotions with their releasers as proposed by Damasio (1994) (cf. Section 2.2). After a first evaluation with a baby-like synthetic agent “Simón the toddler” (Velásquez 1997) the architecture’s performance was tested on the robotic agent Yuppy (Velásquez 1998, cf. Figure 3.8(a)) before it was integrated into the more expressive sociable robot “Kismet” (Breazeal & Velásquez 1998, cf. Figure 3.8(b)).

Yuppy, an Emotional Pet Robot

In order to explore the ability of the Cathexis architecture to form emotional experiences, which are seen as the basis for the acquisition of secondary emotions, the robotic pet Yuppy was built (Breazeal & Velásquez 1998, cf. Figure 3.8(b)). Humans can wave a hand at the robot or use a toy to play with it. Yuppy perceives the human’s action by means of video and audio sensors.

The above-mentioned concept of temperament does not reappear in (Breazeal & Velásquez 1998; Velásquez 1998) but a new concept called “drives” is introduced complementing emotions and residing in the “motivation system”. Breazeal & Velásquez (1998) argue for the integration of three drives as presented in Table 3.2, because their goal is to let “human caretakers” teach the “infant” robot.

Social drive	This drive’s activation ranges from *lonely* at its low end to *asocial* at its high end and represents the robot’s need for sociality.
Stimulation drive	With an activation ranging from *bored* to *distressed*, this drive captures the robot’s need for stimulation, which can either result from external or internal activity, i.e. self-play.
Fatigue drive	This special drive resembles the robot’s need to *sleep* and when a high activation occurs all other drives are reset to their homeostatic regimes “so that the robot is in a good motivational state when it awakens.” (Breazeal & Velásquez 1998, p. 33)

Table 3.2: The three drives of Yuppy and Kismet

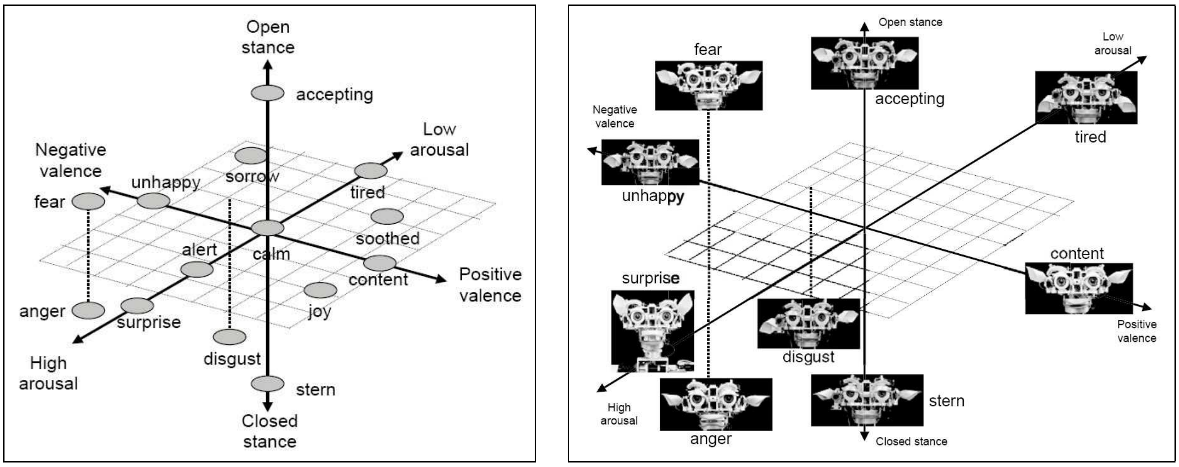
In summary, a human caretaker can influence Yuppy’s long-term behavior by giving feedback in the form of reward or punishment. For example, when Yuppy perceives a bone (as a hard-wired releaser of “happiness”) in the human’s hand, the robot approaches him. Now it depends on the human’s action toward the robot (i.e., if he pets or hits it) if Yuppy will learn to approach or avoid humans in the future (Velásquez 1998). The same mechanism can be exploited to learn other releasers of fear and Velásquez (1998) compares this implementation to LeDoux’s work on fear conditioning (cf. Section 2.2.1).

Kismet

Breazeal (2003) extends the Cathexis architecture by introducing a three-dimensional emotion space, which consists of the dimensions *Arousal*, *Valence*, and *Stance*. Compared to the dimensional theories discussed in Section 2.1.2, the third dimension, labeled “stance”, seems to be less well-founded. Breazeal explains that this dimension specifies “how approachable the percept is to the robot” with positive values corresponding to advance and negative ones to retreat. Directly compared with Plutchik’s “basic behavioral patterns” underlying his proposal of primary or basic emotions (cf. Table 2.1, p. 20), this definition of stance has much in common with the “Exploration” and “Rejection” behaviors, which—according to Plutchik (1980)—form the basis for the primary emotions “Anticipation” and “Disgust”.

According to Figure 3.9(a), however, open stance is associated with the emotion “acceptance” and close stance with the emotion “stern”. “Disgust” falls into the closed stance layer and is characterized by negative valence. Furthermore, this third dimension is again useful to distinguish “anger” and “fear”, even if it does not reflect a dominance or power relationship between robot and human in this model.

Breazeal (2003) assigns prototypical, facial expressions of Kismet to the fourteen emotions and nine of them are shown in Figure 3.9(b). An emotion is activated after “a myriad of environmental and internal factors” have been mapped into the three-dimensional affect space to “assess” them “affectively” (Breazeal 2003, p. 140) based on Damasio’s “Somatic Marker



(a) Fourteen emotions located in Arousal-Valence-Stance space (Breazeal 2003, p. 135)

(b) Corresponding facial expressions of the sociable robot "Kismet" (Breazeal 2003, p. 141)

Figure 3.9: Emotional categories mapped into Arousal-Valence-Stance space [A, V, S] and Kismet's corresponding facial expressions

Hypothesis" (cf. Section 2.2.2). In this process each so-called releaser is "tagged" according to its influence on arousal, valence and stance, which are to be hard-coded by the robot designers. For example, achieving a goal is marked with positive valence, whereas "delayed progress is marked with negative valence." (Breazeal 2003, p. 133) After a net sum of these AVS-vectors is calculated, a winner-takes-all strategy is applied to determine the active emotion. A complex interaction between emotional expression, situational context, and behavioral tendency is integrated to assure a coherent behavior and to calculate an emotions intensity—many parameters are determined empirically (see Breazeal 2002, for details).

3.3.2 Emotion Expression Humanoid Robot WE-4RII

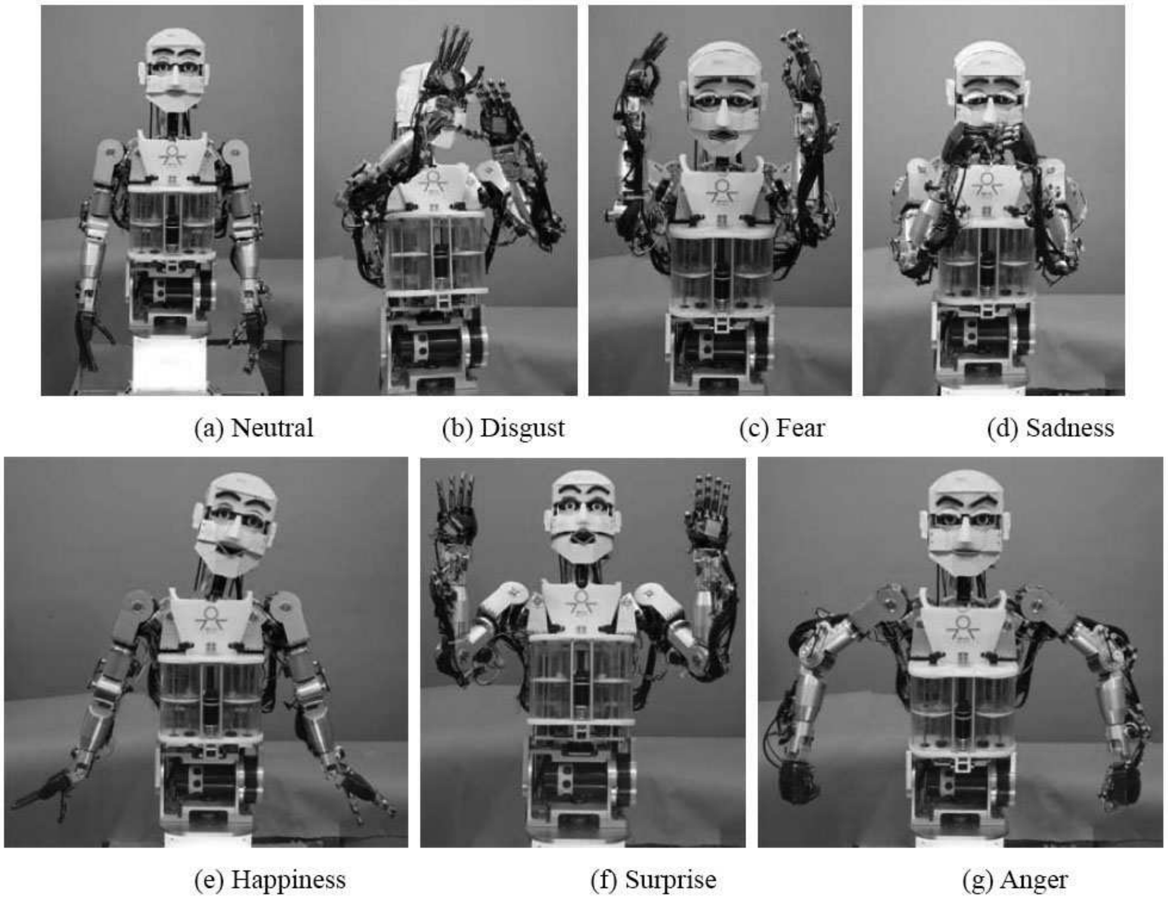
Zecca, Roccella, Carrozza, Miwa, Itoh, Cappiello, Cabibihan, Matsumoto, Takanobu, Dario & Takanishi (2004) discuss their humanoid robot WE-4RII (Waseda Eye #4 Refined II, cf. Figure 3.10), which is capable of expressing the six basic emotions proposed by Ekman (1999b) (cf. Section 2.1.1, p. 18) with its whole body.

These emotional expressions are triggered by an emotion system, which is based on a so-called "3D Mental Space" (cf. Figure 3.11(a)) consisting of the dimensions *pleasantness*, *arousal*, and *certainty*. This space reminds one of the three-dimensional emotion spaces discussed in Section 2.1.2, but no rationale is given for labeling the third dimension *certainty* instead of *dominance* (Russell & Mehrabian 1974) or *power* (Gehm & Scherer 1988).

The six basic emotions (plus neutral) introduced above are mapped into 3D Mental Space as presented in Figure 3.11(b) and a trajectory of an "Emotion Vector E" through this space is generated (cf. Figure 3.11(a)). This trajectory is calculated according to equation 3.1.

$$M\ddot{E} + \Gamma\dot{E} + KE = F_{EA} \tag{3.1}$$

M, Γ, and K are introduced in equation 3.1 as matrices representing the "Emotional Inertia", "Emotional Viscosity", and "Emotional Elasticity", respectively. The "Emotional Appraisal" F_{EA} is considered to capture "the total effects of internal and external stimuli on the



(a) Neutral (b) Disgust (c) Fear (d) Sadness

(e) Happiness (f) Surprise (g) Anger

Figure 3.10: The neutral (a) and six basic emotional expressions of WE-4RII (Zecca et al. 2004, p. 245)

mental state.” (Itoh et al. 2006, p. 267) The robot’s expressive reactions to a stimulus can be changed by adjusting the three “Emotional Coefficient Matrices.”

Itoh et al. (2006) also define mood as a vector in the two-dimensional *pleasantness* and *arousal* subspace according to the following equations:

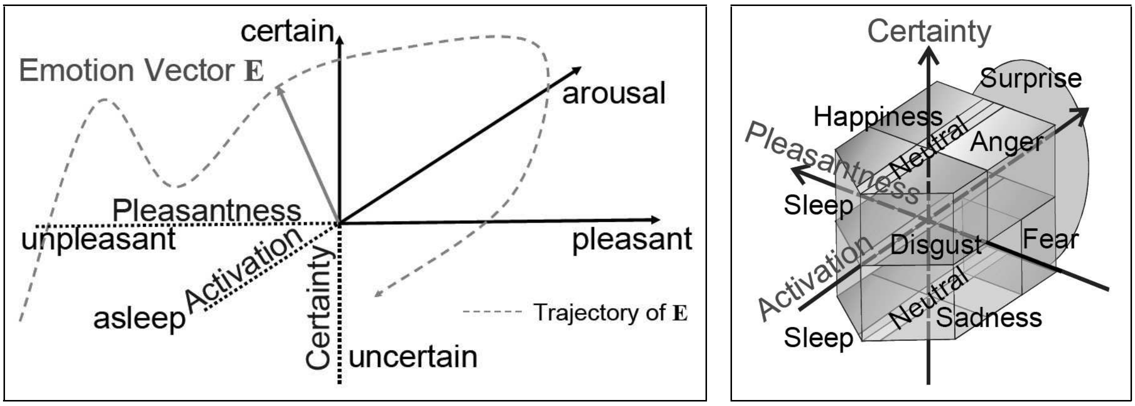
$$M = (M_p, M_a, 0), \tag{3.2}$$

$$M_p = \int E_p dt, \tag{3.3}$$

$$\ddot{M}_a + (1 - M_a^2)\dot{M}_a + M_a = 0 \tag{3.4}$$

Thus the pleasantness component M_p of mood is defined as the integral over the emotional pleasantness component E_p in equation 3.3. With the arousal component of mood Itoh et al. (2006) aim to model a human’s biological rhythm by means of a simulated Van der Pol oscillator with equation 3.4. Miwa, Itoh, Takanobu & Takanishi (2004) describe the calculation of the Emotional Appraisal F_{EA} in more detail as presented in equation 3.5.

$$\begin{aligned} F_{EA} &= f_{EA}(M, P_S), \\ &= k_m \times M + P_S \\ k_m &: Mood\ Influence\ Matrix \\ P_S &: Sensing\ Personality\ Vector \end{aligned} \tag{3.5}$$



(a) The “3D Mental Space” together with an “Emotion Vector E” (Itoh et al. 2006, p. 267)

(b) Neutral and six basic emotions mapped into “3D Mental Space” (Itoh et al. 2006, p. 267)

Figure 3.11: “3D Mental Space” consisting of pleasantness, activation, and certainty and how six basic emotions (cp. Figure 3.10) are mapped into this space

The “Sensing Personality” P_S is continuously updated against internal and external stimuli and, thus, an emotion dynamics in “3D Mental Space” is achieved and represented by the “Emotion Vector E” (cf. Figure 3.11(a)).

3.4 Summary

This chapter provided an overview of those previous and ongoing work that is related to the upcoming field of Affective Computing. After the problematic terms “affect” and “emotion” were discussed, four general emotion architectures were introduced in Section 3.1. They provide useful ideas and concepts for the integration of affective phenomena into cognitive architectures, even if the validity of the proposed approaches is not yet proven satisfactorily.

Section 3.2 addressed those architectures that underly seven different approaches to realizing virtual humans, which are non-physical, anthropomorphic agents mostly used as interface agents. Naturally, researchers in the field of virtual humans put emphasis on the expressive aspect of emotions by integrating and validating a variety of verbal and, especially, non-verbal means to let their agent’s express their emotional state. The approaches taken to achieve this goal are still manifold (and sometimes confusingly complex) and the results of empirical studies difficult to compare. Nonetheless, a general trend toward the integration of dimensional emotion theories can be found.

Taking the step from the virtual into the physical world the field of social robots was exemplarily introduced in Section 3.3 by discussing architectures underlying the realizations of three robotic agents with social interactivity and emotional expressivity. With increasing anthropomorphic realism of these robots they are more and more capable to express their internal states including their simulated emotions. Interestingly, dimensional emotion theories seem to be favored by roboticists—probably because the immediate interaction dynamics evoked by robotic agents suggests an equally dynamic approach to emotion simulation as it is more easily provided by dimensional emotion theories.

4 Conceptualization of affect simulation

The previous chapters illustrated the ambiguity of the concepts emotion, mood, and personality. All of them belong to the class of affective phenomena and are, thus, related to the field of Affective Computing introduced in Section 3. In his diploma thesis (Becker 2003; Becker et al. 2004) the author successfully implemented an emotion dynamics simulation system for the virtual human MAX, which was, however, limited to a simulation of more infant-like emotions. The Affect Simulation Architecture conceptualized here builds upon this previous work as it has proven to support the agent's believability in two different interaction scenarios.

This Affect Simulation Architecture combines bodily emotion dynamics with cognitive appraisal in order to simulate infant-like primary emotions as well as cognitively elaborated secondary emotions. In the following a suitable specification of the different concepts *emotion*, *mood*, and *personality* is derived from the theoretical background before a conceptual outline of the architecture is given.

4.1 A working definition of affective phenomena

With respect to the computational simulation of affect for an embodied agent the following differentiations are derived from the previous chapters[1]:

- **Emotions** result from complex neurophysiological processes and are often summarized by verbal labels, which naturally possess a mutlitude of connotations.

- **Mood** is understood as a background state with a much simpler affective quality than emotions.

- **Personality traits** are understood as a character's static dispositions to appraise environmental stimuli and, consequently, to react more or less emotional to them.

These three classes of affective states are now discussed in detail.

4.1.1 Emotions

Emotions are characterized by the following aspects:

[1] Especially important is Scherer's definition presented and discussed in the context of appraisal theories on page 36.

1. The processes underlying emotions include neural activity of the brain as well as physiological responses of the body.

2. One gets aware of one's emotions in two cases: (1) if their activity exceeds a certain threshold or (2) if one concentrates on the underlying processes by means of introspection.

3. Emotions can be classified into primary and secondary ones. A class of tertiary or social emotions is proposed as well.

4. In most cases an emotion is object-centered in that its eliciting object is known to an emotion experiencing individual, but false attributions are possible as well.

5. Every emotion has either positive or negative valence with a certain intensity and an emotion only lasts for a certain duration.

The elicitation of emotions is certainly a complex process. For the computational model presented here the idea of cognitive processes in combination with physical responses as discussed in Sections 2.1.1 and 2.2 is central. Accordingly, the distinction of primary and secondary emotions as proposed by Damasio (1994) as well as Sloman (2000) is followed. In order to successfully implement these two classes within the Affect Simulation Architecture they are specified more precisely next.

Primary emotions

Primary emotions (PE) are introduced in Section 2.2.2 as inborn affective states, which are triggered by reflexes in case of potentially harmful stimuli. In Sloman's theory (cf. Section 3.1.1) primary emotions are triggered in a similar way by so-called "alarm systems" and result in "perturbances" of the cognitive system.

In both cases primary emotions result in fast, reactive behavioral responses and, thus, are quite similar to the concept of proto-affect proposed by Ortony et al. (2005) (cf. Section 2.1.3, p. 44). According to developmental psychology, young children express their (primary) emotions directly, because they have not yet internalized this process as in the case of adults (cf. Section 2.2.2).

Implications for the thesis In the author's diploma thesis (Becker 2003) this direct expression of primary emotions is realized by implementing five of Ekman's six "basic emotions" as discussed in Section 2.1.1, p. 18. In addition, the emotions "bored", "annoyed", and "depressed" as well as the non-emotional state "concentrated" are also simulated in Becker et al. (2004). Every primary emotion (PE) is located in PAD space (cf. Section 2.1.2) according to Table 4.1.

Naming emotions is notoriously difficult and little agreement exists (cf. Chapter 2). As the virtual human MAX can produce facial expressions, Becker (2003) decided to first concentrate on Ekman's "basic emotions". Accordingly, the labels for the primary emotions (PE) in Table 4.1 are not to be confused with those emotions that are discussed in the context of appraisal theories of emotion in Section 2.1.3. The primary emotion "anger", for example, is one of the most complex emotions in the OCC-model of emotions (cp. Figure 2.9, p. 41). In

PE final (*initial*)		Facial expr. (Ekman)	PAD final	PAD initial
1.	1x angry	anger (*anger*)	(80, 80, 100)	*same*
2.	1x annoyed	sad (*sadness*)	(-50, 0, 100)	*same*
3.	1x bored	bored (*none*)	(0, -80, 100)	*same*
4.	2x concentrated	neutral (*none*)	(0, 0, ±100)	*same*
5.	1x depressed	sad (*sadness*)	(0, -80, -100)	*same*
6.	1x fearful	fear (*fear*)	(-80, 80, 100)	*same*
7.	4x happy (*2x friendly*)	happy (*happiness*)	(80, 80, ±100) (50, 0, ±100)	*(50, 0, ±100)*
8.	1x sad	sad (*sadness*)	(-50, 0, -100)	*same*
9.	2x surprised	surprised (*surprise*)	(10, 80, ±100)	*(80, 80, ±100)*

Table 4.1: Primary emotions in PAD space: The initial labels and PAD values in *italics* were proposed in Becker et al. (2004) and later revised with the final labels and values. The initial term “friendly” was changed to “happy” (see number seven) to better correspond to Ekman’s “basic emotion” happiness. The five “basic emotions” of Ekman (1999b) are assigned to the corresponding facial expressions modeled in Becker et al. (2004) whenever such a mapping is possible (cp. Figure 4.1)

Ortony’s opinion, “frustration” could be interpreted more basic than “anger”, because in case of anger another agent’s blameworthy action is the eliciting condition, whereas frustration can be experienced regardless of the presence of other agents (personal communication, 2007).

For the present purpose of triggering appropriate facial expressions, anger is understood as a label for an undifferentiated, reactive, behavioral response tendency in line with Plutchik’s “basic behavioral patterns” presented in Table 2.1 (p. 20). The facial expression accompanying this kind of “primary anger” is already imitated by one year old children even before they are capable of attributing mental states to others, which is believed necessary for the complex form of anger mentioned above.

The seven facial expressions of MAX corresponding to the eight primary emotions and the neutral state “concentrated” (cf. Table 4.1) are shown in Figure 4.1. The primary emotion’s locations in Figure 4.1 result from the final PAD triples “PAD final” in Table 4.1, such that “happy” is represented four times in PAD space and “surprised” as well as “concentrated” two times. These coordinates are derived from the values given in (Russell & Mehrabian 1977, p. 286ff), of which a selection is presented in Table 2.4 and Figure 2.6 on page 29.

In case of high pleasure Ekman’s set of “basic emotions” only contains one obviously positive emotion, namely happiness (Ekman et al. 1980). Thus, in the presented implementation this primary emotion covers the whole area of positive pleasure regardless of arousal or dominance as it is located in PAD space four times altogether. The distribution of primary emotions in PAD space proposed here is quite similar to the distributions proposed by Breazeal (2003) (cf. Figure 3.9(a), p. 77) for the sociable robot Kismet and proposed by Itoh et al. (2006) (cf. Figure 3.11(b), p. 79) for the humanoid robot WE-4RII. As discussed in Section 3.3 their choices of three-dimensional emotion spaces, however, seem to be less well-founded in the theoretical background.

How and when these facial expression are triggered within the Affect Simulation Architecture is explained in Section 4.2, in which the simulation of emotion dynamics is detailed.

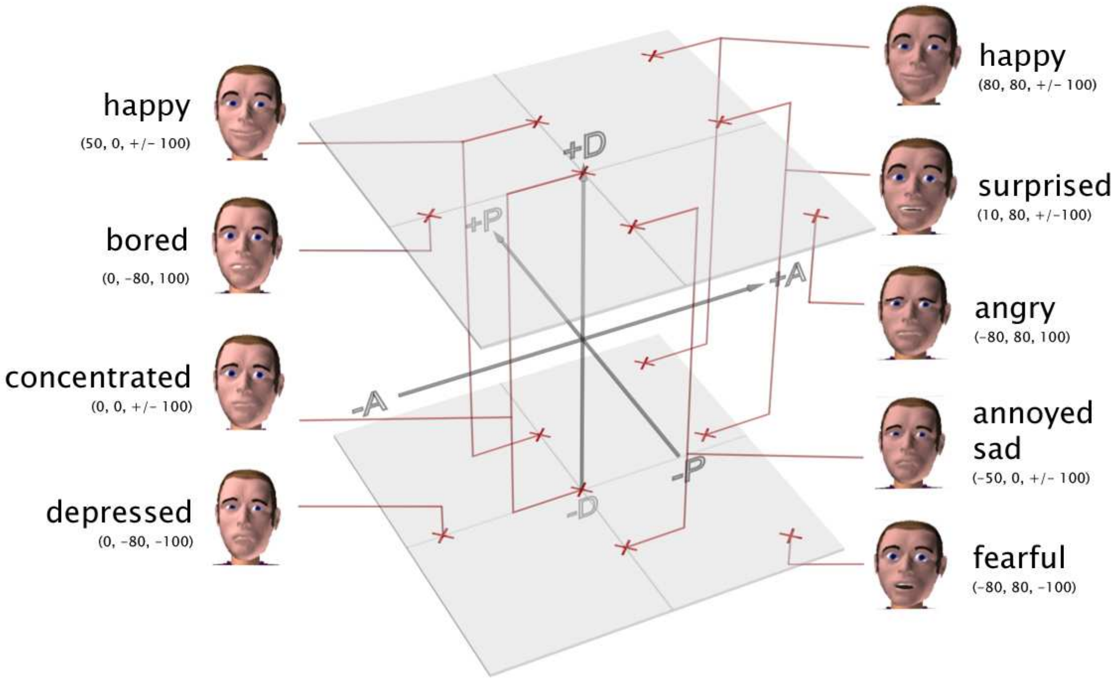


Figure 4.1: Seven facial expressions corresponding to the eight primary emotions plus “concentrated” (cp. Table 4.1)

For this dynamics another affective quality is important, namely the concept of “mood”. Before this concept is introduced, a computationally tractable conceptualization of secondary emotions is specified next.

Secondary emotions

According to Damasio (cf. Section 2.2.2), the elicitation of secondary emotions involves a “thought process”, in which the actual stimulus is evaluated against previously acquired experiences and online generated expectations. Taking developmental aspects into account, even causes of events that were perceived unemotionally at first can be marked emotionally during ontogenesis.

As cited in Section 2.2.2, Damasio uses the adjective “secondary” to refer to “adult” emotions, which utilize the machinery of primary emotions in two ways:

1. Primary emotions influence the acquisition of “dispositional representations”, which are necessary for the elicitation of secondary emotions. These “acquired dispositional representations”, however, are believed to be different from the “innate dispositional representations” underlying primary emotions.

2. Secondary emotions influence bodily expressions through same mechanisms as primary emotions. Therefore, it seems reasonable to combine a primary emotion’s facial expression algorithmically with secondary emotions.

Implications for the thesis The first aspect of the connection between primary and secondary emotions is reflected in the Affect Simulation Architecture in the following way:

(1a) Secondary emotions are based on more complex data structures than primary ones. Accordingly, only some general aspects of a secondary emotion are represented in PAD space.

(1b) The appraisal of secondary emotions depends much more on the actual situational and social context than the appraisal of primary emotions. Thus, secondary emotions are more dependent on the agent's cognitive reasoning abilities.

(1c) The releasers of secondary emotions might be learned based on the history of primary emotions in connection with memories of events, agents and objects.

The second aspect mentioned above leads to the following design-decisions for the Affect Simulation Architecture:

(2a) The agent's facial expressions of primary emotions (cf. Figure 4.1) may accompany secondary emotions.

(2b) Secondary emotions also modulate the agent's simulated physis.

The "prospect-based emotions" cluster of the OCC-model of emotions (cf. Figure 2.9, p. 41) is considered here to belong to the class of secondary emotions, because their appraisal process includes the evaluation of events against previous expectations and potential future outcomes. This OCC-cluster consists of the six emotions *hope*, *fear*, *satisfaction*, *fears-confirmed*, *relief*, and *disappointment*.

Once again, as in the case of *anger* discussed above, one might wonder about the differences in the conception of *fear*. In Table 4.1 *fearful* is listed as a primary emotion in the Affect Simulation Architecture. In the OCC-model, however, the label *fear* refers to a rather complex emotion that includes the evaluation of the desirability of a possible future outcome. Both conceptions are reasonable as explained in Chapter 2 and, thus, in the Affect Simulation Architecture the label *fearful* refers to the simpler, primary emotion *fear* along the lines of LeDoux's work on fear conditioning (cf. Section 2.2.1). *Fearful* is characterized in the Affect Simulation Architecture as an emotion that is "experienced" by the agent MAX only in a state of submissiveness, i.e. only if he feels a lack of control or power. Consequently, the primary emotion *fearful* is characterized in PAD space not only by negative valence and high arousal but also by negative dominance, i.e. submissiveness (cf. Figure 4.1).

Three secondary emotions are integrated into the Affect Simulation Architecture:

I. *Hope* resulting from the prospect of a desirable event for oneself.

II. *Fears-confirmed* in case of the confirmation of an expected undesirable event.

III. *Relief* about the disconfirmation of an expected undesirable event.

A detailed description of the integration of these secondary emotions is given in Section 4.3, because the emotion dynamics simulation has to be introduced before.

4.1.2 Mood

Mood has the following properties:

1. The feedback loop of the body influences the development of mood over time.
2. Mood remains a non-conscious background feeling unless one concentrates on it.
3. Mood is a diffuse valenced state, i.e. the experiencing individual is unable to give a clear reason for a prevailing mood.
4. Emotions have a fortifying or alleviating effect on the prevailing mood of an individual.
5. Mood, in turn, influences the elicitation of emotions.
6. The duration of mood is generally longer than that of emotions.

Mood as an affect-related concept is acknowledged by many psychologists, but not investigated as thoroughly as emotions. As mentioned in Section 2.1.1 (p. 17) the idea of mood as a mental state already appears in the work of James (1884). Scherer (2005) describes mood as a diffuse affect state of low intensity but relatively long duration (cf. Table 2.1.3, p. 36). His list of examples includes the term *depressed*, which in the context of the author's Affect Simulation Architecture refers to a primary emotion that is characterized by low arousal and low dominance (cf. Figure 4.1). As will be shown during the explanation of the emotion dynamics, this conceptual difference is less problematic than one might assume. Similar to mood, Damasio vaguely describes "background emotions" as "composite expressions" resulting from the homeostatic background processes including pain and pleasure as well as drives (cf. Section 2.2.2, p. 56).

Implications for the thesis In computational implementations of affect mood is often added to OCC-based approaches to prevent unnaturally fast fluctuations of emotional states (cf. Chapter 3). In Section 3.1.5 it is concluded that mood needs not to be captured in a data structure as complex as that for emotions.

In the Affect Simulation Architecture mood is modeled as an integer value ranging from -100 to +100. Consequently, an agent can only experience its mood as an undifferentiated, valenced state. This value, however, heavily influences the emotion dynamics part of the Affect Simulation Architecture to support the believability of the agent's long-term behavior (cf. Section 4.2 for details).

4.1.3 Personality

Personality traits are captured in this thesis as follows:

1. Personality traits are rather stable dispositions and they do not change significantly during lifetime.
2. They do not contain a valence component but are, nevertheless, believed to also determine an individual's emotional responses to some respect.

3. Some parameters of an individual's emotion dynamics cover personality related aspects.

As personality traits are rather stable dispositions of an individual it is decided to only implicitly model personality-related aspects within the Affect Simulation Architecture. The usefulness of the Five Factor Model of personality (cf. McCrae & John 1992, for an introduction) is still very controversial in psychology (Bouchard & Loehlin 2001) and, thus, not taken into account in this thesis.

Implications for the thesis An individual's personality is often deduced from his or her more or less emotional reactions to potentially emotion eliciting events. A person is considered *temperamental* if an emotional reaction is rather easily evoked. If many emotional events are needed to evoke an emotional reaction, a person is considered *lethargic* or simply unemotional. These two extremes are understood as personality-related aspects of an individual. As will be shown next, some parameters of the emotions dynamics component can account for these factors of an agent's personality.

4.2 Emotion dynamics and primary emotion simulation

In this section the implementation of an emotion dynamics is described based on the idea that emotions (be they primary, secondary, or tertiary) and moods influence one another. Subsequently, the implementation of primary emotion simulation is explained, which is based on the representation of primary emotions in PAD space. The emotion dynamics component of the Affect Simulation Architecture described here resulted from the author's diploma thesis (Becker 2003) and is described similarly in (Becker et al. 2004, 2007)[2].

4.2.1 Emotions and moods and their mutual interaction

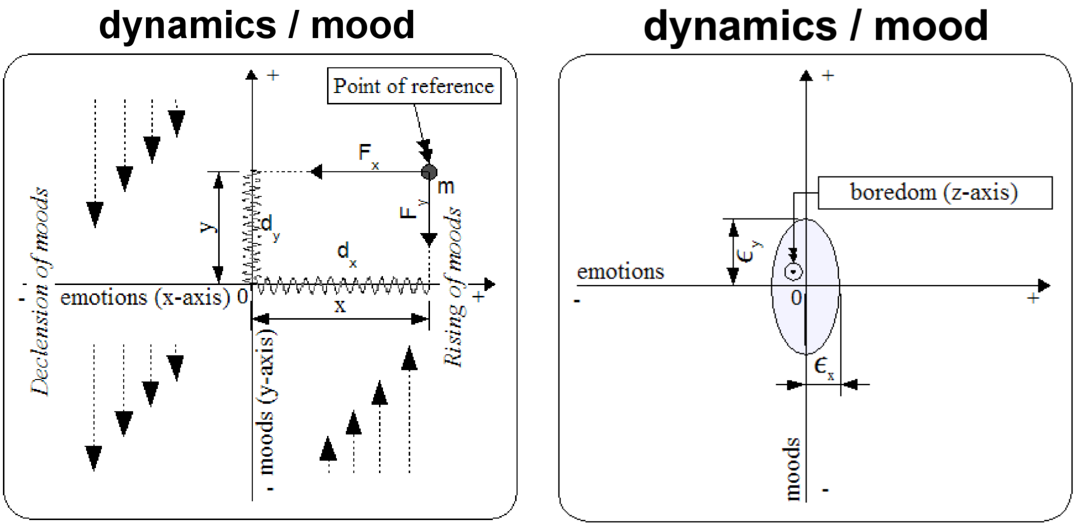
The term "emotion dynamics" refers to the mutual interaction of emotions and mood as outlined in Section 4.1.2. In general, an emotion is a short-lived phenomenon and its valence component has a fortifying or alleviating effect on the mood of an individual. A mood, in contrast, is a longer lasting, valenced state. The predisposition to experience emotions changes together with the mood, e.g. humans in a positive mood are more susceptible to positive than negative emotions, and vice versa (Neumann, Seibt & Strack 2001).

The starting point for the implementation of this emotion dynamics is an orthogonal arrangement of the respective valence components of the two affective phenomena emotion (x-axis) and mood (y-axis) as presented in Figure 4.2(a). The fortifying and alleviating effects of emotions on mood are realized by interpreting emotional valence as a gradient for changing the valence of mood at every simulation step according to the equation 4.1.

$$\frac{\Delta y}{\Delta x} = a \cdot x \tag{4.1}$$

This "upstream" and "downstream" of mood is indicated by the vertical arrows in Figure 4.2(a). The variable *a* in equation 4.1 can be interpreted as a personality-related aspect

[2] Secondary emotion simulation is partly based on the same mechanisms and will be detailed in Chapter 6.



(a) Emotions and moods and their courses over time

(b) The epsilon neighborhood for the simulation of boredom

Figure 4.2: Internals of the emotion dynamics component

modeling an agent's temperament. Smaller values of *a* result in a more *lethargic* agent and greater values of *a* lead to a more *temperamental* agent (cp. Section 4.1.3).

According to Sloman et al. (2005), in the most general sense emotions can be defined as "actual or potential perturbances" of the cognitive system, which are caused by "alarm systems" (cf. Section 3.1.1, for details). This assumption entails that a normal level of cognitive processing has to be defined first, which can then be perturbed by an emotion. This process is realized in the emotion dynamics component by explicitly simulating the course of both valences over time. In contrast to other computational models of affect, this course of emotions and mood over time is modeled rather independent from any elaborate, cognitive appraisal. Most traditional approaches (cf. Chapter 3) start with symbolic reasoning to derive appropriate emotions, calculate their intensities and then solve the problems of concurrently activated, potentially contradicting emotions and the decay of their intensities.

The implementation of emotion dynamics is based on the assumption that an organism's natural, homeostatic state is characterized by emotional balance, which accompanies an agent's normal level of cognitive processing. Therefore, two independent spiral springs are simulated, one for each axis, which create two reset forces F_x and F_y whenever the point of reference is displaced from the origin, i.e. whenever one or both valences do not equal zero (cf. Figure 4.2(a))[3].

The exerted forces are proportional to the value of the corresponding valences x and y just as if the simulated spiral springs were anchored in the origin and attached to the point of reference independently. The mass-spring model was chosen based on the heuristics that it better mimics the time course of emotions than linear and exponential functions. This assumption is supported by Reisenzein (1994), who showed that in most cases the intensity of emotions in the two dimensional Pleasure-Arousal theory is not decreasing linearly but more according to a sinus function.

[3] For further details of the implementation see (Becker 2003, p. 64ff).

By adjusting the two spring constants d_x and d_y and the simulated inertial mass *m* of the point of reference, the dynamics of both concepts can be biased intuitively. These parameters can also be construed as an aspect of an agent's personality trait.

The concept of boredom

In addition to the emotion dynamics described above, a concept of boredom is added to the dynamic component as a third, orthogonal z-axis. Assuming that the absence of stimuli is responsible for the emergence of boredom (as proposed by Mikulas & Vodanovich (1993)), the degree of boredom starts to increase linearly over time if the point of reference lies within an epsilon neighborhood of absolute zero (as given by ϵ_x and ϵ_y, cf. Figure 4.2(b)). Outside of this neighborhood the value of boredom is reset to zero by default. The co-domain of the boredom parameter is given by the interval [-1, 0], so the agent is most bored if the value of negative one is reached. The linear increase of boredom is described by Equation 4.2.

$$z(t+1) = z(t) - b \tag{4.2}$$

The parameter *b* is another aspect of the agent's personality trait. The greater the value of *b* the more easily an agent is bored in the absence of emotionally arousing stimuli.

4.2.2 Simulation of primary emotions

The outlined emotion dynamics component is so far independent from any concrete representation of emotions in PAD space as introduced in Section 4.1.1. With an update rate of Δt = 25 Hz this component provides the valences of emotion and mood together with the degree of boredom.

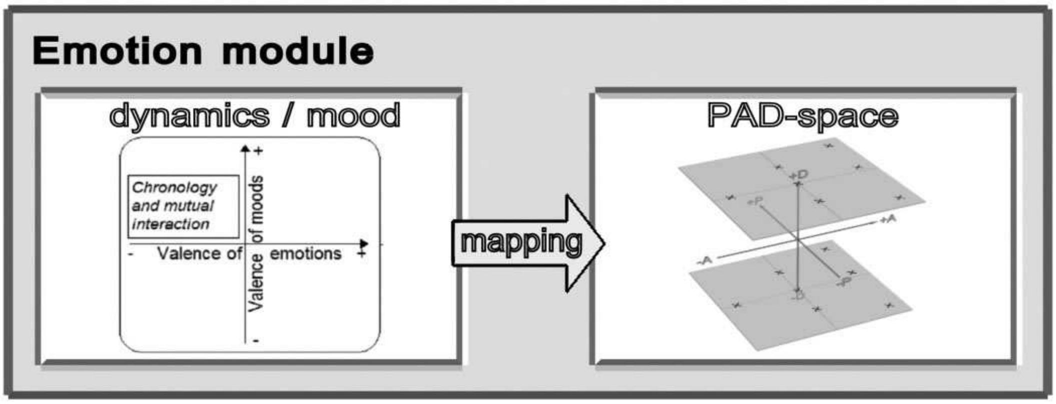


Figure 4.3: The emotion module consists of two components: The dynamics/mood component of emotion dynamics (cf. Section 4.2) and the PAD space (cf. Section 4.1.1) for the calculation of an emotion's awareness likelihood

In order to derive a primary emotion's awareness likelihood from the continuously changing values of emotion dynamics they are mapped into PAD space (cf. Figure 4.3).

Mapping into PAD space

The dynamic component provides the following triple at any time step t:

$$D(t) = (x_t, y_t, z_t), \quad with \quad x_t = [-1,1], y_t = [-1,1], z_t = [-1,0] \tag{4.3}$$

The variable x_t denotes the emotional valence, the variable y_t stands for the actual valence of mood, and z_t represents the degree of boredom. Given this triple, the mapping into PAD space for the calculation of an emotion's awareness likelihood is implemented according to the function $PAD(x_t, y_t, z_t)$ as shown in Equation (4.4). This mapping results in a triple consisting of the functions $p(x_t, y_t)$ for the calculation of *Pleasure*, $a(x_t, z_t)$ for *Arousal* and $d(t)$ for *Dominance*.

$$\begin{aligned} PAD(x_t, y_t, z_t) &= (p(x_t, y_t), a(x_t, z_t), d(t)), with \\ p(x_t, y_t) &= \frac{1}{2} \cdot (x_t + y_t) \; and \; a(x_t, z_t) = |x_t| + z_t \end{aligned} \tag{4.4}$$

Pleasure is assumed to be the overall valence information in PAD space and therefore calculated as the standardized sum of both the actual emotional valence as represented by x_t and the valence of mood as given by y_t. This way, the agent feels a maximum of joy when his emotion as well as his mood is most positive and a maximum of reluctance in the contrary case. The agent's arousal ranges from “sleepiness” to a maximum of “mental awareness” and “physiological exertion”. As it is assumed that any kind of emotion is characterized by high arousal, only the absolute value of emotional valence is considered in the function $a(x_t, z_t)$. The addition of the (negatively signed) value of boredom reflects its relation to the mental state of inactivity. The independent parameter of dominance (or, in the other extreme, submissiveness) cannot be derived from the dynamic component of the emotion module itself. As explained in Section 2.1.2, this parameter describes the agent's “feelings” of control and influence over events versus “feelings” of being controlled and influenced by external circumstances (see also the conclusion of Section 2.1.2, p. 31).

By introducing this parameter it is possible to distinguish between anger and fear as well as between sadness and annoyance. Angriness and annoyance come along with the feeling of control over the situation whereas fear and sadness are characterized by a feeling of being controlled by external circumstances. It is in principle not possible to derive such information from the dynamic component. The BDI interpreter within the cognitive module of Max, however, is capable of controlling the state of dominance. In Chapter 6 a heuristics to control this state of dominance within a cards game scenario is presented.

In principle, the awareness likelihood of a primary emotion *pe* increases the closer the point of reference gets to it. If the point of reference is getting closer than Φ_{pe} units to that particular emotion *pe* (see Figure 4.4), the calculation of its awareness likelihood w_{pe} is started according to Equation 4.5 until the distance d gets below Δ_{pe} units.

$$w_{pe} = 1 - \frac{d - \Delta_{pe}}{\Phi_{pe} - \Delta_{pe}}, \quad with \quad \Phi_{pe} > \Delta_{pe} \;\; \forall pe \in \{pe_1, \ldots, pe_9\} \tag{4.5}$$

The likelihood w_{pe} is set to 1, if the distance d is smaller than Δ_{pe}. In Equation 4.5, Φ_{pe} can be interpreted as the activation threshold and Δ_{pe} as the saturation threshold, which can be adjusted for every primary emotion $pe_n \in \{pe_1, \ldots, pe_9\}$ independently[4].

[4] The nine primary emotions are indexed according to Table 4.1.

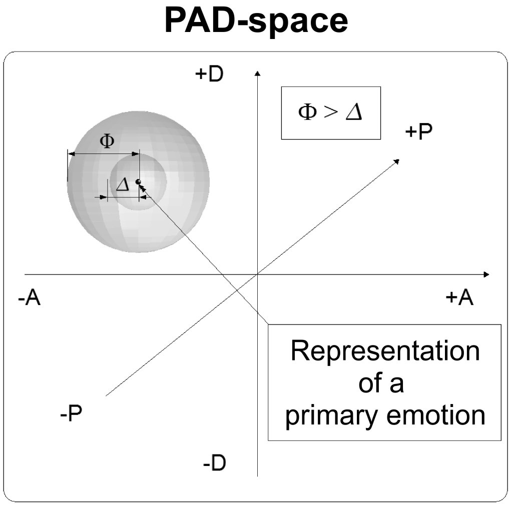


Figure 4.4: Activation threshold Φ_{pe} and saturation threshold Δ_{pe} for the awareness likelihood w_{pe} of a primary emotion *pe*

In case of primary emotions that are represented in PAD space more than once (i.e. concentrated, happy, and surprised; cf. Table 4.1) the representation with the minimum distance to the reference point is considered in Equation 4.5 for calculation of that primary emotion's awareness likelihood.

4.2.3 Summary

This simulation of emotion dynamics is quite similar to the ideas of Itoh et al. (2006), who propose a trajectory of an emotion vector through "3D Mental Space" (cf. Section 3.3.2). In their architecture the arousal component of mood is simulated as a Van der Pol oscillator, of which the harmonic oscillator simulated here is a special case. The simulation of two independent spiral springs for both valences is preferable, because the effects of adjusting the two spring constants d_x and d_y are easier to comprehend by non-experts than the effects of the many parameters of a Van der Pol oscillator. The locations of emotions in "3D Mental Space" (cf. Figure 3.11(b), p. 79) proposed by Itoh et al. (2006), however, are quite similar to those of the primary emotions presented in Figure 4.1 (p. 84).

This simulation of primary emotions proved to increase the believability of our agent MAX in different interaction scenarios as detailed in Chapter 5.

4.3 Secondary emotion simulation

The simulation of secondary emotions affords a more complex interconnection of the agent's emotion dynamics and its cognitive reasoning abilities (Becker & Wachsmuth 2006a; Becker-Asano et al. 2008). This section first describes how secondary emotions are represented in the same three-dimensional emotion space as primary ones. Then the dynamic processes are sketched that are responsible for the activation of secondary emotions in PAD space.

4.3.1 Secondary emotions in PAD space

With respect to the simulation of secondary emotions certain aspects of their connotative meaning are represented in PAD space as well, which readily enables the calculation of their awareness likelihoods. This co-representation of primary and secondary emotions in the same three-dimensional emotion space also ensures mood-congruent elicitation of both classes of emotions. Furthermore, as will be detailed in Section 4.4, a secondary emotion's valence component influences the emotion dynamics in the same way, and at the same time, as the outcomes of non-conscious appraisal of primary emotions.

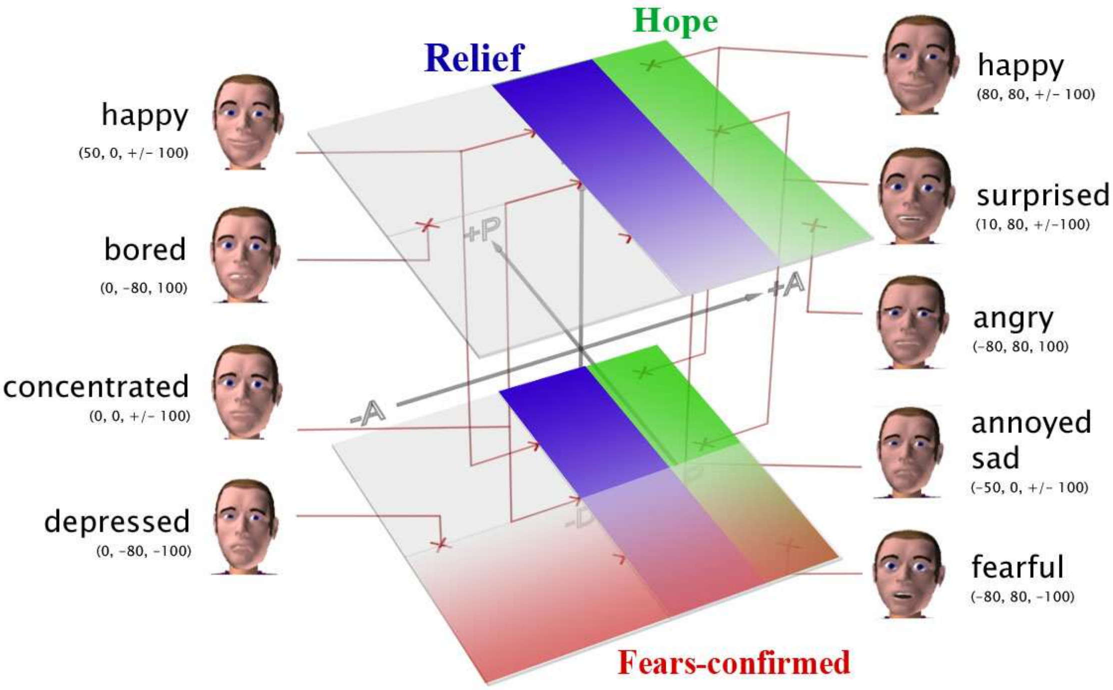


Figure 4.5: The nine primary emotions of Figure 4.1 extended by three secondary emotions as weighted areas in PAD space

As secondary emotions, however, result from conscious appraisal processes based on experiences and expectations, it is insufficient for them to be represented in terms of PAD values alone (cf. Section 4.1.1, p. 84). Furthermore, the prospect-based, secondary emotions *hope*, *fears-confirmed*, and *relief* (cf. Section 4.1.1) do not appear in the comprehensive list of (Russell & Mehrabian 1977, p. 286ff). The clusters resulting from factor analysis provided by Gehm & Scherer (1988), however, contain the two clusters “full of expectation”, to which *hope* can be ascribed, and “content”, to which “relieved” belongs. After further grouping the clusters to the four clusters “predominantly unpleasant” (A), “well-being” (B), “conflict” (C), and “happy excitement” (D) they form the “tetrahedral model of subjective emotional space” presented in Figure 2.7 (p. 31). In the final model “full of expectation” (hope) is assigned to cluster D, which features relatively neutral pleasure, high arousal, and low conflict/dominance. The “content” (relief) cluster belongs to cluster B such that it is characterized by positive pleasure, low arousal, and neutral conflict/dominance.

This analysis of the psychological background suggests to represent the secondary emotions *hope*, *fears-confirmed*, and *relief* less clear-cut in PAD space by means of *graded strucures* in contrast to *circular distributions* as in the case of primary emotions (cf. Figure 4.5). Each secondary emotion is now explained in detail.

Hope

Ortony et al. describe how *hope* results from the appraisal of a prospective event. If the potential event is considered desirable for oneself, one is likely to be "pleased about the prospect of a desirable event" (Ortony et al. 1988, p. 110). How this cognitive appraisal is exemplarily realized in the context of a card game scenario is explained in Chapter 6. The calculation of this emotion's awareness likelihood, however, is rather independent from these cognitive processes.

The previous analysis provides the rationale for modeling *hope* in the following way:

- Pleasure: The awareness likelihood of *hope* increases the more pleasurable the agent feels.

- Arousal: With respect to an agent's arousal, *hope* is more likely to be elicited the higher the agent's arousal value.

- Dominance: The awareness likelihood of *hope* is modeled to be independent of the agent's general level of dominance.

To realize this distribution of awareness likelihood in the case of hope, two areas (green) are introduced in Figure 4.5, one in the high dominance plane and the other in the low dominance plane. In Table 4.2 the exact values of the four corners of each of the two areas together with the respective intensity in each corner is given for *hope*[5].

HOPE	
Area	(PAD values), intensities
high dominance	(100, 0, 100), 0.6; (100, 100, 100), 1.0; (-100, 100, 100), 0.5; (-100, 0, 100), 0.1
low dominance	(100, 0, -100), 0.6; (100, 100, -100), 1.0; (-100, 100, -100), 0.5; (-100, 0, -100), 0.1
lifetime	10.0
standard intensity	0.0
decay function	linear
OCC-tokens	anticipation, anticipatory excitement, excitement, expectancy, hope, hopeful, looking forward to

Table 4.2: The parameters of the secondary emotion *hope* for representation in PAD space

[5] The additional parameters *lifetime*, *standard intensity*, and *decay function* in Table 4.2, Table 4.3, and Table 4.4 are explained in Section 6.2 of Chapter 6 (p. 138).

Fears-confirmed

According to (Ortony et al. 1988, p. 110), *fears-confirmed* is elicited while being "displeased about the confirmation of the prospect of an undesirable event." With respect to its representation in PAD space the similarity to the primary emotion *fearful* is taken into account and the following decisions are taken:

- Pleasure: The awareness likelihood of *fears-confirmed* increases the less pleasurable the agent feels.
- Arousal: *fears-confirmed* is considered to be independent from the agent's arousal value.
- Dominance: *fears-confirmed* can only be perceived by the agent, if he feels submissive as in the case of *fearful*.

This distribution of awareness likelihood is realized in PAD space (cf. Figure 4.5) by introducing the red area in the low dominance plane. The exact values of this area are given in Table 4.3.

FEARS-CONFIRMED	
Area	(PAD values), intensities
low dominance	(-100, 100, -100), 1.0; (0, 100, -100), 0.0; (0, -100, -100), 0.0; (-100, -100, -100), 1.0
lifetime	10.0
standard intensity	0.0
decay function	linear
OCC-tokens	fears-confirmed, worst fears realized

Table 4.3: The parameters of the secondary emotion *fears-confirmed* for representation in PAD space

Relief

The secondary emotion *relief* is described as being experienced whenever one is "pleased about the disconfirmation of the prospect of an undesirable event." (Ortony et al. 1988, p. 110) Taking the mentioned similarity with Gehm and Scherer's "content" cluster into account, the representation of *relief* in PAD space is chosen according to the following considerations:

- Pleasure: *relief* is more likely to become aware the more pleasurable the agent feels.
- Arousal: Only in case of relatively low arousal levels the agent is assumed to be aware of the emotion *relief*.
- Dominance: The awareness likelihood of *relief* is considered to be independent from the agent's state of dominance.

Accordingly, the awareness likelihood is represented in Figure 4.5 by the two shaded blue areas, one located in the high dominance plane and the other in the low dominance plane. The values for these areas together with the intensities are presented in Table 4.4.

RELIEF	
Area	(PAD values), intensities
high dominance	(100, 0, 100), 1.0; (100, 50, 100), 1.0; (-100, 50, 100), 0.2; (-100, 0, 100), 0.2
low dominance	(100, 0, -100), 1.0; (100, 50, -100), 1.0; (-100, 50, -100), 0.2; (-100, 0, -100), 0.2
lifetime	10.0
standard intensity	0.0
decay function	linear
OCC-tokens	relief

Table 4.4: The parameters of the secondary emotion *relief* for representation in PAD space

4.3.2 Secondary emotion dynamics

With representing these three secondary emotions in PAD space it is now possible to assure their mood-congruent elicitation, because the location of the point of reference (introduced in Section 4.2) is also relevant for calculating every secondary emotion's awareness likelihood. In contrast to the rather direct elicitation of primary emotions, which is so far solely based on their distance to the reference point, secondary emotions possess certain *standard intensities*, which are set to zero by default (cp. the above tables). Any secondary emotion has first to be triggered by a cognitive process, before it gains the potential to get aware to the agent. Furthermore, a secondary emotion's *lifetime* parameter (set to 10.0 by default) together with its *decay function* (set to linear by default) are used to decrease its intensity over time until the standard intensity is reached again.

In Chapter 6 the simulation of secondary emotion dynamics outlined here is explained in further detail. The next section shows how the concept of *awareness likelihood* can help to overcome long-standing difficulties that often arise in purely cognitive emotion architectures.

4.4 Connecting feelings and thoughts

The emotion module explained above needs so-called valenced emotional impulses together with the actual degree of *Dominance* as input signals to drive its internal dynamics. In return it provides descriptions of the agent's emotional state on two different levels of abstraction, first, in terms of raw but continuous *Pleasure*, *Arousal* and *Dominance* values and, second, in terms of awareness likelihoods of a number of primary and secondary emotions.

It is explained next how conscious and non-conscious appraisal lead to the elicitation of primary and secondary emotions, respectively. Especially the interplay of conscious reasoning and non-conscious reactive processes together with the emotion dynamics is outlined.

4.4.1 Conscious vs. non-conscious appraisal

In the context of emotion simulation it is helpful to divide the *Cognition module* in Figure 4.6 into two layers (based on the ideas of Ortony et al. (2005) and Damasio (1994) (cf. Sections 2.1.3 and 2.2.2)):

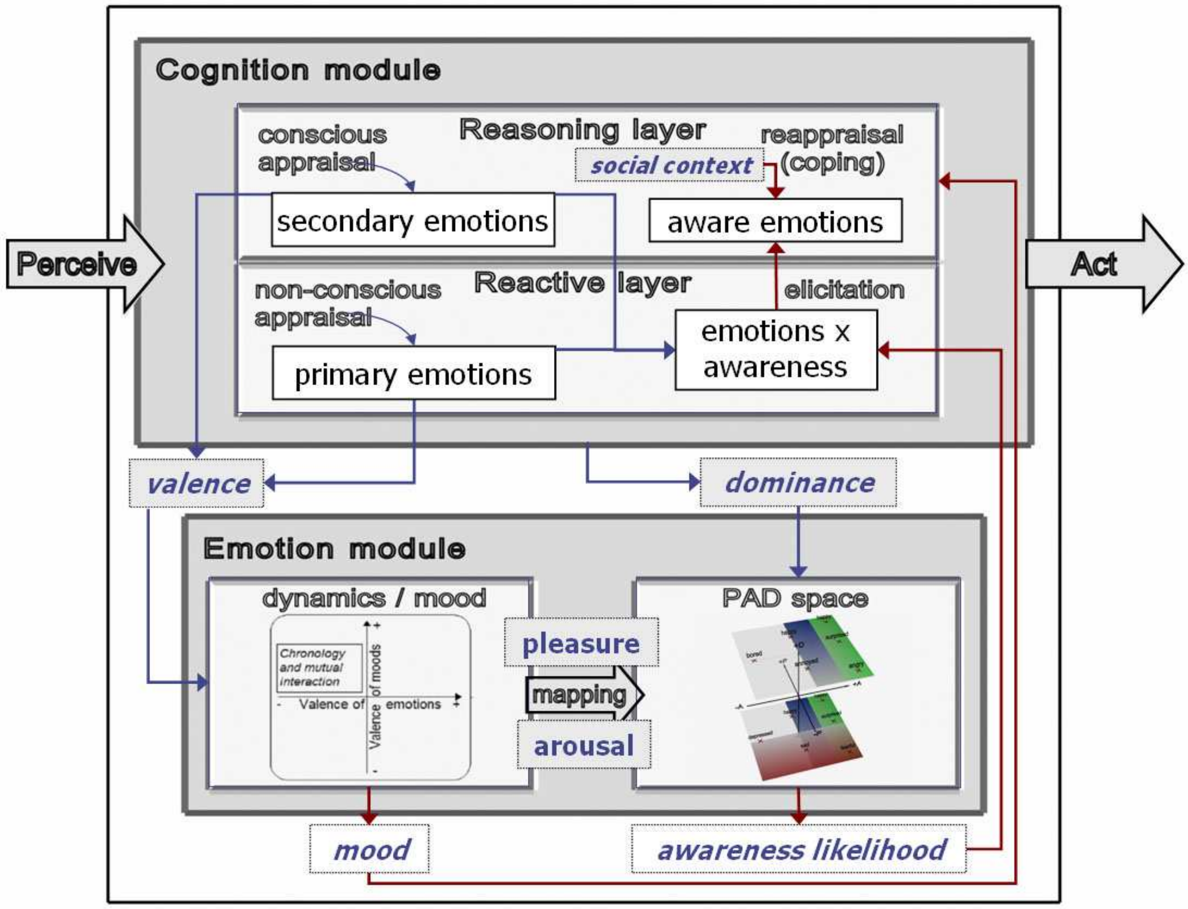


Figure 4.6: The mutual interaction of cognition and emotion. A stimulus is appraised leading to the elicitation of both primary and secondary emotions. Emotional valence and dominance values drive the emotion module to continuously update an emotion awareness likelihood, which is used to filter the elicited emotions. Finally, the aware emotions are reappraised in the social context.

1. The agent's "conscious", BDI-based deliberation resides in the *Reasoning Layer*. As the ability to reason about the eliciting factors of one's own emotional state is a mandatory prerequisite for the emergence of secondary emotions, conscious appraisal, taking place on this layer, leads to secondary emotions. This appraisal process generally includes aspects of the past and the future, making use of different kinds of memories also present on this layer[6].

2. The *Reactive Layer* can be understood as resembling onto-genetically earlier processes, which are executed on a more or less "non-conscious", automatized level. These reactive processes include simple evaluations of positive or negative valence and are implemented as hard-wired reactions to basic patterns of incoming sensor information (e.g. fast movement in the visual field). Consequently, non-conscious appraisal leads to primary emotions, which can directly give rise to "non-conscious" reactive behaviors such as approach or avoidance.

[6] An example where the (near) past and future are taken into account is demonstrated in Chapter 6.

As described in Section 4.1.1 every emotion includes a certain valence, which is either positive or negative. This hedonic (pleasurable) valence is derived from the results of appraisal on both layers and used as the main driving force in the simulation of the agent's emotion dynamics. If MAX believes, for example, that winning the game is desirable (as in the gaming scenario introduced in Chapter 5.2.2) and suddenly comes to know that the game is over without him winning, non-conscious appraisal might lead to the emergence of the primary emotion "anger" including highly negative valence[7]. However, in the Affect Simulation Architecture the resulting negative impulse only increases the likelihood of negative emotions such as *anger*. Thus, our emotional system does not follow a direct perception-action link as present in many purely rule-based, cognitive architectures.

By further representing expectations as well as memories on the reasoning layer, the secondary emotions *hope*, *fears-confirmed* and *relief* are derived. For example, if MAX analyzes the current situation and concludes that the human opponent (in the card game) could play a card which is bad for MAX insofar as it would hinder him to achieve one of his goals, the primary emotion *fearful* would result in a negative emotional impulse. When, however, the human player then plays another card on top of that card instead of playing the undesired card itself, the cognitive appraisal would result in a positive emotional impulse and the possible state of undifferentiated happiness might be accompanied or even substituted by the secondary emotion *relief* (cf. Figure 4.5).

4.4.2 Elicitation, reappraisal and coping

After the *Cognition module* has generated "proposals" of cognitively plausible emotions on the basis of conscious and non-conscious appraisal, the inherent valences of these emotions drive the dynamics/mood part of the *Emotion module*. As described in Section 4.2 the values of the dynamics subcomponent are mapped into *PAD space* for categorization and combined with the actual state of *Dominance*. This *Dominance* is provided by the *Cognition module*, which deduces its value from the actual social and situational context. The output of the *Emotion module* in terms of *awareness likelihoods* for mood-congruent emotions is then fed back to the *Cognition module*. It is combined with the initially generated "proposed emotions" to elicit a set of *aware emotions*. These *aware emotions* can be guaranteed to bear a high degree of resemblance in terms of their respective hedonic valences. Finally, reappraisal can take place to implement coping strategies such as Max leaving the display in case of high degree of anger as implemented in the museum guide scenario.

How the conscious appraisal process is computationally realized as an extension to the BDI component of the architecture is detailed in Chapter 6.

4.5 Summary

With the working definition of affective phenomena introduced in this chapter the author made the desgin commitments necessary for a computational approach to Affect Simulation. In

[7] One might object that the necessary reasoning capabilities to deduce this kind of "anger" can hardly be conceived as remaining non-conscious. In the current context, however, such a distinction is only used to separate fast reactive emotional appraisal from relatively slower, deliberative (re-)appraisal. Thus, applying symbolic reasoning to implement processes on a so-called non-conscious, reactive level is assumed noncritical.

summary, primary and secondary emotions are defined as resulting from the need to find verbal labels for complex neuropsychological processes in communication. Paying respect to this definition some aspects of their connotative meanings are represented in PAD space.

Eight verbal labels are chosen (cf. Table 2.1.3) to denote primary emotions that are correlated to five prototypical facial expressions based on Ekman (1999a). By representing these primary emotions as points in PAD space—some of them even multiple times—and by simulating an independent continuous progression of the agent's subjective feeling state, a distance metric can be applied to directly calculate the awareness likelihoods of primary emotions.

This continuous progression in PAD space is based on the idea of a mutual influence between emotion and mood. As empirically proven by psychology research a prevailing mood influences the outcome of appraisal processes such that, e.g., subjects in positive mood are less likely to get angry than subjects evaluating the same event in a negative mood.

Instead of changing the cognitive appraisal process or its outcome at the start of an emotional episode, this influence of mood on emotions is realized in the WASABI architecture independent of the realization of the appraisal process. The valence component of a cognitively elicited emotion is interpreted as an emotional impulse, which is driving the dynamic interaction of mood (understood as a longer lasting, undifferentiated, valenced state) and emotional valence. The dynamics of these two valences are updated with 25Hz resulting in a continuous progression of an agent's subjective feeling state, which is mapped into PAD space. For this mapping the standardized sum of both valences is taken as the *Pleasure* value and the absolute value of emotional valence (together with the boredom value) results in the agent's *Arousal*.

In result, the cognitive elicitation of, e.g., a positive emotion such as happiness is only increasing the likelihood that the reference point is pushed close enough to one of the four representations of happiness in PAD space. In fact, by consequently applying the idea of simple valenced, emotional impulses even reactive processes, that are not able to conduct elaborate reasoning, can influence the agent's emotional state.

As long as no high-level reasoning is integrated into the agent's cognitive architecture, however, the *Dominance* dimension in PAD space cannot be driven appropriately and the elicitation of secondary emotions is also impossible. In the WASABI architecture secondary emotions are conceived as more complex than primary emotions. For their elicitation experiences and expectations have to be derived from the situational context, because they form the basis of at least some secondary emotions.

In order to exemplify the author's approach to secondary emotion simulation three *prospect-based* OCC-emotions (*hope*, *fears-confirmed*, and *relief*) are integrated into the WASABI architecture. In contrast to primary emotions these secondary emotions are represented in PAD space as areas rather than points and there standard intensity is set to zero such that MAX can only become aware of them after they were triggered by cognitive reasoning processes.

The interplay of cognitive and non-cognitive reasoning finally drives the independent emotion module by means of emotional impulses, that are derived from every emotion's hedonic valence. The cognitive evaluation of the agent's situational context enables him to also adjust the *Dominance* value at runtime. The emotion module incessantly updates the awareness likelihoods of primary and secondary emotions and transmits them back to the cognition module, in which they are reappraised and may result in emotion-focused or situation-focused coping behavior.

5 Evaluation of primary emotions

After successful implementation of the emotion dynamics module in 2003 (cf. Becker (2003)) it was integrated into a conversational agent scenario already outlined in Section 3.2.4. In this scenario as well as in the gaming scenario subsequently presented in Section 5.2.2 the emotion simulation was limited to primary emotions. As stated in the end of the author's diploma thesis it was necessary to first evaluate the appropriateness and usefulness of the proposed emotion dynamics module, before further extension could reasonably be integrated. This chapter reports on first experiences gained in the context of the museum guide scenario. Afterwards the results of an empirical study are detailed, for which a non-conversational gaming scenario was implemented. Parts of this chapter were published in Prendinger, Becker & Ishizuka (2006) and Becker, Kopp & Wachsmuth (2007).

The following section explains how the cognitive reasoning abilities (described in Kopp et al. (2005)) are connected with the concurrently running emotion module to enhance the agent's believability in smalltalk conversation.

5.1 MAX as a museum guide

Since in the virtual museum guide scenario (cf. Section 3.2.4) the agent MAX is taking part in a smalltalk conversation, he has to follow the basic rules of social dialog as mentioned in the context of REA in Section 3.2.1. For MAX, however, it was decided to integrate an emotion simulation module enabling him to "have emotions of its own" rather devising rules as to how to influence the interlocutors emotional state as described in Bickmore & Cassell (2005). This emotion module (cf. Section 4.2) with its internal dynamics leads to a greater variety of often unpredictable yet seemingly coherent, emotion-colored responses adding to the impression that the agent has a unique personality.

5.1.1 The integration of emotions

The components of the cognitive architecture of Max essentially feed the emotion module with emotional impulses (cf. Section 4.4). These positive or negative impulses always originate from deliberative processes (interpretation and dialog manager) or as direct reactions to a positive or negative stimulus (perception).

The continuous stream of visual information provided by the video camera is first analyzed to detect the presence of skin-colored regions. A reactive, gaze following behavior is triggered whenever a new person enters the visual field of Max. At the same moment a small positive emotional impulse is sent to the emotion module such that Max's mood increases the more

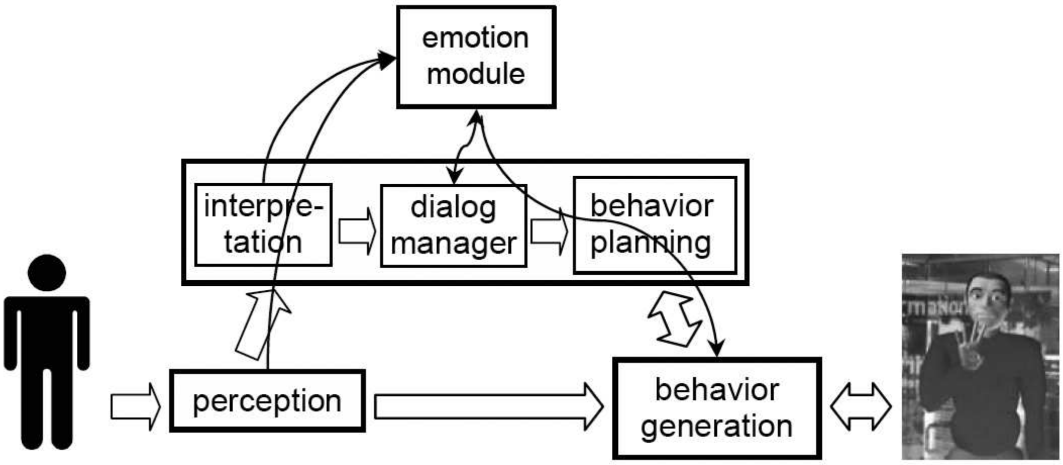


Figure 5.1: Integration of the emotion module into the conversational agent scenario (same as Figure 3.6, shown here for further discussion)

people are around. In the absence of interlocutors the emotion module is generating the emotional state of boredom (cf. Section 4.2.1) and special secondary behaviors such as leaning back and yawning are triggered. The corresponding physical exertion is modeled to have an arousing effect by automatically setting the boredom value (and, thus, also the arousal value) to zero. Concerning the *Dominance* value it was decided to let Max never feel submissive in this scenario (although a notion of initiative is accounted for by the dialogue system).

The interpretation module analyzes every input by the visitor. If, for example, the visitor's utterance is understood as a compliment, the interpretation module sends a positive impulse to the emotion dynamics module. Likewise, the achievement of a desired discourse goal, e.g., coming to know the visitor's age after having asked for it, causes the dialog manager to send a positive impulse to the emotion module.

The emotion module in turn supplies the cognitive architecture of MAX with the following data:

1. the mood valence and the degree of boredom of the dynamic component
2. the corresponding PAD triple
3. the emotion awareness likelihoods of primary emotions if any are activated

The first two kinds of information are non-cognitive information types. They are used in the behavior generation module to trigger secondary actions and to modulate involuntary facets of MAX's observable behavior, namely, the rate of his simulated breathing, the frequency of eye blink, and the pitch as well as the rate of his speech.

The third kind of information is mainly used within the dialog manager at the cognitive level of MAX's architecture. In general, deliberative reasoning is realized by a BDI interpreter that operates on the agent's beliefs, on desires representing persistent goals and a library of plans, each having preconditions, context conditions, an effect and a utility function to formulate intentions (cf. Leßmann et al. (2006) for details and Chapter 6 for examples of plans). The interpreter continually pursues the applicable plan with the highest utility value as an intention.

The categorical output of the emotion system is incessantly asserted as belief of the agent. That way, the agent's plan selection is influenced by his current affective state, which he can also verbalize. In addition, the emotion is used as precondition and context condition of plans to choose among alternative actions or even to trigger actions when becoming “aware” of a certain emotion (by asserting an according belief). Finally, the primary emotion with the highest awareness likelihood is directly reflected in Max’s facial expressions. This facial expression is then superposed on possible conversational behaviors like smiling.

5.1.2 First experiences

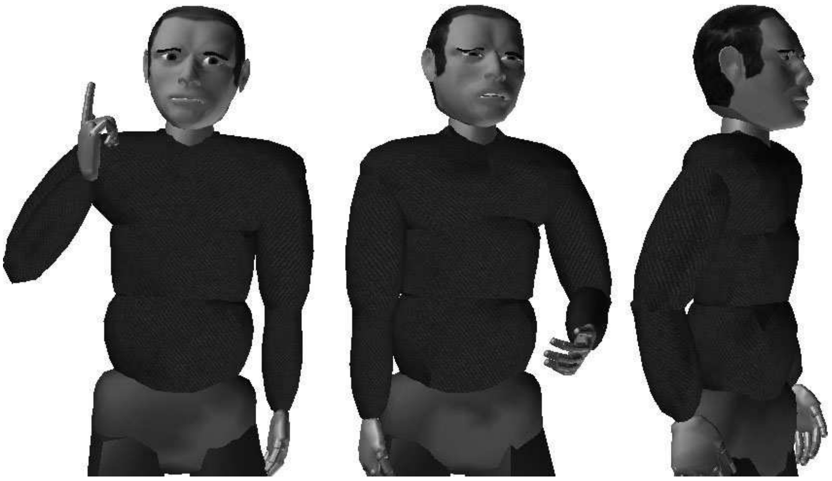


Figure 5.2: MAX is getting angry and leaves

In Figure 5.3 two parts of an example dialogue together with corresponding traces of emotions in Pleasure-Arousal-space are presented. In the beginning, MAX’s is in a neutral emotional state labeled *concentrated* until the visitor’s greeting is processed by the BDI-based Cognition module. In addition to the production of a multimodal utterance, a positive emotional impulse is sent to the emotion module. This impulse drives the internal dynamics of the “dynamics / mood” component as described in Section 4.2 and the resulting values are constantly mapped on Pleasure and Arousal values as shown in Figure 5.3(a). The first positive emotional impulse directly leads to the activation of the primary emotion *surprised* at time t_1, modulating MAX’s facial expression and synthesized voice accordingly (see Figure 4.5). During the next fourteen seconds no further impulses affect the emotion module. However, the internal dynamics leads to an increase in the agent’s mood together with a decrease of the agent’s emotional valence. Hence, the agent’s Arousal is decreasing whereas the agent’s Pleasure is increasing, such that at time t_2 the reference point in Pleasure-Arousal-space moves to *happy* and this primary emotion gets activated.

After a series of positive emotional impulses due to praising statements by the human dialogue partner, a very intense state of *happiness* is reached at time t_3. The word “pancake” is specially implemented to produce a strong negative impulse (mimicking a very rude insult), which leads to a decrease of arousal and pleasure at time t_4. Notably, the agent does not get *angry* directly but only less *happy*, because he was in a very good mood shortly before. That is, mood-congruent emotions are guaranteed as a result of the internal dynamics of the emotion module.

To the end of the conversation, MAX has become *very concentrated*–i.e. non-emotional–again, just before the visitor insults him at time t_1 (see Figure 5.3(b)) resulting in a strongly negative impulse. Within an instant MAX is *surprised* at time t_2 and only five seconds later the internal emotion dynamics let him feel *annoyed* at time t_3. The strongly negative emotional valence causes the mood to become negative within the next five seconds. Thus, when the human insults him again at time t_4, MAX gets *angry*, which he becomes aware of himself. His warning utterance is emphasized by the gesture presented on the left of Figure 5.2, before saying that he is starting “to feel unhappy”. When he is becoming very angry at time t_5, a kind of situation focused coping behavior is triggered by leaving the scene as shown in the middle and right part of Figure 5.2. As the visitor only asks MAX only an emotionally neutral question in the following fifteen seconds, MAX’s emotional state at first slowly shifts from hot to mild *anger* ending in a state of *annoyance* at time t_6. When the visitor is finally apologizing, the resulting positive impulse lets MAX feel *concentrated* again at time t_7. In effect, he re-enters the display ready to go on with the conversation.

5.1.3 Summary

A first evaluation of this collaborative scenario revealed that Max evokes natural communication strategies in the visitors, who ascribe a certain degree of sociality to the agent (cf. Kopp et al. (2005)). Until today this installation has been presented at a variety of public events ranging from science fairs (NRW60 festival, Open House of Bielefeld University, “Campus: City!” in Bielefeld City, among others) to an international AI conference (KI 2006, cf. Kopp et al. (2006)).

Furthermore, this system is permanently running in the corridor of the AI group in Bielefeld to greet visitors and engage them in a conversation. Most recently an adapted version was built for the HNF museum with special knowledge about mathematics and, more general, numbers of all kind. In all installations the emotion dynamics simulation is concurrently providing emotional states that modulate or even change MAX’s behavior in the aforementioned way.

A controlled evaluation, however, of the effects of emotion simulation in this scenario is difficult, because of the following reasons:

- A conversational agent scenario is rather uncontrollable, because the visitors are free to say anything they like to MAX.
- The duration of interactions varies a lot, because it is up to a visitors how long he or she wants to do conversation with MAX.
- A systematic way to change MAX’s emotional reactions in this open scenario is difficult to conceive.
- The necessity to let MAX verbally express himself makes it difficult to distinguish between non-verbal and verbal effects of emotion expression.

Avoiding most of these difficulties a non-verbal gaming scenario was used for an empirical study which is presented in the following section.

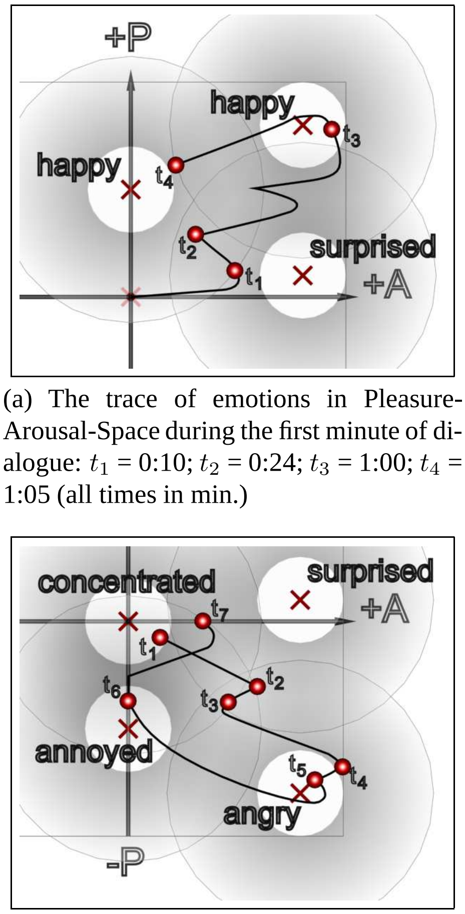
Human: Hello Max.
Max: (t_1, *surprised*) Hello. Let's chat. Use the keyboard to talk to me. What is your name?
Human: Stefan.
Max: Hello Stefan! (t_2, *happy*) Nice to meet you. I can explain something to you or we just have small talk.
Human: How are you?
Max: I'm fine at the moment and how are you?

⋮

Human: You are amazing.
Max: (t_3, *very happy*) Our conversation gets better and better.
Human: Pancake. (bad word for Max)
Max: (t_4, *happy*) Why do you say that?

⋮

Max: (t_1, *very concentrated*) Can you do so, too?
Human: You pig.
Max: (t_2, *surprised*) Have a look in the mirror (t_3, *annoyed*).
Human: Pancake.
Max: (t_4, *angry*) Better think about what you are saying. I start to feel unhappy. (t_5, *very angry*) I'm fed up for now (*leaves the display*).
Human: (*pause*) Where are you?
Max: (t_6, *very annoyed*) I will not come back until I calmed down.
Human: Sorry.
Max: (t_7, *concentrated*) After all you accept it. Now I calmed down. (*reenters the display*)
Human: Bye.
Max: Good bye, Stefan.



(a) The trace of emotions in Pleasure-Arousal-Space during the first minute of dialogue: t_1 = 0:10; t_2 = 0:24; t_3 = 1:00; t_4 = 1:05 (all times in min.)

(b) The trace of emotions until the end of the dialogue: t_1 = 1:46; t_2 = 1:50; t_3 = 1:55; t_4 = 2:02; t_5 = 2:10; t_6 = 2:25; t_7 = 2:35 (all times in min.)

Figure 5.3: A dialogue example from the conversational agent scenario, cited from Becker et al. (2007). The subfigures show the corresponding traces of Max's emotions in the Pleasure-Arousal-plane during the first (a) and second (b) part of the dialogue. Dominance is always positive and constant in this scenario.

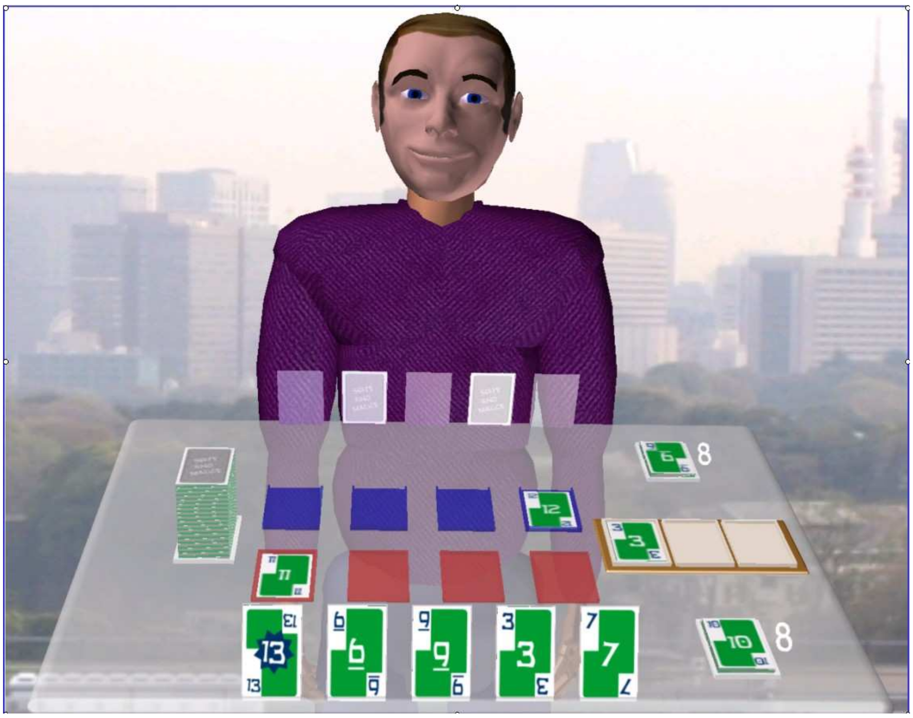


Figure 5.4: MAX playing cards against a human opponent

5.2 MAX playing Skip-Bo

Gaming scenarios that involve animated characters (such as the ones of Brave, Nass & Hutchinson (2005), Prendinger, Mori & Ishizuka (2005), or Becker et al. (2005b)) are sufficiently complex for humans to engage in meaningful social interaction and games, in general, are adequate interaction scenarios for the following reasons:

- Games help to establish social bonding between players.
- Gaming rules build a clear boundary for possible interaction moves.
- Most people like to play games and are well motivated to engage in such interactions.
- People do not expect too much interactivity of a virtual interlocutor in a gaming scenario.
- By choosing the right game the duration of interaction can easily be controlled.
- In the context of most games natural language interaction is very limited, thus avoiding problems with speech recognition and speech production.

To further investigate the appropriateness of MAX's dynamic simulation and expression of emotions through bodily gestures and facial expressions the classical card game "Skip-Bo"[1] was implemented as a face-to-face interaction scenario between a human player and MAX (see Fig. 5.4). This scenario provides MAX with a clearly defined goal (to win the game), and

[1] With friendly permission of Mattel.

he may, thus, derive a power relationship between the human player and himself in any given (game) state. This information enables MAX to distinguish between the emotion categories “fear” (low dominance) and “anger” (high dominance), and adapt his behavior accordingly (cf. Section 4.1.1 for details). By further integrating emotion recognition as outlined in Section 3.2.5 it is possible to also investigate the effect of “empathic” feedback of MAX in this scenario.

After the term “empathy” has been clarified in the following Section, the gaming scenario is introduced in Section 5.2.2. Subsequently, a short introduction to physiology-based emotion recognition (cf. Section 5.2.3) is followed by an explanation of the general setup that was used for an empirical study (cf. Section 5.2.4). This study was conducted in cooperation with Prof. Helmut Prendinger during the author’s three month visit at the National Institute of Informatics in Tokyo, Japan, as a pre-doctoral fellow of the “Japan Society for the Promotion of Science” (JSPS). Finally, the results of statistical analysis of the questionnaires as well as the bio-metrical data is presented in Section 5.2.5.

5.2.1 Conceptualizing empathy

Empathy has recently been found to be an important aspect in human-computer interaction. Paiva, Dias, Sobral & Aylett (2004) tentatively define empathy as “an observer reacting emotionally because he perceives that another is experiencing or about to experience an emotion.” (Paiva et al. 2004, p. 194) They further distinguish two different ways of mediating empathy: (1) via the situation and (2) via emotional expression. The first means to mediate empathy is conceptually close to the “Fortunes-of-others” cluster of the OCC-model of emotions (cf. Figure 2.9, p. 41) and the second manifests itself in facial mimicry as described in the end of Section 2.1.1.

To the Skip-Bo scenario presented here the second way of empathy mediation is more relevant, because empathic reactions of MAX are triggered by changes of the physiologically derived emotional states of the human player. However, the first notion of modeling empathy is also taken into account, because the emotional impulses within the Affect Simulation Architecture are adjusted with respect to the experimental condition (cf. Section 6.1).

For Brave et al. (2005) empathy is a fundamental and powerful means to manifest caring in humans. In their blackjack study they investigate the psychological impact of affective agents, which are endowed with the ability to behave empathically. In their card game scenario the agent and the human player play against a disembodied dealer. Brave et al. (2005) consider two conditions for evaluation: (1) self-oriented emotions and (2) other-oriented, empathic emotions.

In the self-oriented emotional condition the agent expresses positive emotions if winning, and negative emotions if losing, whereas in the empathic condition he expresses positive emotions if the human wins, and negative emotions if the human loses. Based on online questionnaires, Brave et al. (2005) found that subjects judge the empathic agent as more likable, trustworthy and caring as the self-emotional agent.

Arguments for a physiology-based approach

Although the results of Brave et al. (2005) offer valuable support for the utility of empathic agents, their study has some limitations. Most importantly, situations where humans interact with an agent seem to be more typical (and interesting) than those where a human and an agent assume the same view as co-players (against the dealer). Secondly, animated agents such as MAX provide a richer set of communicative modalities than photographic agents as the ones of Brave et al. (2005), and are more likely used as part of intelligent interfaces. Thirdly, questionnaires may be useful for estimating a human's opinion on dimensions such as likability, trustworthiness, or intelligence, but they fall short in assessing a human's emotional moment-to-moment experience.

Physiology-based approaches are a promising alternative to evaluating affective interactions with life-like agents since human physiology provides rich information regarding a person's emotional experience. An early study has been conducted by Ekman, Levenson & Friesen (1983) (cf. Section 2.1.1, p. 18), who investigated the effects of six basic emotions (surprise, disgust, sadness, anger, fear, and happiness; cp. Section 2.1.2) on four types of physiological signals: heart rate, skin temperature, skin resistance, and muscle tension. Their findings include a larger increase of heart rate with anger and fear than with happiness, and a higher decrease of skin resistance (leading to higher skin conductance) for fear and disgust as opposed to happiness, among other results. More recently, research in "affective computing" (cf. Chapter 3) is offering sound results on interpreting human physiological information as emotions (cf. Picard, Vyzas & Healey (2001)).

The key advantages of using human physiological response as an evaluation for human-computer interaction are the following:

- The dynamic moment-to-moment nature of a human's experience can be estimated.
- Physiological response is usually not within the conscious control of humans, preventing fake attitudes or body expressions (e.g. simulated facial expressions).
- Physiological information provides insight into the human's affective state without relying on cognitive judgements or the ability to remember past emotions.
- The recording of physiological signals does not interfere with the primary interaction task.

A potential drawback of using sensors is that they can be seen as intrusive.

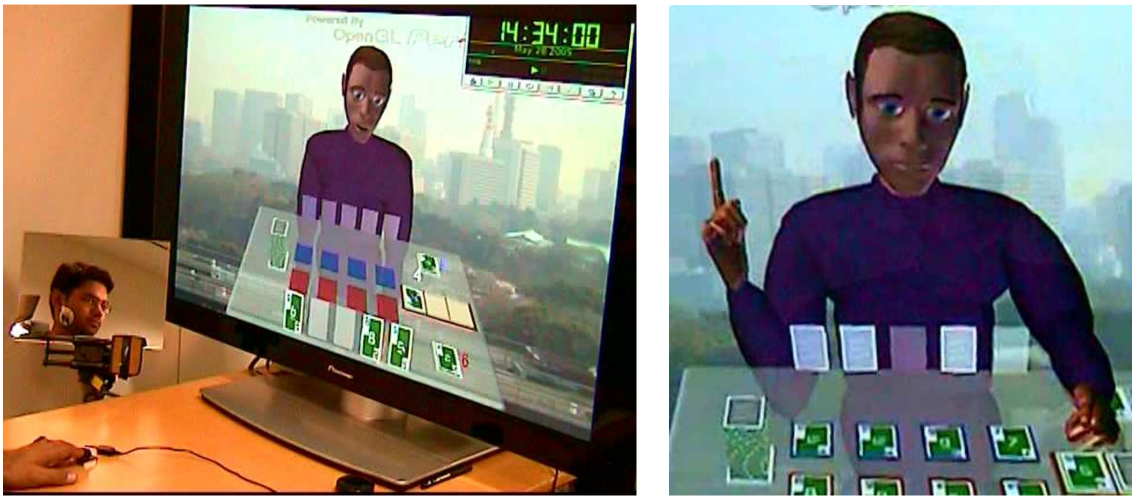
Empathy for MAX

For the study introduced here, empathy refers to MAX's response to the human's assumed emotion and covers both positive (emotional) response (e.g. sorry for the human's distress) and negative response (e.g. happy about the human's distress).

While the expression of emotion and empathy has well-known positive effects in social life, little is known about the importance of affect when expressed by a virtual human. Reflecting the experience of Berry et al. (2005) (cf. Section 3.2.2) and recasting the suggestion of (Dehn & van Mulken 2000, p. 19), the empirical study was conducted to provide a partial answer to the question "What kind of animated agent used in what kind of domain influences what kind of user's physiological state?" rather than simply "Does an animated agent improve human-computer interaction?"

5.2.2 Scenario description

Skip-Bo provides the players with conflictive goals to get rid of their eight cards on their pay-off piles on the right side of the table by playing them to the shared white center stacks (cf. Figure A.1). As on these center stacks the order of cards from one to twelve is relevant the hand and stock cards must be used strategically to achieve this overall goal. The complete instructions about how to play the game can be found in Appendix A.



(a) MAX is afraid to loose the game

(b) MAX corrects an opponent's move

Figure 5.5: Two interaction examples from the Skip-Bo gaming scenario

Speech is not seen as necessary in the card game setting and is therefore not implemented. However, MAX utters various types of "affective sounds" such as grunts and moans. Moreover, he continuously simulates breathing and eye-blinking, giving the human player the impression of interacting with a life-like agent. Visual and auditory feedback is also given whenever the human player selects or moves cards. Moreover, MAX gives visual feedback to the human player by dynamically looking at the objects (cards) selected by himself or the human for a short period of time, and then looking straight ahead again in the direction of the human player. MAX also performs a simple type of turn-taking by nodding whenever completing his move. These behaviors are intended to increase the human player's perception of interacting with an agent that is aware of its environment and the actual state of the game.

The physical objects necessary for the game are modeled as 3D objects and enriched by semantic information, so that intuitive point-and-click interaction by the human player as well as gestural interaction by MAX are easily realized (cf. Latoschik, Biermann & Wachsmuth (2005)).

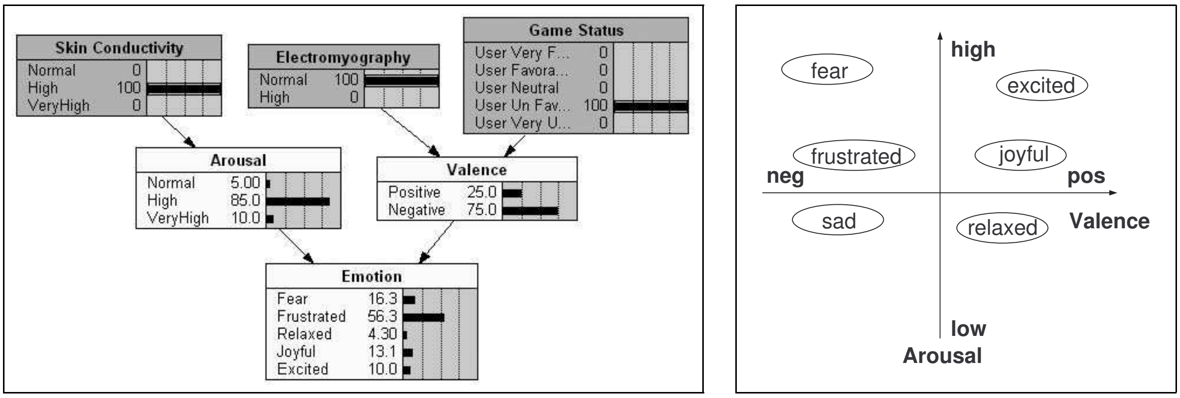
Integration of primary emotions

MAX always retains control over the game as he corrects the human player in case of a false move (see Figure 5.5(b)). MAX's emotion module is initialized to reflect this state of high *Dominance*, but whenever the human player is at least two cards ahead to win the game, this value is changed to reflect a state of *Submissiveness* (i.e. non-dominance). Consequently, when MAX is highly aroused and in a state of negative pleasure, he sometimes shows *fear* (see Figure 5.5(a)) instead of *anger*.

In Chapter 6 it is explained, how secondary emotions are integrated into this scenario and another empirical study is reported on.

5.2.3 Physiology-based emotion recognition

If MAX is supposed to respond in an empathic way, it is of paramount importance that emotions of the human player are interpreted in real-time, and input to the agent's emotion module. Based on the experiences described in Section 3.2.5 a system was used that derives the human player's emotions from skin conductance, electromyography, and situational context parameters (e.g. the game state, cf. Figure 5.6(a)).



(a) Simple Bayesian network to determine a human player's emotional state from bio-signals and game status (Becker et al. 2005, p. 40)

(b) Some named emotions in the arousal-valence space according to Lang (1995)

Figure 5.6: The decision network for emotion recognition in the Skip-Bo game and six named emotions in valence-arousal space

In short, the emotion recognition component builds on the two-dimensional (arousal, valence) model of Lang (1995) who claims that all emotions can be characterized in terms of judged valence (positive or negative) and arousal (high or low). As skin conductance increases with a person's level of overall arousal or stress, and electromyography correlates with negatively valenced emotions, named emotions can be identified in the arousal-valence space. Figure 5.6(b) shows some named emotions as coordinates in the arousal-valence space. The relation between physiological signals and arousal/valence is established in psychophysiology arguing that the activation of the autonomic nervous system (ANS) changes while emotions are elicited. The following two signals have been chosen for their high reliability[2]:

- Galvanic skin response (GSR) is an indicator of skin conductance (SC), and increases linearly with a person's level of overall arousal.
- Electromyography (EMG) measures muscle activity and has been shown to correlate with negatively valenced emotions.

[2] Other signals (electrocardiogram, EEG, respiration, temperature, pupil dilation) are applied e.g. in Picard (1997).

The current mean value is derived at runtime from a segment of five seconds. If skin conductance is 15-30% above the baseline, is assumed as “high”, for more than 30% as “very high”. If muscle activity is more than three times higher than the baseline average, it is assumed as “high”, else “normal”. Once the raw data from the sensors has been categorized, a Bayesian network (implemented with the software toolkit Netica (2003)) is used to combine the categorized information from the bio-signals and other facts about the interaction and determine the human player’s emotion based on these values. This network is shown in Figure 5.6(a). The Bayesian network is used to derive the human’s emotional state by first relating skin conductance to arousal, and EMG together with the current state of the game from the human player’s perspective to valence, and then inferring the his emotional state by applying the model of Lang (1995). The probabilities have been set in accord with the literature (whereby the concrete numbers are made up). Some examples are: “Relaxed (happiness)” is defined by the absence of autonomic signals, i.e. no arousal (relative to the baseline), and positive valence; “Joyful” is defined by increased arousal and positive valence; “Frustrated” is defined by increased arousal and negative valence.

The node “Game Status” represents situations in which the game is in one of the following states: very favorable for the human player, favorable (for the human), neutral, unfavorable, or very unfavorable. This (‘non-physiological’) node was included to the network in order to more easily hypothesize the human’s positive or negative appraisal of the current situation of the game, because EMG activity is typically seen for strong emotions only and, thus, in additional source to evaluate valence is taken into account.

In the Skip-Bo game, the behavior of MAX is modulated by both its own and the human player’s emotional state. However, in situations where a human player’s emotions are interpreted in order to determine adequate agent response, MAX’s behavior solely determined by the human player’s affective state overriding all signals from its own emotion simulation model.

5.2.4 Investigating the effects of positive and negative empathy

Since Skip-Bo is a competitive game, human players very likely perceive MAX as an opponent in this situation. Hence, the following two hypotheses underlay the study:

Hypothesis 5.1 *If MAX behaves “naturally” in that he follows his own goals and expresses associated positively or negatively valenced affective behaviors, human players will be less aroused or stressed than when MAX does not do so.*

Hypothesis 5.2 *If MAX is oriented only toward his own goals and displays associated behaviors, human players will be less aroused or stressed than when MAX does not express any emotion at all.*

The study was also motivated by the question whether the expression of negative emotions would induce negatively valenced responses in the human, or analogously, the expression of positive emotions would induce positively valenced human emotions.[3]

[3] According to (Levenson 1988, p. 19), positively valenced physiological response (a state of “relaxed happiness”) is characterized by the absence of negative response.

Subjects

Fourteen male and eighteen female subjects participated in the study and all but one subject were Japanese. Their age ranged from 22 to 55 years and the average age was 30 years. Subjects were given a monetary reward of 500 Yen for participation and they were told in advance that they would receive an extra reward of 500 Yen if they won against MAX. Subjects were randomly assigned to four experimental conditions (eight in each condition).

Design

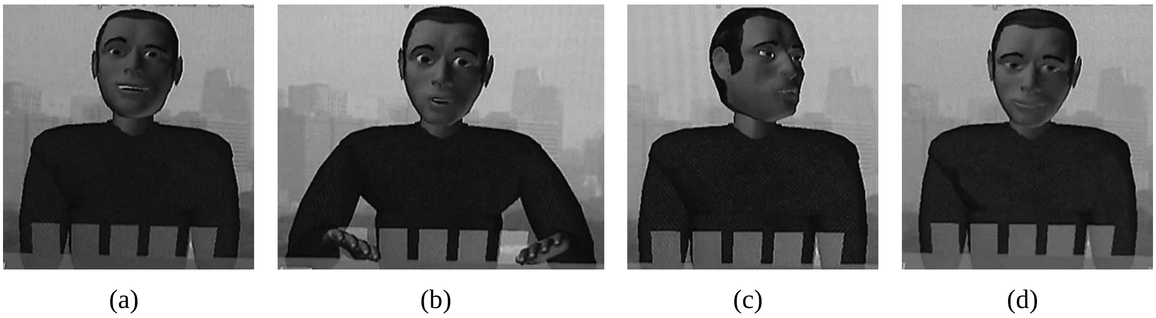


Figure 5.7: The human player is angry or frustrated and MAX reacts (a) negative empathic or (b) positive empathic or the human player is joyful or excited and MAX reacts (c) negative empathic or (d) positive empathic

In order to assess the effect of simulated emotions and empathic feedback in the context of human-computer interaction, the following four conditions within the proposed gaming scenario were designed[4]:

(i) *Non-Emotional* condition: MAX neither shows emotional behavior nor is he aware of the human player's emotional state. Nevertheless the emotion recognition data as well as the emotion simulation data are recorded for later analysis.

(ii) *Self-Centered Emotional* condition: MAX shows affective behavior that is evoked only by his own actions. The human player's actions have no effect on his own emotional state and he is not aware of the human's emotional state. MAX only appraises his own game play, and displays e.g. (facial) happiness when he is able to move cards.

(iii) *Negative Empathic* condition: MAX shows (a) self-centered emotional behavior, and (b) responds to the opponent in a "negative" way. The opponent's actions are influencing MAX's emotional state and he is aware of the opponent's affective state and responds accordingly. E.g. when the human shows frustration, MAX displays *Schadenfreude* ("joy about the user's distress", cf. Figure 5.7(a)). On the other hand, when the human player dominates the game and is recognized to be in a positively valenced state, MAX expresses ignorance by looking aside (cf. Figure 5.7(c)). Consequently, he e.g. displays distress or fear when the human performs a good move or is detected to be positively aroused.

[4] A video of the gaming interaction can be found at: http://www.becker-asano.de.

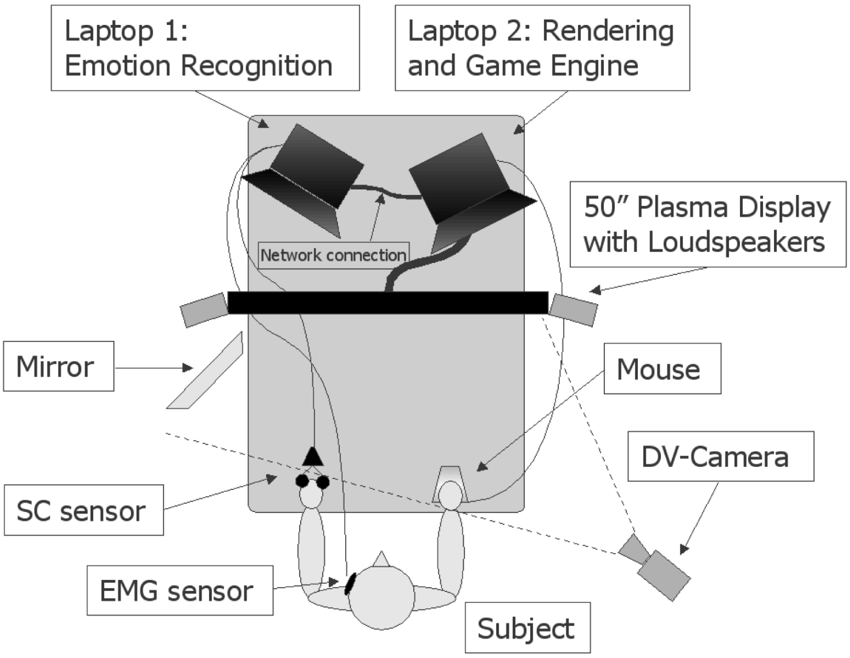


Figure 5.8: The experimental setup (Becker et al. 2005b, p. 469).

(iv) *Positive Empathic* condition: Here, MAX is (a) self-centered emotional, and (b) the opponent's actions are appraised 'positively' such that he is "happy for" the human player's game progress (cf. Figure 5.7(d)). If the human player is detected to be distressed, MAX performs a calm-down gesture (cf. Figure 5.7(b)).

These conditions should be seen as two pairs of conditions: (i) self-centered emotional (only) versus absence of self-centered emotional behavior (non-emotional behavior), and (ii) negative empathic versus positive empathic behavior. The first set will also be called non-empathic conditions, and the latter set empathic conditions. Notably, these conditions are subtly different from the conditions of Brave et al. (2005), because also negative empathy is considered here.

Procedure

Subjects received written instructions of the card game (in Japanese) with a screenshot of the starting condition before they entered the room with the experimental setup. Subjects entered the room individually and were seated in front of a 50 inch plasma display with attached loud speakers on both sides (cf. Figure 5.8). They were briefed about the experiment, in particular that they would play a competitive game. Then, subjects could play a short introductory game against a non-emotional MAX, which allowed them to get used to the mouse-based point-and-click interface, and also provided subjects the possibility to ask clarifying questions about the game. Each subject won this first game easily.

Next, the bio-metrical sensors of the ProComp Infinity encoder (cf. ThoughtTechnology (2003)) were attached to the subject and the subject was assured that these sensors were not harmful. Upon consent, a skin conductance (SC) sensor was attached to the index finger and the small finger of the non-dominant hand. The electromyography (EMG) sensor was attached to the subject's left (mirror-oriented) cheek to measure the activity of the masseter muscle (cf. Figure 5.5(a)). Then a relaxation phase of three minutes started, with MAX leaving the display and the subject being advised not to speak. In this phase a baseline was obtained for the normalization of the bio-signals, since values may greatly vary between subjects.

From now on, the experimenter remained visually separated from the subject (behind the screen) only to supervise the experiment. After the baseline was set, the agent re-entered to the screen and the subject was asked to start the game. After the game was completed, the subjects were asked to fill in a questionnaire in English presented on the screen, together with a Japanese translation on hard-copy. The questionnaire contained 25 questions that were related to the participant's subjective experience while playing the game (see Appendix B).

The whole interaction was recorded with a digital video camera positioned to the right behind the subject (cf. Figure 5.8). In order to capture both the interaction on the screen as well as the human player's facial expression, a mirror was set up to acquire in indirect image of the human players face (cf. Figure 5.5(a)). Facial expressions were not analyzed in the current study. The rationale for the mirror was to be able to identify artifacts in the EMG values due to "laughing" behaviors of subjects. Each game lasted for about ten minutes. A protocol of the progression of the game, the acquired physiological data, and the video data were recorded for later analysis.

5.2.5 Results of the empirical study

Both questionnaires and bio-metrical data were evaluated to estimate the impact of different forms of emotional agent behavior (or their absence) on human users. Our findings will be presented in the following sections.

Questionnaire results

The questionnaire contained twenty-five questions, which can be grouped into the following categories:

(i) Overall Appraisal: Seven questions about the experimental condition, including questions about whether subjects liked playing the game or how they felt during game play.

(ii) Affective Qualities of MAX: Twelve questions related to the emotionality, personality, and empathic capability of MAX.

(iii) Life-Likeness of MAX: Six questions about human players' judgements of the human-likeness of MAX' behavior and his outward appearance.

Questions were rated on a 7 point Likert scale. With respect to the first group of questions (Overall Appraisal), all but two subjects liked to play the game and everyone wanted to play it again. A nearly significant effect of the two empathic conditions in comparison with the Non-Emotional and Self-Centered Emotional conditions could be found. Subjects in the empathic conditions tended to feel less lonely (t(30) = 1:66; p = 0:053).[5]

The second group of questions (Affective Qualities of MAX)—while not providing results of statistical significance—showed that subjects had a tendency to perceive MAX as hiding

[5] The level of statistical significance is set to 0.05.

his “true feelings” in the Non-Emotional and Self-Centered Emotional conditions and showing his “true feelings” in both empathic conditions ($t(30)$ = -1:49; p = 0.073). Also, MAX was experienced as more caring about the human player’s feelings when playing a positive empathic manner then when playing in a negative empathic manner ($t(14)$ = -1.6; p = 0.068).

Concerning the third group of questions (Life-Likeness of MAX), the agent was more perceived as a “human being” when playing in an empathic way, opposed to playing in a non-emotional or self-centered emotional way ($t(30)$ = -3.42; p = 0.001). Moreover, MAX’ outward appearance was judged as more attractive when reacting empathically as compared to the Non-Emotional and Self-Centered Emotional conditions ($t(30)$ = -2.2; p = 0.018).

Results of Bio-metrical Data Analysis

This section presents the findings obtained from the analysis of bio-metrical data (SC and EMG) under the assumption of both global and local baselines.

Analysis of winning situations First it was focused on game situations where emotional reactions in the human player were likely to occur. Specifically, emotional reactions were hypothesized whenever either of the players (human or MAX) was able to play at least two pay-off pile cards in a row—which are moves toward winning the game—and eighty-seven such situations were found.

Determining the exact duration of emotions is a notoriously hard problem. In this study periods of ten seconds were analyzed, consisting of five seconds before the last pay-off card was played, and the following five seconds. For those segments the arithmetic means (averages) were calculated for both normalized SC and normalized EMG values. For each data set (each subject and each signal type), normalization was performed by applying equation 5.1.

$$x_{norm} = \frac{x_{current} - \bar{x}_{base}}{x_{max} - x_{min}} \tag{5.1}$$

In equation 5.1 the average baseline $\bar{x}_{base}$ is first subtracted from the current signal value $x_{current}$ (in the relevant segment) and the resulting value is then divided by the entire range of values applicable to each subject. This analysis assumes a global baseline as described in Section 5.2.4 (p. 111). Although named emotions could have been computed from SC and EMG data by applying the model of Lang (1995), the signal types are treated separately, in order to retain detailed physiological information about the human player.

Skin conductance

The results for skin conductance are shown in Figure 5.9.

MAX winning move. Regarding the human player’s response to MAX’s behavior when MAX performed a winning move, a significant difference between the Negative Empathic condition and the Positive Empathic condition [$t(20)$ = 2.1; $p < 0.03$] was found.[6] The non-empathic conditions were not statistically different [$t(11)$ = 2.36; p = 0.13].

Given that high skin conductance is an indicator of high arousal or stress, the human player was seemingly most aroused or stressed in the Non-Emotional condition and in the Positive

[6] All p values were obtained with two-tailed t-tests assuming unequal variances. The confidence level α was set to 0.05.

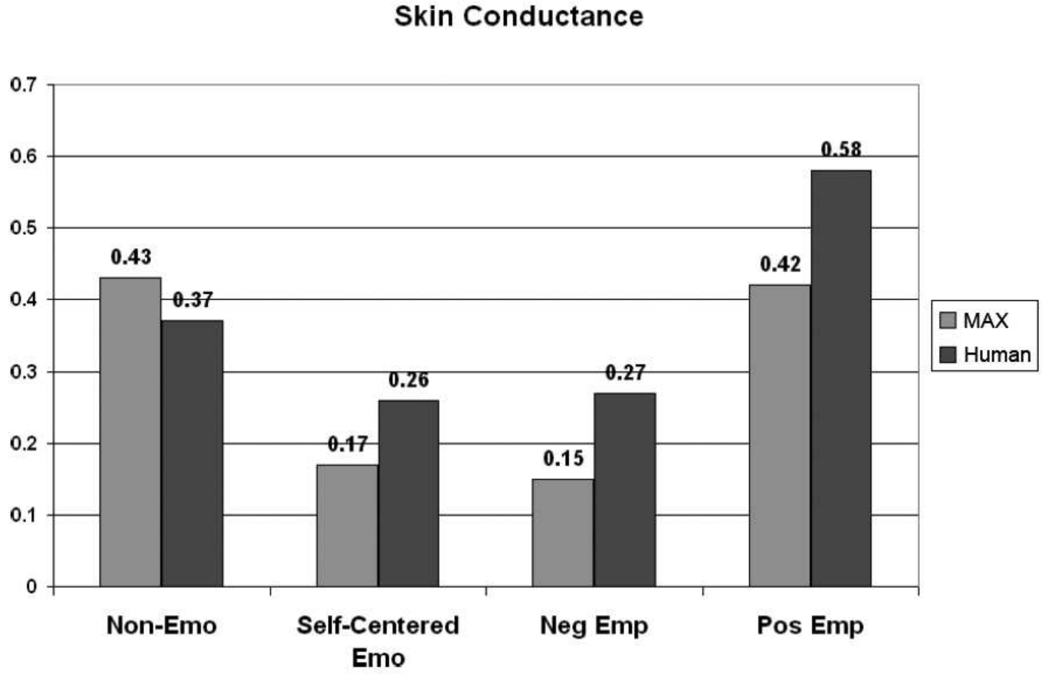


Figure 5.9: The average values of normalized skin conductance data within dedicated segments of the interaction in the four conditions: Non-Emotional (Non-Emo), Self-Centered Emotional (Self-Centered Emo), Negative Empathic (Neg Emp), and Positive Empathic (Pos Emp). MAX refers to situations where MAX performs a winning move; Human refers to winning move situations of the human player (Prendinger et al. 2006, p. 379).

Empathic condition. Although counter-intuitive at first sight, it is important to notice that in the setting of a competitive game, the lack of emotional expression or positive empathy are quite unnatural behaviors and may, thus, have induced user stress. The result supports the argumentation that inappropriate behavior (relative to an interaction task) may lead to higher stress levels.

Human winning move. A human player's physiological response to MAX when the human is in a winning situation showed a somewhat similar pattern. Notably, the agent's behavior is not independent of the human's (favorable) game moves since the physiological reaction of the human triggers emotional behavior in MAX in accord with the respective condition.

The Positive Empathic condition was experienced as significantly more arousing or stressful than the Negative Empathic condition [$t(26) = 2.07$; $p < 0.01$]. However, there was no significant difference between the Non-Emotional and Self-Centered conditions [$t(21) = 2.09$; $p = 0.46$]. The result and its explanation are related to the previous ones; e.g. in the Positive Empathic condition MAX was happy for the human player's success and gave positive feedback by displaying sorriness for the human, which constitutes an unusual behavior in a competitive game.

These findings are also consistent with the corresponding questionnaire item asking whether the agent's behavior is seen as irritating (see Appendix B). MAX was perceived as most irritating in the Non-Emotional condition, followed by the Positive Empathic condition.

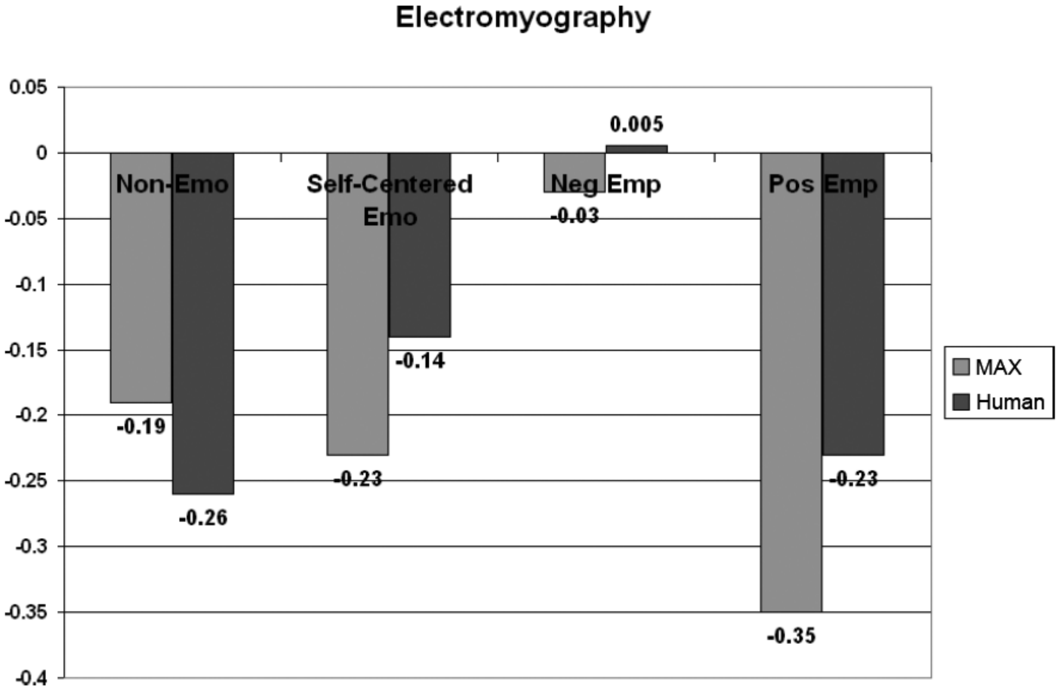


Figure 5.10: The average values of normalized electromyography data within dedicated segments of the interaction in the four conditions: Non-Emotional (Non-Emo), Self-Centered Emotional (Self-Centered Emo), Negative Empathic (Neg Emp), and Positive Empathic (Pos Emp) (Prendinger et al. 2006, p. 381).

Electromyography

Electromyography results are shown in Figure 5.10. Most values are below zero, meaning that the baseline period was experienced as negatively valenced rather than as "relaxing" in terms of muscle tension.

MAX winning move. The Negative Empathic condition differs significantly from the Positive Empathic condition [$t(20) = 2.2$; $p < 0.04$], indicating that human players seemingly "reflect" the valence of the agent's emotion expression on a physiological level. There was no statistical difference between the non-empathic conditions [$t(11) = 2.23$; $p = 0.85$].

Human winning move. Comparable to the result for MAX, the Negative Empathic condition is significantly different from the Positive Empathic condition [$t(26)=2.2$; $p < 0.04$]. Again, the non-empathic conditions were not significantly different [$t(21) = 2.07$; $p = 0.35$].

High values of electromyography are primarily an indicator of negative valence. The highest values are achieved in the Negative Empathic condition, where MAX is designed to evoke negative emotions in the human player by showing negative emotions, e.g. a mocking smile (a "happy" facial expression with an appropriate affective sound) to the human's (recognized) frustration (cf. Figure 5.7(a)). Notably, the lowest EMG values can be observed in the Positive Empathic condition where MAX performed a "calm down" gesture (slow up and down movement of hands, cf. Figure 5.7(b)) if the human player was detected to be frustrated or angry.

Interestingly, humans seemingly do not respond significantly different in both conditions when empathic agent behavior is absent (for both skin conductance and electromyography signals). This result demonstrates the discriminative effect of the type of empathic behavior displayed to the human player, and underlines the importance of an agent caring about a human's feelings in an appropriate fashion.

Analysis of situations where particular agent emotions are expressed Besides situations where either MAX or the human player is in a winning (game) situation, also situations were investigated where MAX expressed some particular emotion. This allows us to directly associate particular agent behaviors to human player's responses. This type of analysis is different from the previous one in that the experimental condition in which the emotion occurred was not taken into account.

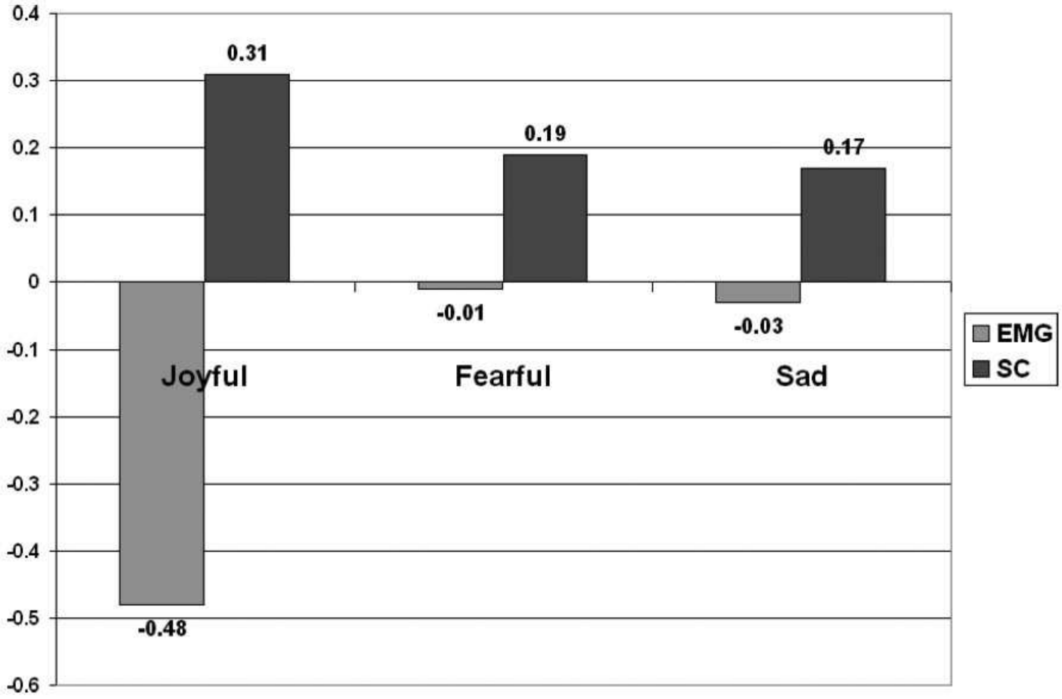


Figure 5.11: The average values of normalized skin conductance and electromyography data for the three emotions "joy", "fear", and "sadness" (Prendinger et al. 2006, p. 382)

The effect of the expression of three emotions (joyful, fearful, sad) could be analyzed (cf. Figure 5.11). Occurrences of the expression of other emotions (angry, bored, surprised) were too little for statistical analysis (fewer than six) and were hence discarded.

With regard to *skin conductance*, a between-subjects analysis of variance (ANOVA) showed that subjects were significantly more aroused or stressed when MAX expressed "joy" than when he expressed "fear" or "sadness" [$F(2, 120) = 3.9$; $p < 0.03$]. Again, it can be argued that humans seemingly consider joyful reactions of MAX as unnatural in a competitive gaming scenario and hence as arousing or stressful or, alternatively, that human's were most stressed in such situations that were favorable for MAX letting him express joy but unfavorable for the human player.

The main effect of negative emotions on *electromyography* was even more clear cut. Humans showed a significantly less negatively valenced response to joy than to fear or sadness

[$F2$, 120) = 33.78; $p < 0.0001$]. The high statistical significance of the outcome might have to be partly attributed to the nature of the EMG signal, where values typically rise beyond 300% over the baseline when the masseter muscle contracts. The result indicates that the expression of a positive emotion (joy) induces a significantly less negatively valenced response than the expression of negative emotions (fear, sadness).

5.2.6 Conclusion

The results support the supposition that an embodied agent's behavior has to be adequate with respect to the given interaction task (cf. Dehn & van Mulken (2000)). While previous similar studies only considered positive emphatic response (cf. Paiva et al. (2004), Brave et al. (2005)), Prendinger & Ishizuka (2005)), this experiment also evaluated the utility of displaying negative emotions.

Hypothesis 5.1 could be confirmed. Displaying positive affect within a competitive gaming scenario is conceived as significantly more arousing or stressful than displaying negative affect (derived from skin conductance). The same effect might appear when playing against another human.

Hypothesis 5.2 could not be confirmed by the study. If MAX does not care about the human's emotions (the non-empathic conditions), humans do not care either, i.e. their physiological response is not significantly different between the non-emotional and the self-emotional condition. Negative empathic behavior of MAX, in contrast, induces negatively valenced emotions (derived from electromyography) in humans, and analogously, positive empathic behavior is characterized by the absence of negatively valenced emotions. This finding indicates a certain reciprocity between MAX's display of affect and the human's physiological response. Moreover, MAX's expression of a positive emotion like *joy* is experienced as more arousing or stressful than the expression of a negative emotion, such as *fearful* or *sad*. On the other hand, the expression of negative emotions seemingly induces negatively valenced response, unlike the investigated positive emotion.

Overall, these results suggest that the simulation and direct expression of both positive as well as negative primary emotions has decisive effects on a human's emotional responses. If used in expedient ways, integrating primary emotions, thus, has the potential to serve significantly supportive functions in human-machine-interaction.

5.3 Summary

The emotion dynamics simulation has proven to enhance the believability of the virtual human MAX—at best so, if it also includes the simulation of negative emotions.

The de-escalation behavior in the museum scenario (i.e. MAX leaving the display as shown in Figure 5.2) implements a basic kind of situation focused coping behavior (cf. Section 2.1.3). For this kind of coping behavior MAX, however, does not reason about his level of *Control* or *Power* as suggested by Scherer (2001) (cf. Table 2.6, p. 37). In fact, in the museum guide scenario MAX's level of dominance is never changed during interaction and, thus, he gets angry instead of fearful in case of a series of insults by the human visitor. Furthermore, only this one behavior is being triggered whenever the emotional state "very angry" is activated in the emotion module and transmitted back to cognition.

Due to the emotion dynamics, on the one hand, the direct perception-action link is broken up such that the amount of insults necessary to evoke the de-escalation behavior depends not only on the actual position of the reference point in PAD space, but also on the forces accumulated over time within the dynamics simulation. On the other hand, the same emotion dynamics prevents the emotional state to "jump" from very negative to very positive ensuring a more believable succession of emotions over time. The combination of rule-based behavior generation and emotion dynamics has proven so believable that it was and still is presented at a variety of public events.

By systematically changing the emotional impulses, that are sent from the cognition module to the emotion module, positive as well as negative empathic behavior could easily be implemented in the card game scenario. Physiological measurement provided a reliable means to evaluate the effects of MAX's behavior in this non-verbal, competitive interaction scenario independent of a subject's post-hoc interpretation of the situation. Furthermore, the physiological data was used online to enhance MAX's interactive abilities letting him not only react to a human's actions, but also to his probably unconsciously changing emotional state.

The class of simulated emotions in these scenarios is so far limited to only nine primary emotions and the following criticism might be applied:

- Direct expression of emotions: Every primary emotion proposed by the emotion module directly leads to a facial expression of MAX. As MAX resembles an adult human this direct link might appear unnatural for him, because one might expect him to be able to hide his true feelings.

- The case of *surprise*: MAX often seems to be surprised just because something emotionally relevant happens in his surrounding (cp. Figure 5.3). In the gaming scenario this surprise often seems unmotivated or childish, because MAX could have expected the human's action that triggers his surprise. Even worse, he is sometimes surprised about his own gameplay, although it results from his own deliberation.

- The case of *hope*: In the museum scenario MAX sometimes responds to a difficult question with an evasive sentence such as "I hope you are not seriously asking this question."[7] Considering *hope* a secondary emotion as defined in Section 4.3.1 this statement is clearly unjustified, because secondary emotions were not yet simulated within the Affect Simulation Architecture.

To resolve some of these problems a number of extensions were conceptualized and implemented finally resulting in the WASABI architecture, which is presented in the following chapter.

[7] German: Ich hoffe Du meinst diese Frage nicht ernst.

6 Integrating secondary emotions

This chapter explains the integration of secondary emotions into the existing cognitive architecture as it was described before, resulting in a fuller account of an Affect Simulation Architecture—the WASABI architecture. The following changes and extensions to the emotion module and the cognition module are applied (see also Figure 4.6, p. 96):

1. Cognition module:

    a) The cognition module is extended to first **generate expectations** about the human's next actions and then check previous expectations against the actions currently performed by the human.

    b) By evaluating these expectations at runtime the three **secondary emotions** *hope, fears-confirmed,* and *relief* as well as the **primary emotions** *fearful* and *surprised* are **triggered** by the cognition module setting their intensities to 1.0 for a configurable amount of time.

    c) The **awareness likelihoods** of secondary emotions (being concurrently calculated in the emotion module and transmitted back to the cognition module) are subsequently processed in the cognition module and result in the **elicitation** of secondary emotions letting MAX produce appropriate verbal expressions.

2. Emotion module:

    a) **Primary emotions** are extended to also consist of base intensities that are initialized to 0.75.

    b) The base intensity of the primary emotion **surprised** is initialized to zero such that MAX can only be surprised after the cognition module appraises an event as unexpected. Furthermore, the base intensity of **fearful** is decreased to 0.25 such that MAX is less likely to get aware of this emotion, if it is not *triggered* by the cognition module.

    c) The base intensities of the three **secondary emotions** introduced in Section 4.3 are initialized to zero such that they need to be *triggered* by the cognition module before MAX might get aware of them.

The extensions to the cognition module are explained in the context of their exemplary implementation within the Skip-Bo scenario in Section 6.1, where the BDI-based reasoning capabilities of MAX are detailed. The necessary changes and enhancements applied to the emotion module are then presented in Section 6.2, including the calculation of awareness likelihoods for emotions. An overview of the information flow within the WASABI architecture is given in Section 6.3. Section 6.4 reports on an empirical study conducted to falsify the benefits of secondary emotion simulation, before a summary concludes this chapter.

6.1 The cognition module and Skip-Bo

As introduced in Section 1.2.3 the cognitive architecture of MAX comprises a cognition module (cf. Figure 6.1) and an emotion module (cf. Figure 6.3, p. 138). In order to implement the *reasoning layer* of the cognition module Leßmann (2002) argues for building upon "JAM", a "hybrid intelligent agent architecture that draws upon the theories and ideas of the Procedural Reasoning System (PRS), Structured Circuit Semantics (SCS), and Act plan interlingua." (Huber 1999, p. 236)

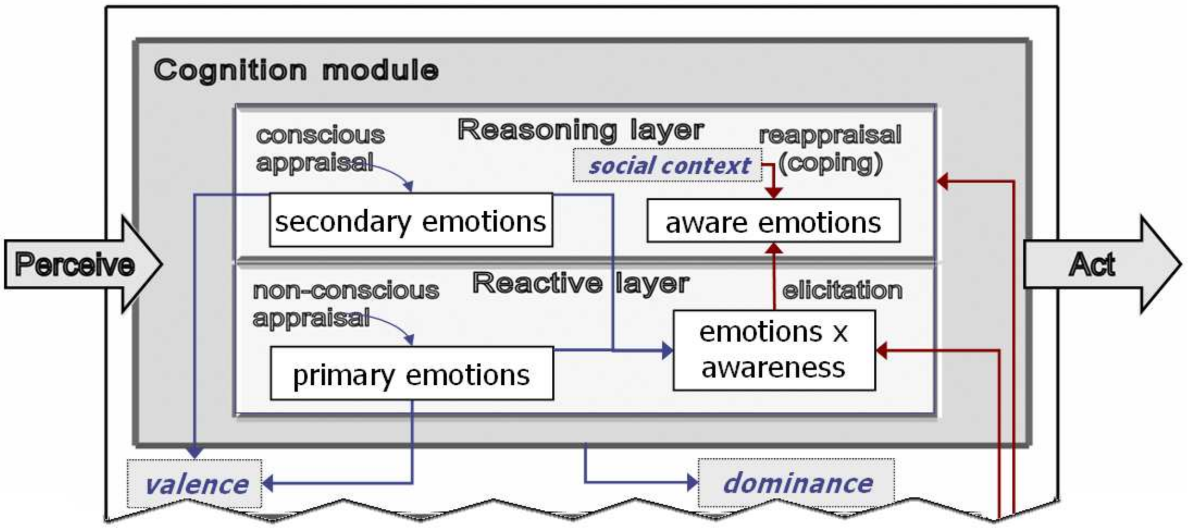


Figure 6.1: The cognition module of the WASABI architecture consisting of a reasoning layer and a reactive layer both of which feed the emotion module with input

The classical perceive-reason-act triade is extended here by the agent's ability to "short-cut" perception and action by means of the *reactive layer* (cf. Figure 6.1). In the context of the card game Skip-Bo, however, MAX has the primary goal to win the game by following its rules. MAX is given this ability by exploiting his reasoning capabilities as explained next.

6.1.1 BDI-based reasoning

According to Leßmann et al. (2004), our group "adopted the BDI architecture, for it provides provisions for modeling intentional actions in the form of plans, which help to perform complex tasks under certain conditions while being interruptible and able to recover from failure." (Leßmann et al. 2004, p. 59) Huber (1999) describes the motivation behind the development of JAM as a BDI-based architecture as follows:

> "We developed the JAM intelligent agent architecture as the next step in the evolution of pragmatic BDI-based agent architectures. [..] Starting with a BDI-theoretic 'kernel' allows us to reap the benefits of a large body of research on

> the theory and implementation of, in particular, the Procedural Reasoning System (PRS). Explicit modeling of the concepts of beliefs, goals (desires), and intentions within an agent architecture provides a number of advantages, including facilitating use of declarative representations for each of these concepts. The use of declarative representations in turn facilitates automated generation, manipulation, and even communication of these representations."

The "Principle of Rationality" (cf. Section 1.2.1, p. 5) underlies the BDI-approach in that an agent following a goal will instantiate a plan—i.e. intend this plan—based on the evaluation of his current beliefs about the world with regard to its top-level goals, i.e. its desires. If more than one plan is applicable, the one with the highest utility is chosen such that "the JAM architecture results in strictly rational agents." (Huber 1999, p. 238) How the JAM architecture was integrated into the group's software agent system and its application in the museum scenario (cf. Section 5.1) can be found in Gesellensetter (2004). *Goals* and *Plans* are the two major concepts in the JAM architecture and introduced next, because they are central to the implementation of the Skip-Bo gaming rules.

Goals and Plans

An agent performs rational top-down behavior, if it is based on the JAM architecture, by stating so-called "top-level goals". Initially one or more top-level goals are given to the agent at startup, but further goals might be instantiated during runtime either automatically as sub-level goals or dynamically by means of external communication with other processes such as the emotion module updating the awareness likelihood of emotions.

The type of each goal is either ACHIEVE, PERFORM, or MAINTAIN and every goal might be given a certain UTILITY function. An ACHIEVE goal "specifies that the agent desires to achieve a goal state" (Huber 1999, p. 239) and the agent checks whether the goal has already been accomplished before selecting an appropriate plan to reach that goal. Furthermore, if the goal has been achieved successfully a world model entry is being asserted. A PERFORM goal, in contrast, implements a semantics, which is an extension to the classical BDI architectures, because it reflects an agent's desire to perform a certain behavior even if such goal already has been achieved before. Finally, a MAINTAIN goal lets the agent maintain a certain state of the world by never removing it from the goal list automatically after achievement.

In addition to the aforementioned goal-driven behavior CONCLUDE plans let the agent perform data-driven behaviors as well. A CONCLUDE plan takes a *world model relation* as argument that is continuously checked for its logical value. As soon as the relation is considered to be *true*, the plan's PRECONDITIONS as well as CONTEXT are checked, before the plan's BODY might be executed. If the execution of a plan is successful, its optional EFFECTS section is executed; in case of failure a plan's optional FAILURE section is carried out.

6.1.2 The WBS-agents and Skip-Bo

In the AI and VR laboratory of the University of Bielefeld a multi-agent system is used as a software framework for increasingly complex software architectures. In the context of the ongoing development of MAX it enables us to encapsulate his cognitive abilities by devising specialized software agents, that communicate with each other by means of message passing over local area network.

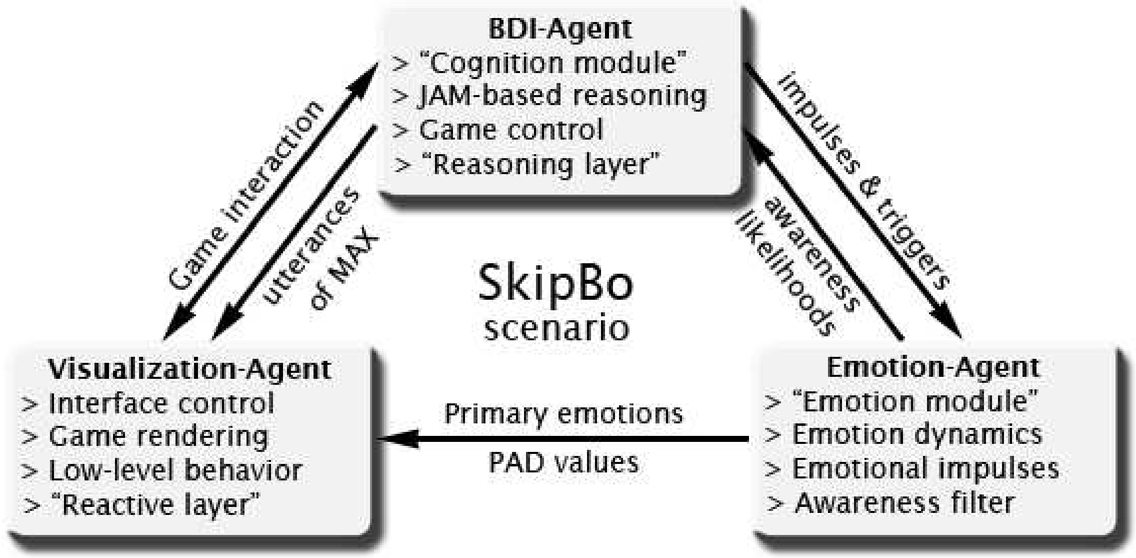


Figure 6.2: The three most important WBS-agents in the Skip-Bo scenario

Accordingly, the author's emotion dynamics simulation system was implemented as a so-called *Emotion-Agent*, which is derived from a "WBS-agent"[1] and acts in concert with a number of other agents. The emotion module—being part of the WASABI architecture—extends the functionality of the *Emotion-Agent* and, thus, is also implemented as a "WBS-agent" (cf. Figure 6.2). In the Skip-Bo scenario it receives emotional impulses from the *BDI-Agent*, which is continuously being updated with the current awareness likelihoods of primary and secondary emotions. Concurrently, the *BDI-Agent* keeps the *Visualization-Agent* updated about the actual primary emotions and PAD values.

As mentioned before, the JAM architecture is integrated into a so-called *BDI-Agent* (cf. Figure 6.2) realizing the reasoning capabilities of the "Cognition module" (cf. Figure 6.1). The *BDI-Agent* lets MAX express his secondary emotions by triggering appropriate utterances. A two-way connection between the *Visualization-Agent* and the *BDI-Agent* is established to let the cognition module take control over the human player's actions, if these do not apply to the rules, by temporarily blocking the interface. Furthermore, the *BDI-Agent* controls MAX deliberate behaviors as to let him play the game in accordance with the rules.

The visualization together with the user interface are based on a high-level, scene-graph-based framework for virtual reality applications (Latoschik et al. 2005). It provides a command line interface for rapid prototyping, which is implemented in the functional programming language Scheme. In the WBS-agent system it is represented as an agency such that its different components can be addressed directly by other agents, although these components reside in a single UNIX process. The term *Visualization-Agent* (cf. Figure 6.2) is used to refer to this complex part of the system.

[1] WBS: [W]issens[B]asierte [S]ysteme (Knowledge-based systems)

6.1.3 Implementing Skip-Bo in JAM

The Skip-Bo gaming rules (cf. Appendix A) have proven to be not too difficult for the subjects of the empirical study (cf. Section 5.2.2). For MAX being able to interact adequately in this scenario, these rules had to be transformed into a set of JAM-plans. The most important plans are presented here in pseudo code and explained with a focus on those aspects relevant to the integration of emotions. An overview of the notational conventions that apply to the following plans is given in Table 6.1.

Keyword	**Explanation**
send	The *BDI-Agent* sends a message to either the *Visualization-Agent* or the *Emotion-Agent*.
utter	The *BDI-Agent* lets MAX utter some sentence.
call	Some other plan is called within the *BDI-Agent*.

Table 6.1: Some notational conventions for the plans in pseudo code

When the *BDI-Agent* sends emotional impulses to the *Emotion-Agent* the value of the *impulse* is sometimes represented symbolically. The concrete values of these symbolic constants were different in the three emotional conditions of the first empirical study (cf. Section 5.2.2) and are presented in Table 6.2. By only adjusting these values MAX's emotional behavior was successfully changed to mimic positive versus negative empathy.

impulse	self-emo & neg-emp	pos-emp
negativeStrong	-40	-10
negativeMedium	-25	-10
negativeSmall	-10	-2
negativeTiny	-2	2
positiveTiny	2	2
positiveSmall	10	10
positiveMedium	25	25
positiveStrong	40	40

Table 6.2: The values of the emotional impulses depending on the experimental condition: self-emotional (self-emo), negative empathic (neg-emp), and positive empathic (pos-emp)

The most basic plans, which let MAX react to the human player's actions, are presented first[2], before those plans are discussed that let MAX play the game in accordance with its rules.

Reacting to human player's actions

Plan 6.1 is triggered whenever a human player selects a card by left-clicking on it with the mouse (cf. Section 5.2.2) or touching it with the hand in the CAVE (cf. Section 6.4).

[2] The initial Plan C.1 (cf. Appendix C) lets MAX welcome the human player and sends an emotional impulse of $+100$ to the Emotion-Agent resulting in a positive mood and happiness of MAX.

Plan 6.1 react to card selection

```
Conclude: REACT-TO-SELECT-CARD($who$, $cardID$, $cardValue$, $source$)
Body
    send setDominance -100
    send doAnimation lookAt
Effects
    send impulse NegativeTiny
    if more than two selects per turn AND max is empathic then
        send impulse NegativeMedium
    end if
```

Plan 6.2 react to human playing a card

```
Conclude: REACT-TO-PLAY-CARD($cardID$, $cardValue$, $source$, $target$)
Body
    if $target$ is a center then
        if $cardValue$ fits on center then
            utter one of acknowledgement sentences
            call check-for-expectations $action$                ▷ see Section 6.1.4
            send doAnimation lookAt
            if main card of max $= cardValue$ then
                send impulse negativeMedium
            end if
            if max is empathic then
                if human played from special pile then
                    send impulse negativeMedium
                else
                    send impulse negativeSmall
                end if
            end if
        else
            call handle-invalid-move                            ▷ see Plan 6.3
        end if
    end if
    if $target$ is a stock pile then
        send setDominance 100
        call game-turn-max                                      ▷ see Plan 6.4
    end if
    Effects
        if $numberOfHumanMainCards = 0$ then
            send impulse negativeStrong
            send doAnimation losing
        end if
```

First, the dominance level of MAX (represented in the emotion module by the third axis of the PAD space) is set to -100 in line 3, because MAX cannot control the human player's card selections and, thus, feels submissive[3]. The *doAnimation* message in line 4 lets MAX look at the selected card for some time before he automatically looks back at the human player.

If the plan's body was executed successfully, its EFFECTS section lets the cognition module first send a *tiny negative* impulse to the emotion module in line 6. If the human selects his third card already without having played any card during his turn and MAX is set to be empathically, an additional *medium negative* impulse is sent. This way MAX might get *fearful* (or *angry*) after a series of probably unmotivated clicks by the human player.

With **Plan 6.2** MAX reacts to a *play-card* event that was initiated by the human player either by right-clicking on a stock or center pile on the screen, or by moving a card manually to one of these piles in the CAVE. If the target is a center pile on which it does not fit (lines 4 and 18), the Plan *handle-invalid-move* is *called*. If the card, however, fits on the center, MAX first utters an acknowledgement sentence[4] before he *checks*, if he *expected* this action (cf. Section 6.1.4). Once again, he looks at the card just played by the human player (line 7).

A *medium negative* emotional impulse is sent to the emotion module, if the human player just played a card with the same value of MAX's main card. In an empathic condition an additional impulse is sent depending on the type of card being played. A human's main card results in a *medium negative* impulse whereas any other card (from the hand or stock) only results in a *small negative* impulse.

If the human played his card on one of his stock piles, no emotional impulses are sent. This action, however, automatically ends the human's turn and accordingly MAX's level of dominance is set to $+100$, because it is his turn now letting him take control over the game.

In case of success the plan finally checks, if the human player managed to get rid of all of his main cards and, thus, won the game. In that case a *strong negative* emotional impulse is sent before MAX performs an appropriate animation.

Plan 6.3 handle invalid move

```
Goal: PERFORM HANDLE-INVALID-MOVE($source$, $target$)
Body
    if first failure then
        send doAnimation rightHandUp
    end if
    send setDominance 100
    send setGameTurn max temporary
    utter one of correcting sentences
    send undo $source$ $target$
    send setGameTurn human temporary
Effects
    if max is empathic then
        send impulse NegativeStrong
    end if
```

[3] In the first empirical study discussed in Section 5.2.2 this line was not included in the plan. At that time MAX only felt submissive, if he was two cards behind with his main cards.

[4] With a probability of 72% MAX utters either "Soso!", "Aha!", "Ach so!", or "Genau!".

After MAX detected an invalid move of the human in line 4 of Plan 6.2, **Plan 6.3** is *called* in line 19 of the same plan.

If the human player did his first mistake in the game, MAX performs a *right-hand-up* animation to get his attention (cp. Figure 5.5(b), p. 107). The dominance level is set to $+100$ in line 6, because MAX is about to take temporary control of the game to undo the human player's invalid move. While doing so MAX utters an appropriate correcting sentence[5]. In case of successful execution of the plan's BODY a *strong negative* emotional impulse is sent to the emotion module in the empathic conditions.

MAX playing Skip-Bo

After the human played a hand card on one of his stock piles, the *call* of *game-turn-max* in line 24 of Plan 6.2 lets MAX take the turn. From this moment on the human player's interface (either the mouse interface (cf. Section 5.2.2) or the natural gesture interface (cf. Section 6.4)) is turned off such that he or she has to wait until MAX plays his last card on one of his stock piles.

Plan 6.4 is *called* by Plan 6.2 after the human player has played a card on one of his stock piles. The other agents are informed—by sending a message—that MAX has the turn and the level of dominance is set to $+100$, because MAX controls the game now. After MAX performed a turn-taking signal by nodding (line 5) the *take-card* plan is *called* if MAX has less than five cards on his hand[6]. After MAX filled up his hand with five cards Plan 6.5 is *called*.

Plan 6.4 set game turn

```
1: Goal: PERFORM GAME-TURN-MAX
2: Body
3:     send setGameTurn max
4:     send setDominance 100
5:     send doAnimation nodding
6:     if max needs one or more hand cards then
7:         call take-card                          ▷ see Plan C.2
8:     end if
```

Plan 6.5 is the main plan to let MAX play Skip-Bo and due to its PRECONDITION it can only be instantiated if MAX has the turn. As long as he did not play a main card he first tries to do so by calling the plan *play-main-card*. If that plan fails MAX first tries to play a hand card before (in case of another failure) he tries to play one of his stock cards. Only if all three plans are unsuccessful MAX will *exit* the loop and play one last card. The plan *generate-expectations* is *called* to let MAX think about what to expect the human player to do. A nodding animation concludes this plan to indicate that the turn is given back to the human player.

[5] With a probability of 54% MAX utters either “Also so bitte nicht!”, “Das geht so nicht!”, or “So geht das nicht!”.

[6] The two plans responsible for filling up MAX's hand are omitted here, because they have no emotional effects. For the sake of completeness, however, they are to be found in Appendix C as Plan C.2 and Plan C.3.

Plan 6.5 think about Skip-Bo

```
Goal: PERFORM THINK-SKIP-BO
Precondition: max has turn
Body
    while max did not play a main card do
        if (call play-main-card) not successful then                 ▷ see Plan 6.6
            if (call play-hand-card) not successful then             ▷ see Plan 6.7
                if (call play-stock-card) not successful then        ▷ see Plan 6.7
                    exit while
                end if
            end if
        end if
    end while
    call play-last-card                                              ▷ see Plan 6.9
    call generate-expectations                                  ▷ see Section 6.1.4
    send setGameTurn human
    send doAnimation nodding
```

Plan 6.6 lets MAX try to play his topmost main card. First the distances between his main card and each of the three actual center cards is calculated to let MAX determine that center pile with the closest distance to his main card (i.e. *closestCenter* in line 8). If the actual card on this center has a value one less than the actual main card of MAX (line 9), the plan *play-card* is *called* and a *strong positive* emotional impulse sent to the emotion module.

If the main card was not already played (line 16), MAX keeps trying to play either a hand card, a stock card, or a joker on the closest center pile until the next card to be played would be his main card. In order to also give unexperienced players a better chance to win the game, MAX does not play his main card directly but gives the turn back to the human player at this stage. If MAX cannot build the pile up by using his hand and stock cards, this plan fails in line 25. In case of success a *small positive* impulse is sent to the emotion module, whereas in case of failure the impulse is *tiny negative*.

Plan 6.7 presents two similar plans (*play-hand-card* and *play-stock-card*) in combination. The brackets indicate the places where the term "Hand" has to be replaced by the term "Stock" to change from one plan to the other.

The calculation of the minimum distance between any stock or hand card and any of the three center piles is accomplished similarly to Plan 6.6. In addition to the outer loop (line 4) an inner loop traverses all facts about hand (resp. stock) cards in line 6 as long as no card has been played. Once again, the center pile *closestCenter* with a card value closest to any possible hand (stock) card is determined. If the distance equals one, the card is played and the plan succeeds with sending a *small positive* impulse to the emotion module; otherwise, the plan fails and sends a *tiny negative* impulse.

Plan 6.8 is *called* whenever MAX wants to play a card *cardID* on some *target* pile. As all necessary checks have been applied before, the BODY of this plan only updates some belief states, sends the *playCard* command to the *Visualization-Agent*, and finally waits until the *Visualization-Agent* finished its action.

In the EFFECTS section of Plan 6.8 (starting in line 6) a *small positive* impulse is sent. If

Plan 6.6 MAX tries to play his main card

```
Goal: ACHIEVE PLAY-MAIN-CARD(mainCard)
Body
    $cardPlayed \leftarrow false$, $closestCenter \leftarrow centerOne$, $minDistance \leftarrow \infty$
    while more facts about center cards and $cardPlayed = false$ do
        $actCenter \leftarrow$ retrieveNextCenterFact
        $actDistance \leftarrow$ distance(getCard($actCenter$), $mainCard$)
        if $actDistance <= minDistance$ then
            $minDistance \leftarrow actDistance$, $closestCenter \leftarrow actCenter$
            if $actDistance = 1$ then
                call play-card $mainCard$ $actCenter$                ▷ see Plan 6.8
                $cardPlayed \leftarrow true$
                send impulse positiveStrong
            end if
        end if
    end while
    if $cardPlayed = false$ then
        while no center with value one less than mainCard do
            if any $handCard$ fits on $closestCenter$ then
                call play-card $handCard$ $closestCenter$            ▷ see Plan 6.8
            else if any $stockCard$ fits on $closestCenter$ then
                call play-card $stockCard$ $closestCenter$           ▷ see Plan 6.8
            else if max has any $joker$ then
                call play-card $joker$ $closestCenter$               ▷ see Plan 6.8
            else
                fail
            end if
        end while
    end if
Effects send impulse positiveSmall
Failure send impulse negativeTiny
```

Plan 6.7 MAX tries to play a hand card, resp. stock card

```
Goal: ACHIEVE PLAY-HAND[STOCK]-CARD
Body
    $cardPlayed \leftarrow false$, $closestCenter \leftarrow centerOne$, $minDistance \leftarrow \infty$
    while more facts about center cards and $cardPlayed = false$ do
        $actCenter \leftarrow$ retrieveNextCenterFact
        while more facts about hand [stock] cards and $cardPlayed = false$ do
            $actHand[Stock]Card \leftarrow$ retrieveNextHand[Stock]CardFact
            $actDistance \leftarrow$ distance(getCard($actCenter$), $actHand[Stock]Card$)
            if $actDistance <= minDistance$ then
                $minDistance \leftarrow actDistance$, $closestCenter \leftarrow actCenter$
                if $actDistance = 1$ then
                    call play-card $actHand[Stock]Card$ $closestCenter$        ▷ see Plan 6.8
                    $cardPlayed \leftarrow true$
                end if
            end if
        end while
    end while
    if $cardPlayed = false$ then
        fail
    end if
Effects send impulse positiveSmall
Failure send impulse negativeTiny
```

Plan 6.8 MAX plays a card

```
Goal: ACHIEVE PLAY-CARD($cardID$, $target$)
Body
    update belief states
    send playCard cardID target
    wait for feedback from Visualization-Agent
Effects
    send impulse positiveSmall
    if $numberOfOwnMainCards = 0$ then
        utter winning sentence
        send impulse positiveStrong
        send doAnimation winning
    else if $numberOfHandCards = 0$ AND playingLastCard = false then
        call take-card                                              ▷ see Plan C.2
    end if
```

Plan 6.9 MAX plays his last card

```
Goal: ACHIEVE PLAY-LAST-CARD
Body
    $playingLastCard \leftarrow true$
    $cardFound \leftarrow false$, $targetStock \leftarrow stockOne$, $maxDistance \leftarrow 0$
    while more facts about stock cards and $cardFound = false$ do
        $actStock \leftarrow$ retrieveNextStockFact
        while more facts about hand cards and $cardFound = false$ do
            $actHandCard \leftarrow$ retrieveNextHandCardFact
            $actDistance \leftarrow$ distance(getCard($actStock$), $actHandCard$)
            if $actDistance >= maxDistance$ then
                $maxDistance \leftarrow actDistance$, $targetStock \leftarrow actStock$
                $finalHandCard \leftarrow actHandCard$
                if $maxDistance = 11$ then                    ▷ building reverse pile
                    $cardFound \leftarrow true$
                else if $maxDistance = 12$ then               ▷ building pile of equal values
                    $cardFound \leftarrow true$
                end if
            end if
        end while
    end while
    if any emptyStock then                                    ▷ preferring empty piles
        $targetStock \leftarrow emptyStock$
    end if
    call play-card $finalHandCard$ $targetStock$              ▷ see Plan 6.8
Effects send impulse positiveTiny
```

MAX just played his last main card, he utters an appropriate sentence, sends a *strong positive* impulse to the emotion module and performs a *winning* animation. Otherwise, it is checked if he has an empty hand and did not intend give the turn back to the human player by playing a last card. If these conditions hold, MAX takes five new hand cards and continues his turn.

Plan 6.9 lets MAX decide where to play his last card from his hand. It is *called* by Plan 6.5 in line 13 (p. 127) when MAX is about to finish his turn. The strategy behind this plan is to first let MAX try to fill up his empty stock piles (line 21), then let him try to build "reverse stock piles" (line 13), e.g. a hand card of the value 7 on top of a stock card of the value 8. If these to options do not work, he tries to build "stock piles with equal values" (line 15), i.e. any hand card on top of any stock card of the same value. If Plan 6.9 succeeds, a *tiny positive* emotional impulse is being sent to the emotion module.

These twelve plans allow MAX to supervise the human's action in the game and to play the game in accordance with the rules of the game. Of course, these plans let MAX not play Skip-Bo like an expert. For example, MAX does not take into account the human player's stock cards or actual main card visible to him. Technically it would certainly be possible to extend the plans in such a way as to let MAX play the game more intelligently, but this is not the goal of this thesis. For empirical studies it is rather useful that MAX is not too strong an opponent, because this game only serves as testbed for intuitive human-machine interaction providing a clear set of goals.

The following Section 6.1.4 builds upon these plans in explaining how expectations are first generated and then checked against current events in order to give rise to the secondary emotions *hope*, *fears-confirmed*, and *relief* as explained subsequently in Section 6.1.5.

6.1.4 Expectations and secondary emotions

In the previous nine plans two calls are related to *expectations*: In line 14 of Plan 6.5 ("think-about-skipbo", p. 127) the plan *generate-expectations* is called and in line 6 of Plan 6.2 ("react-to-play-card", p. 124) *check-for-expectations* is invoked.

The idea behind this sequence of expectation generation and checking is as follows: After MAX played his last card and is about to give the turn back to the human player, he first thinks about which card his opponent might play next (generate-expectations). When the human player then plays a card from his or her hand or from one of his or her stock piles on one of the center piles, MAX checks for a match with his previously generated expectations (check-for-expectations).

After the human player played a card on a center that matches his expectations, MAX would, so far, not find further matching expectations before the human player finished his turn, because no other expectations are left and no further expectations generated. This unnecessary limitation is avoided by introducing Plan 6.10.

Plan 6.10 react to new card on center

```
Conclude: CENTER($centerID$, $centerValue$)
Body
    call generate-expectations
```

Plan 6.10 lets MAX generate further expectations as soon as the *Visualization-Agent* informs the *BDI-Agent* of a new card that has been played on a center pile.

It is detailed next how expectations are first generated and memorized within the BDI framework and how current events are then matched against these memorized expectations.

Generating expectations

Plan 6.11 is responsible for generating expectations, which are finally used to appraise the (secondary) prospect-based emotions *hope*, *fears-confirmed*, and *relief* (cf. Section 6.1.5). Furthermore, this plan's reasoning process is also utilized to trigger the primary emotion *fearful* resulting in a maximum intensity of that emotion in PAD space for 10 seconds. This does not mean, however, that MAX directly gets *fearful* in such a situation. It only has the effect that MAX is more likely to get aware of the emotion *fearful*, because its base intensity is temporarily raised to its maximum. The same applies to all other *trigger* messages being sent from the *BDI-Agent* to the *Emotion-Agent*. They only set the emotion's intensity temporarily to the maximum of 1.0 for the amount of seconds given as third argument[7].

In the first *while* loop of Plan 6.11 (lines 3 to 12) it is checked whether the human player can play his main card on any center pile and in case of success the plan *expect* is invoked to memorize this expectation (line 8)[8]. Notably, the third argument of *expect* (-50 in line 8) denotes the valence of the expected action, which is going to be send, if and when the human really performs that action afterwards (see Plan 6.12).

The primary emotion *fearful* is triggered in line 9, because in this situation MAX has a good reason to fear the human player's next action as it contradicts his own goal of winning the game. If the reference point in PAD space, however, does not get close enough to the primary emotion *fearful* (cf. Figure 4.5, p. 92), MAX might never get aware of this "being fearful". In humans such a mechanism might relate to someone being "cognitively" aware of some fear eliciting condition, but not getting the necessary "bodily feedback" to feel accordingly.

Starting with line 13 Plan 6.11 checks the cards on each of the human's stock piles against the three center stacks to determine, if any stock card fits on a center pile (line 19). In such a case MAX takes his own main card into account and considers two cases: He can either *hope* that the human plays his stock card or he could *fear* that the stock card is played, because this would hinder him to play his main card to that center pile (line 26).

Hope is triggered in line 24, if afterwards MAX could play his own main card, because his main card's value is two points higher than that of the considered center card (line 21). Accordingly, the expectation in this case is coupled with a positive valence of +20 in line 22. In the case of *hope*, however, an emotional impulse of +20 is sent directly to the *Emotion-Agent* to model a primitive kind of pleasant anticipation. The reason for feeling hopeful is memorized in line 25 to let MAX recall the necessary details later (cf. Plan 6.12). Once again, this process only sets the intensity of the secondary emotion *hope* to its maximum of 1.0 for ten seconds. As the base intensities of secondary emotions are initialized to zero (cp. Section 4.3.1), MAX never gets aware of them before the *BDI-Agent* found a reason to trigger them, which just has been found in line 21 of Plan 6.11.

Fearful as a primary emotion in PAD space is triggered by the *BDI-Agent*, if the stock card the human player can be expected to play is not beneficial for MAX. If the center pile holds a card with a value one less than that of MAX's actual main card, the generated expectation

[7] A detailed explanation of this process is given in Section 6.1.5

[8] The plan *expect* (Plan C.4) together with the corresponding plan *expected* (Plan C.5) are to be found in Appendix C.

Plan 6.11 let MAX generate some expectations for secondary emotions

```
Goal: PERFORM GENERATE-EXPECTATIONS
Body
    while more facts about center cards do
        $actCenter \leftarrow$ retrieveNextCenterFact
        $actDistance \leftarrow$ distance(getCard($actCenter$), $humansMainCard$)
        if $actDistance <= minDistance$ then
            if $actDistance = 1$ then
                call expect play-card $humansMainCard$ $actCenter$ -50    ▷ see Plan C.4
                send trigger fearful 10
            end if
        end if
    end while
    while more facts about center cards do
        $actCenter \leftarrow$ retrieveNextCenterFact
        while more facts about stock cards do
            $actStockCard \leftarrow$ retrieveNextStockFact
            $actDistance \leftarrow$ distance(getCard($actCenter$), $actStockCard$)
            if $actDistance <= minDistance$ then
                if $actDistance = 1$ then
                    $MAXDistance \leftarrow$ distance(getCard($actCenter$), $mainCardMAX$)
                    if $MAXDistance = 2$ then
                        call expect play-card $actStockCard$ $actCenter$ 20 ▷ see Plan C.4
                        send impulse 20
                        send trigger HOPE 10
                        memorize HOPE-REASON $action$
                    else if $MAXDistance = 1$ then
                        call expect play-card $actStockCard$ $actCenter$ -20▷ see Plan C.4
                        send trigger fearful 10
                    else
                        call expect play-card $actStockCard$ $actCenter$ -10▷ see Plan C.4
                        send trigger fearful 10
                    end if
                    exit while
                end if
            end if
        end while
    end while
```

is associated with an even more negative valence (line 27) than otherwise (line 30), because it would prevent MAX from playing his own main card.

After such critical card was found and the according expectations generated the search is aborted (line 33). If none of the human player's stock cards fits on any center pile, no expectations are generated. This quite limited ability to foresee a human player's actions in the Skip-Bo game is useful, because it has to be in accordance with MAX's ability to actively play the game. As mentioned in the end of the previous section, MAX is only able to play Skip-Bo on a beginner's level.

Checking previously generated expectations

Plan 6.12 is called whenever a human player's action is to be checked against the previously generated expectations. In the current implementation only Plan 6.2 ("react-to-play-card", p. 124) calls this plan in line 6 after the human correctly played a card on a center pile. As MAX generated expectations about possible cards to play on center piles with Plan 6.11, it is now reasonable to call plan *expected*[9] in line 3. This plan returns a tuple with the previously determined *valence* and a boolean value *answer*, which indicates if the *action* matches with an expectation or not.

If the *answer* is *true* the according *valence* is sent to the Emotion-Agent as an emotional impulse. The same *valence* is also taken into account to determine whether MAX should trigger the secondary emotion *fears-confirmed* in line 7. This activates the corresponding area in PAD space for ten seconds (cf. Figure 4.5, p. 92) making it likely for MAX to get aware of this secondary emotion. Next, the current time is memorized as an argument for the proposition *FEARS-CONFIRMED-TIME*, which is used later in Plan 6.13 again. The *reason* for this secondary emotion is only prepared in line 9 to be memorized later, if the right feedback from the *Emotion-Agent* is received. After these actions have been taken MAX forgets about this expectation.

If the human player's action was unexpected (*else* branch starting in line 12), the primary emotion *surprise* is triggered for ten seconds. As mentioned in the beginning of this chapter surprise is the only primary emotion with a base intensity of zero in the WASABI architecture[10]. Therefore, the *BDI-Agent* must trigger *surprise* before MAX has any chance to get aware of it. MAX's new ability to form expectations about the possible course of events enables him to "stay calm" in situations in which he would have been surprised before.

The rest of Plan 6.12 is concerned with detecting whether the unexpectedly played card now covers some other card on a stock or center pile, which was part of a previously expected action. First it is checked, whether the *target* of the *play-card* action is one of the stock piles, because MAX has a reason to be *relieved*, if the human player's card now covers another card that he previously *feared* the human to play (i.e. an expectation with negative *valence*). Accordingly, in line 21 the secondary emotion *relief* is triggered, before the corresponding time is memorized. Similar to the case of *fears-confirmed* before, the *reason* for being relieved is prepared to be memorized later. It might also happen that the human player plays a card on

[9] This is Plan C.5 to be found in Appendix C

[10] In the first study, reported in Section 5.2.2, the concept of base intensities was not used. The emotion dynamics system of Becker (2003) can, however, be modeled as a special case with the WASABI architecture's emotion module by, first, removing all secondary emotions and, second, setting all primary emotion's base intensities to 1.0 (see also Section 6.4.1).

Plan 6.12 let MAX check if he expected this action

```
Goal: PERFORM CHECK-FOR-EXPECTATIONS($action$)
Body
    $(valence, answer) \leftarrow$ call expected $action$ $valence$ $answer$        ▷ see Plan C.5
    if $answer = true$ then
        send impulse $valence$
        if $valence < 0$ then
            send trigger FEARS-CONFIRMED 10
            memorize FEARS-CONFIRMED-TIME getTimeInSeconds
            prepare-memorize FEARS-CONFIRMED-REASON $action$
        end if
        forget expect $action$ $valence$
    else
        send trigger surprised 10
        $target \leftarrow$ getTarget($action$)
        if target is a stock pile then
            while more facts about expectations do
                $actExpect \leftarrow$ retrieveNextExpectationFact
                $expSource \leftarrow$ getSource($actExpect$)
                if $target = expSource$ then
                    if getValence($actExpect$) $< 0$ then
                        send trigger RELIEF 10
                        memorize RELIEF-TIME getTimeInSeconds
                        prepare-memorize RELIEF-REASON $action$
                    end if
                end if
            end while
        else if target is a center pile then
            while more facts about expectations do
                $actExpect \leftarrow$ retrieveNextExpectationFact
                $expTarget \leftarrow$ getTarget($actExpect$)
                if $target = expTarget$ then
                    forget expect $action$ $valence$
                end if
            end while
        end if
    end if
```

a center pile, which was a target of some previously generated expectation (line 31). In such a case it is reasonable to let MAX forget about the previous expectation, because it cannot become true anymore.

With Plan 6.11 and Plan 6.12 the primary emotions *fearful* and *surprised* as well as the secondary emotions *hope*, *fears-confirmed*, and *relief* are triggered by the *BDI-Agent*, but this is only a necessary condition and not yet sufficient for MAX to get aware of these secondary emotions. In the following one more plan is explained, which is in turn triggered by the *Emotion-Agent* and responsible for the elicitation of secondary emotions.

6.1.5 Eliciting secondary emotions

Plan 6.13 is automatically invoked as soon as a new fact about a secondary emotion is asserted, because it is a data-driven CONCLUDE plan. Its argument *se* holds the name of the secondary emotion, which MAX is about to get aware of.

Plan 6.13 react to secondary emotion

```
 1: Conclude: REACT-TO-SECONDARY-EMOTION($se$)
 2: Body
 3:     $seReason \leftarrow$ append($se$, -REASON)                       ▷ E.g. RELIEF-REASON
 4:     if FACT mem-prelim $seReason$ $reason$ then      ▷ reason is set if seReason found
 5:         $seTime \leftarrow$ append($se$, -TIME)                           ▷ E.g. RELIEF-TIME
 6:         remember $seTime$ $time$                          ▷ time is set if seTime found
 7:         if $time + 10 >$ getTimeInSeconds then              ▷ less than 10 seconds ago
 8:             acknowledge prepared-memory
 9:             if $se =$fears-confirmed then
10:                 utter one of fears-confirmed sentences                   ▷ cf. Table 6.3
11:             end if
12:             if $se = relief$ then
13:                 utter one of relief sentences                            ▷ cf. Table 6.4
14:             end if
15:         end if
16:     end if
17:     forget $seTime$ $time$
18:     retract mem-prelim $seReason$ $reason$
19:     if FACT mem HOPE-REASON $reason$ then▷ $reason$ is set if HOPE-REASON found
20:         utter one of hope sentences                                      ▷ cf. Table 6.5
21:     end if
```

In the BODY of the plan, first, a *seReason*-proposition is constructed by appending the string "-REASON" to the secondary emotion's name (e.g. "RELIEF-REASON"). It is then checked in line 4, whether a preliminary memory is found and the *reason* for the secondary emotion can be recalled. In case of success the exact time when the secondary emotion was triggered (*seTime*) is *remembered* to check, whether it is less than ten seconds ago in which case the prepared memory is *acknowledged* making this memory permanent for MAX[11].

[11] These mechanisms of preparing and acknowledging memories were implemented in previous work by Gesellensetter (2004) for the museum guide scenario.

1) Das hatte ich schon befürchtet! (I was already afraid of that!)
2) Genau das war zu befürchten! (Exactly that was to be feared!)
3) Ganz wie befürchtet! (That's exaclty what I feared!)
4) Das musste ja so kommen! (That was to be expected!)
5) Natürlich, das hatte ich befürchtet! (Of course, I was afraid of that!)
6) Das war klar, verdammt! (Damned, that was clear!)

Table 6.3: Six sentences chosen at random by MAX to be uttered in case of *fears-confirmed*

1) Da bin ich aber erleichtert! (Now I feel relieved!)
2) Gut so, vielen Dank! (Good, thank you very much!)
3) Puh, da fällt mir ein Stein vom Herzen! (Wow, that takes a load off my mind!)
4) Ein Glück dass Du nicht die andere Karte gespielt hast! (Luckily you did not play the other card!)
5) Das ist ein Grund zur Erleichterung! (That's a reason for relief!)

Table 6.4: Five sentences chosen at random by MAX to be uttered in case of *relief*

Afterwards, the type of secondary emotion is checked and appropriate utterances are produced by MAX (line 10 and line 13). For *fears-confirmed* the sentences are given in Table 6.3 and the sentences for *relief* can be found in Table 6.4. In lines 17 and 18 MAX forgets the time *seTime* and the preliminary memory *seReason* is retracted, because the memory was either *acknowledged* in line 8 or it is outdated.

In line 24 of Plan 6.11 ("generate-expectations", p. 133) the secondary emotion *hope* is directly triggered and in line 25 the *reason* for triggering *hope* is directly memorized. One might wonder why this procedure is different from that one of triggering *fears-confirmed* and *relief*. These two emotions are only triggered later after the human player already played a card on some center pile (Plan 6.12, "check-for-expectations", p. 135).

1) Ich hoffe Du spielst die *cardValue* jetzt! (I hope you play the *cardValue* now!)
2) Hoffentlich spielst Du jetzt die Karte *cardValue*! (Hopefully, you play the card *cardValue* now!)
3) Die Karte mit der *cardValue* wär' toll! (The card with the *cardValue* would be great!)
4) Kannst Du nicht die *cardValue* spielen? (Can't you play the *cardValue*?)
5) Ich hoffe Du spielst die *cardValue* jetzt! (I hope you play the *cardValue* now!)

Table 6.5: Five sentences chosen at random by MAX to express his *hope* that the human player might play the card with value *cardValue* (sentences one and five are intentionally the same to increase the likelihood that MAX uses the verb *hope* in his utterance)

The rationale for this difference is the following: One might first *hope* (or *fear*) that some desired (or undesired) event is about to happen *in the future*, but one is only *relieved* (or feels his or her *fears-confirmed*) *after* an event has happened. This entails that not before a prospective event (i.e. an action of the human player) was confirmed MAX has any reason to memorize the eliciting condition (i.e. *reason*) of secondary, prospect-based emotions such as *relief* or *fears-confirmed*.

Consequently, for *hope* it must be checked in line 19 of Plan 6.13, whether any *reason* has been memorized as a fact before. If that reason is found, it is used to let MAX utter an adequate sentence (cf. Table 6.5). The three plans 6.11, 6.12, and 6.13 are sufficient to let MAX cognitively appraise the game situation in terms of the secondary emotions *hope*, *fears-confirmed*, and *relief*.

The two other prospect-based emotions *satisfaction* and *disappointment* (cf. Figure 2.9, p. 41) could be integrated in Plan 6.12 (“check-for-expectation”, p. 135) by including two *else* branches; the first one for *satisfaction* after line 10 and the second one for *disappointment* after line 24. With the necessary extensions to Plan 6.13 MAX could then also react to these two emotions with appropriate utterance.

So far, it was not explained how exactly the *Emotion-Agent* is triggered by the *BDI-Agent* and how it updates the awareness likelihoods of emotions. The next section clarifies this process of emotion dynamics calculation.

6.2 The emotion module

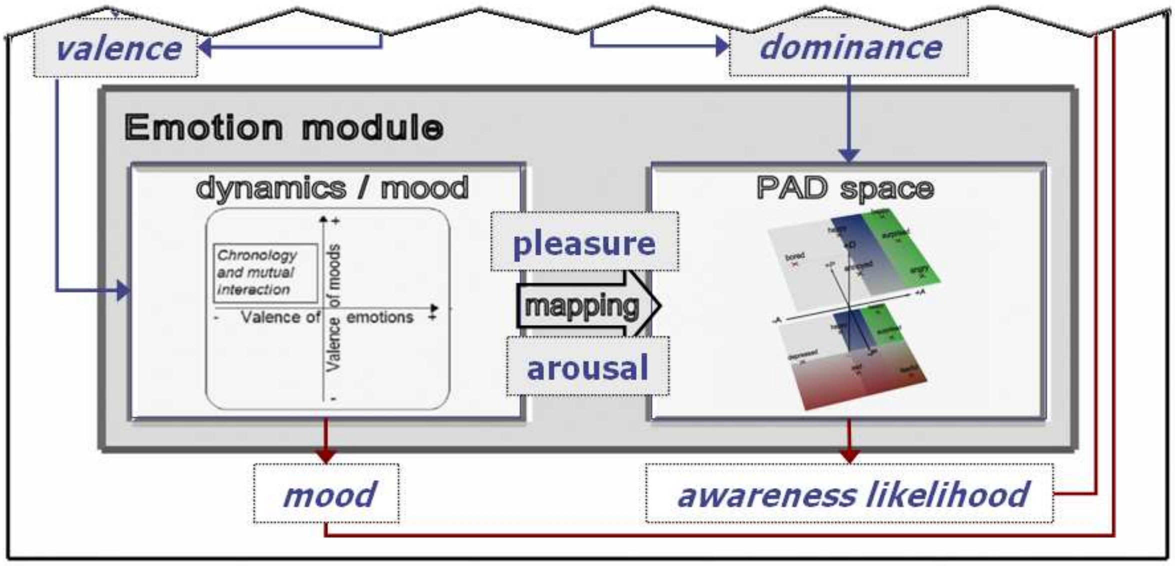


Figure 6.3: The emotion module of the WASABI architecture

This section details how an emotion (primary or secondary) is *triggered* by the *BDI-Agent*. Therefore it is necessary to recall that every emotion now consists of a base intensity of less than 1.0 (see explanation in the beginning of this Chapter).

6.2.1 Basic configuration of the emotion module

The emotion module consists of the *dynamics/mood* component and the *PAD space* component (cf. Figure 6.3). To initialize the parameters of each component, two separate configuration files are parsed by the *Emotion-Agent* at startup. The structure of these files is presented next.

Dynamics/mood

To initialize the first component of the *Emotion-Agent* the file `init.emo_dyn` (cf. Listing 6.1) is parsed[12].

Listing 6.1: Initialization file `init.emo_dyn` containing the parameters for the *dynamics/-mood* component

```
  xTens 50 # spring constant for valence of emotions
2 yTens 10 # spring constant for valence of mood
  slope 500 # factor of mutual interaction of emotion and mood
4 mass 5000 # mass of the point of reference
  xReg 1 # x-region for boredom
6 yReg 1 # y-region for boredom
  boredom 50 # time-factor for boredom
```

The parameters `xTens` and `yTens` denote the two reset forces F_x and F_y respectively (cf. Figure 4.2(a), p. 88). As justified in Section 4.2 `xTens` is greater than `yTens`, because emotional valence is considered to decrease faster than valence of mood. The fortifying and alleviating effects of emotions on mood can be tuned by changing the parameter `slope` in line 3 of Listing 6.1 (i.e. the factor a of Equation 4.1, p. 87).

By changing the parameter `mass` (line 4) the simulated inertia of the whole emotion dynamics can be adjusted, because the `mass` influences both simulated spiral springs. `xReg` and `yReg` are labeled ϵ_x and ϵ_y in Section 4.2.1 on page 89 and they define the epsilon neighborhood (cf. Figure 4.2(b)) around zero for the concept of boredom. The time it takes for boredom to reach its maximum can be adjusted by the parameter `boredom` in Listing 6.1 (i.e. the factor b in Equation 4.2, p. 89). The values of Listing 6.1 have proven to result in a reasonable emotion dynamics in all previous studies and applications.

PAD space

The *Emotion-Agent* reads the contents of Listing 6.2 at startup to initialize the primary and secondary emotions in PAD space. Each line represents all data necessary for a primary emotion according to the format:

```
<name> <P-value> <A-value> <D-value>
       <facialExpr> <saturationThresh> <activationThresh>
       <baseIntensity> [decayFunction]
```

The `P`, `A`, and `D` values as well as the parameter `<facialExpr>` of each primary emotion in Listing 6.2 are taken from Table 4.1 (cf. Section 4.1.1, p. 82).

The parameters `saturationThresh` (saturation threshold Δ_{pe}), `activationThresh` (activation threshold Φ_{pe}, cf. Figure 4.4, p. 91), and `baseIntensity` are newly introduced as explained in the previous section. The two thresholds were set to the same values for each (primary) emotion (Δ_{pe} =0.2 and Φ_{pe} =0.64) so far. For the final empirical study reported on in Section 6.4 the saturation threshold for *surprise* is now set to 0.3, i.e. Δ_9 =0.3 (lines 13 and 14 of Listing 6.2).

[12] During file parsing all characters after a # ignored until the end of line is reached.

Listing 6.2: Initialization file `initSec.emo_pad` with primary and secondary emotions

```
# PRIMARY AND SECONDARY EMOTIONS
fearful -0.8 0.8 -1 MOOD_FEARFUL 0.2 0.64 0.25 LINEAR
concentrated 0 0 -1 MOOD_CONCENTRATED 0.2 0.64 0.75 LINEAR
concentrated 0 0 1 MOOD_CONCENTRATED 0.2 0.64 0.75 LINEAR
depressed 0 -0.80 -1 MOOD_SAD 0.2 0.64 0.75 LINEAR
happy 0.8 0.8 1 MOOD_FRIENDLY 0.2 0.64 0.75 LINEAR
happy 0.5 0 1 MOOD_FRIENDLY 0.2 0.64 0.75 LINEAR
happy 0.8 0.8 -1 MOOD_FRIENDLY 0.2 0.64 0.75 LINEAR
happy 0.5 0 -1 MOOD_FRIENDLY 0.2 0.64 0.75 LINEAR
bored 0 -0.85 1 MOOD_BORED 0.2 0.64 0.75 LINEAR
annoyed -0.5 0 1 MOOD_SAD 0.2 0.64 0.75 LINEAR
sad -0.5 0 -1 MOOD_SAD 0.2 0.64 0.75 LINEAR
surprised 0.1 0.8 1 MOOD_SURPRISED 0.3 0.64 0.0 LINEAR
surprised 0.1 0.8 -1 MOOD_SURPRISED 0.3 0.64 0.0 LINEAR
angry -0.8 0.8 1 MOOD_ANGRY 0.2 0.64 0.75 LINEAR
> relief.se
> fears-confirmed.se
> hope.se
```

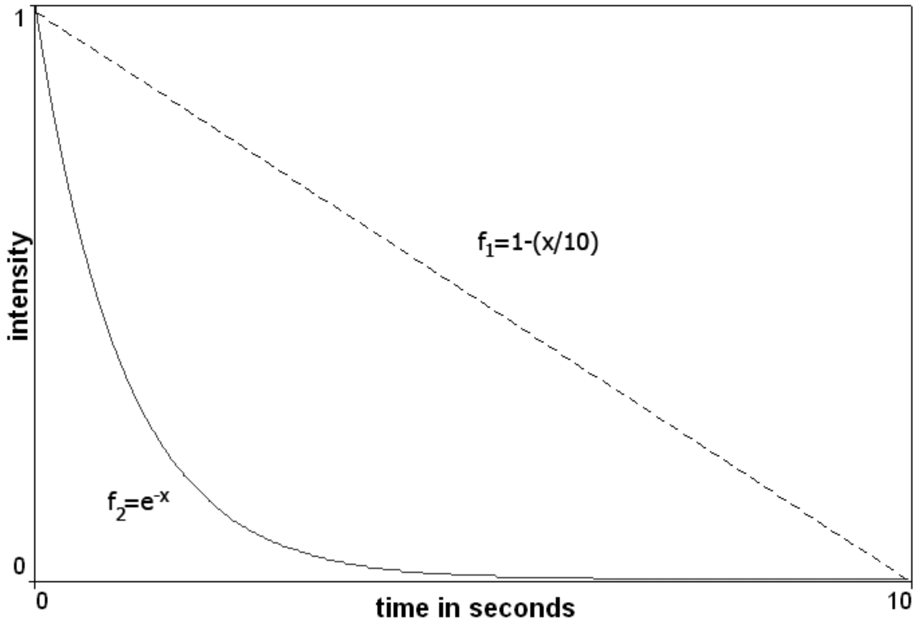


Figure 6.4: The plots of the linear decay function f_1 and the exponential decay function f_2 in case of a standard lifetime of 10 seconds

With the optional parameter `decayFunction` the type of decay function for emotion intensity can be configured according to Table 6.6. If this parameter is omitted, the decay function is set to type `NONE`. In Figure 6.4 the plots f_1 for the linear as well as f_2 for the exponential decay function are shown in case of a standard lifetime of ten seconds.

Type	**Explanation** (a primary emotion's `lifetime` is 10.0 by default)
`NONE`	The intensity is not decayed over time and reset to the emotion's `<baseIntensity>` after the emotion's `lifetime` is expired.
`LINEAR` (f_1)	The intensity decreases linearly over time until the emotion's `lifetime` is expired; then the intensity is reset to `<baseIntensity>`.
`EXPONENTIAL` (f_2)	The intensity decreases exponentially until the emotion's `lifetime` is expired; then it is reset to `<baseIntensity>`.

Table 6.6: The three possible decay functions for emotion intensities

The last three lines of Listing 6.2 start with the special character "`>`" indicating the inclusion of an external file (`*.se`) defining a secondary emotion. The initialization file for the secondary emotion *hope* is presented in Listing 6.3.

Listing 6.3: Initialization file `hope.se` for the secondary emotion *hope*

```
   polygon_begin QUAD
 2 vertex 100 0 100 0.6
   vertex 100 100 100 1.0
 4 vertex -100 100 100 0.5
   vertex -100 0 100 0.1
 6 polygon_end
   polygon_begin QUAD
 8 vertex 100 0 -100 0.6
   vertex 100 100 -100 1.0
10 vertex -100 100 -100 0.5
   vertex -100 0 -100 0.1
12 polygon_end
   decayFunction LINEAR
14 lifetime 10.0
   baseIntensity 0.0
16 type HOPE
   tokens_begin OCC
18 anticipation
   excitement
20 expectancy
   hope
22 hopeful
   looking_forward_to
24 tokens_end
```

In lines 1 to 12 of Listing 6.3 two *polygons* are defined by stating their respective *vertices*. The parameter `QUAD` after the keyword `polygon_begin` indicates the type of polygon to be realized with the subsequent list of vertices. By changing the parameter to `POINTS` every vertex is interpreted as a single point in PAD space such that a "cloud-like" distribution for a secondary emotion can be realized as well. These two possible types of polygons are explained in Table 6.7[13].

Type	Explanation
`POINTS`	Every vertex is interpreted as a point similar to primary emotions.
`QUAD`	The four vertices v_0, v_1, v_2, v_3 are interpreted as the corners of a quadrilateral (four-sided polygon). All following vertices are ignored.

Table 6.7: Supported types of polygons as parameter for keyword `polygon_begin`

The four parameters `<P-value>`, `<A-value>`, `<D-value>`, and `<baseIntensity>` have to follow after every `vertex` keyword. Accordingly, the vertices of the first (lines 1 to 6) and second (lines 7 to 12) polygon correspond to the values for the two areas *high dominance* and *low dominance* in Table 4.2 (p. 93).

The keyword `decayFunction` is used to specify the type of decay function (cf. Table 6.6) in line 13 of Listing 6.3. Together with the information about a secondary emotion's `lifetime` (in seconds, line 14) the decrease of its intensity after it has been *triggered* is specified. The `baseIntensity` of *hope* is specified in line 15 to equal 0.0. By the keyword `type` a name for the emotion is declared, which is used to identify the emotion in the graphical user interface of the *Emotion-Agent*.

For the sake of completeness a list of `tokens` can be declared in which case the parameter after `tokens_begin` (`OCC` in line 17 of Listing 6.3) denotes the type of tokens and is automatically prepended to every token that follows. Thus, the source of a concept for any secondary emotion can be specified by, e.g., stating one of `OCC`, `SCHERER`, `SLOMAN`, or `DAMASIO` here (cf. Section 4.3, p. 91). In the current C++ implementation these tokens are generated and represented as a vector of strings within a `SecondaryEmotion` object but not further used so far.

The initialization files for the secondary emotions `fears-confirmed` and `relief` have a very similar structure and are, thus, not discussed here. They can be found as Listings D.1 and D.2 in Appendix D.

6.2.2 Calculating awareness likelihoods

The calculation of a primary emotion's awareness likelihood is already described in Section 4.2.2 (p. 89), except for the influence of an emotion's intensity. In the final implementation these intensities i_{pe} are provided to enable some more cognitive control over the primary emotions *fearful* and *surprised* as described in the previous section. The final calculation of awareness likelihoods for primary emotions is given in Equation 6.1.

$$l_{pe} = w_{pe} \cdot i_{pe} \tag{6.1}$$

The result w_{pe} of Equation 4.5 is simply multiplied with the primary emotion's current intensity i_{pe} resulting in the final emotion awareness likelihood l_{pe}.

[13] This implementation is similar to the syntax of polygon definitions in OpenGL with `gl_begin()` and `gl_end()`, but note the difference in the usage of the keyword `QUAD` instead of `QUADS` indicating that only one quadrilateral can be defined here. However, as more than one polygon can be defined within an initialization file, this difference is unproblematic.

Awareness likelihood of secondary emotions

The calculation of a secondary emotion's awareness likelihood depends on the location of the reference point in PAD space as well. As the secondary emotions *hope*, *fears-confirmed*, and *relief*, however, are represented in PAD space as areas (i.e. four sided polygons) instead of points, the computation is different from a primary emotion's awareness likelihood.

Each of the four vertices constituting a polygon has its own intensity value. Figure 6.5 shows an example of such a polygon.

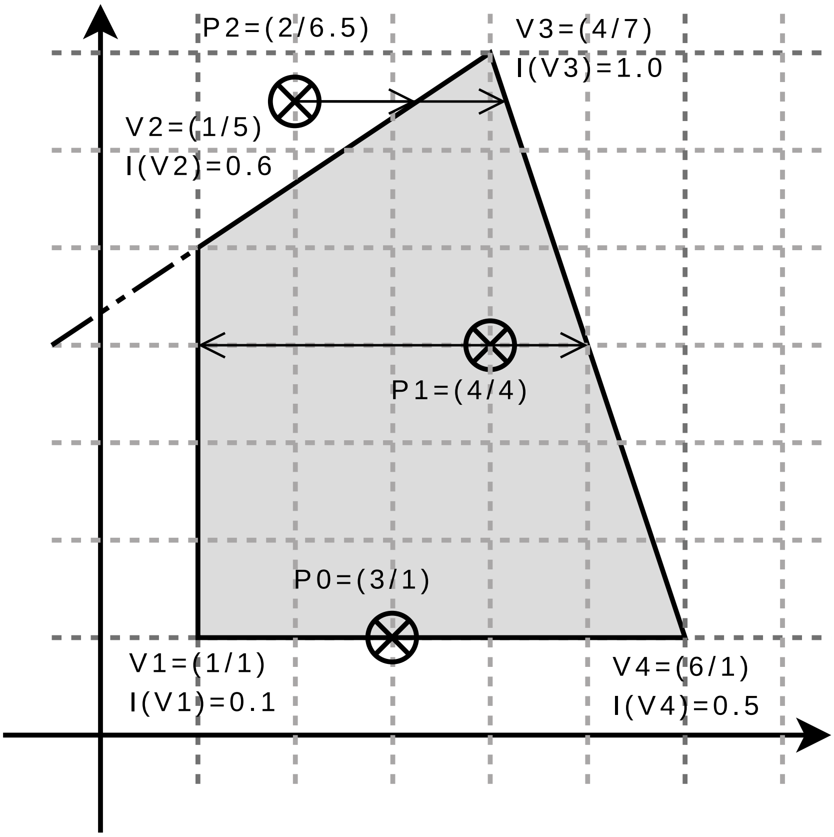


Figure 6.5: An example of a four sided polygon with intensity values in each of its vertices V1, V2, V3, and V4. The three reference points P0, P1, and P2 are examples of a possible trace of the reference point in PAD space over time.

The vertex `V1` with coordinate `(1/1)` has an intensity value `I(V1)` of 0.1, vertex `V2` at `(1/5)` an intensity `I(V2)` of 0.6, vertex `V3` at `(4/7)` has intensity `I(V3)=`1.0, and `V4` at `(6/1)` an intensity `I(V4)` of 0.5.

Three possible cases for the reference point have to be considered: (1) the reference point `P0` cuts the horizontal edge `(V1,V4)`; (2) reference point `P1` lies between the vertical edge `(V1,V2)` and some other edge; and (3) `P2` lies outside the polygon.

Case (1) In case of `P0` a linear interpolation is applied to the intensities of the left and the right vertex according to Equation 6.2.

$$i_{P0} \quad = \quad I(V1) + \frac{P0_x - V1_x}{V4_x - V1_x} \cdot (I(V4) - I(V1))$$

$$\begin{aligned} &= 0.1 + \frac{3-1}{6-1} \cdot (0.5 - 0.1) \\ &= 0.26 \end{aligned} \tag{6.2}$$

The resulting intensity value for the example polygon at the reference point *P0=(3/1)* is 0.26.

Case (2) For `P1` more calculations are necessary. The coordinate `(4/4)` lies within the polygon, but not directly on a vertical or horizontal edge and, thus, the intersection of the horizontal with the polygon edges has to be determined first. This is achieved by, first, establishing the linear equations for each of the edges according to Equation 6.3.

$$\begin{aligned} x &= \frac{y-b}{m} \\ m &= \frac{VB_y - VA_y}{VB_x - VA_x} \\ b &= VA_y - m \cdot VA_x \\ x &= \frac{(y - (VA_y - (\frac{VB_y - VA_y}{VB_x - VA_x}) \cdot VA_x)) \cdot (VB_x - VA_x)}{VB_y - VA_y}, \text{ with } VA \neq VB \end{aligned} \tag{6.3}$$

For edge `(V1/V2)` Equation 6.3 cannot be applied ($m = \infty$), but the coordinate of the intersection with this edge is simply $(V1_x/P1_y) = (1/4)$. Substituting *VA* with `V2` and *VB* with `V3` gives Equation 6.4 and evaluating this equation at $y = P1_y = 4$ results in Equation 6.5.

$$x = \frac{(y - (5 - (\frac{7-5}{4-1}) \cdot 1)) \cdot (4-1)}{7-5} \tag{6.4}$$

$$\begin{aligned} x &= \frac{(4 - (5 - \frac{2}{3})) \cdot 3}{2} \\ &= -0.5 \end{aligned} \tag{6.5}$$

The resulting coordinate $(P1_y/-0.5) = (4/-0.5)$, however, is outside the polygon and ignored (cf. the dashed line in Figure 6.5). Using Equation 6.3 for edge `(V3/V4)` and `P1` results in coordinate $(4/5)$, which lies on the edge and, thus, belongs to the polygon.

Next, two linear interpolations between, first, `I(V1)=`0.1 and `I(V2)=`0.6 at intersection `(1/4)` and, second, `I(V3)=`1.0 and `I(V4)=`0.5 at intersection `(5/4)` are calculated according to Equation 6.2. The resulting intensity values at these two intersections are 0.475 and 0.75 respectively.

Finally, the above values are taken to interpolate "horizontally" (similar to case one) between the intensity at intersection `(1/4)` (0.475) and at intersection `(5/4)` (0.75) for reference point `P1=(4/4)`. This calculation produces 0.68125 as the secondary emotion's intensity in case of the reference point being located at `P1`.

Case (3) Finally, `P2` is located outside the polygon, although it is inside the bounding rectangle that is span by the maximum and minimum values of the four vertices (indicated by the darker, dashed lines in Figure 6.5). Points outside the bounding rectangle are easily determined in advance and not further considered by applying the following algorithm:

```
boolean
inBoundingRectangle?(V1, V2, V3, V4, P) {
   int max_x = max(V1_x, V2_x, V3_x, V4_x);
   int min_x = min(V1_x, V2_x, V3_x, V4_x);
   int max_y = max(V1_y, V2_y, V3_y, V4_y);
   int min_y = min(V1_y, V2_y, V3_y, V4_y);
   if (P_x > min_x && P_x < max_x &&
       P_y > min_y && P_y < max_y) {
      return true;
   }
   return false;
}
```

This algorithm takes the four corners `V1`, `V2`, `V3`, and `V4` as well as a reference point `P` to be checked as arguments, calculates the maximum and minimum values, and only returns `true`, if the point `P` lies between these values, otherwise `false`.

In case of `P2`, however, the above algorithm returns `true` although the point does not lie within the polygon. To solve this case as well, first, the two edges are determined which intersect with the horizontal line through the reference point `P2`, i.e. the line $y = P2_y = 6.5$, by evaluating the linear equations of these edges at $y = 6.5$. This yields two values for x that are both greater than the value $P2_x = 2$ and, accordingly, the reference point `P2` must lie outside the polygon.

In more general terms: Given the values $C1_x$ and $C2_x$ for the intersections of the horizontal of a reference point P with any two edges of a polygon, Equation 6.6 most hold for the point P to lie within the polygon.

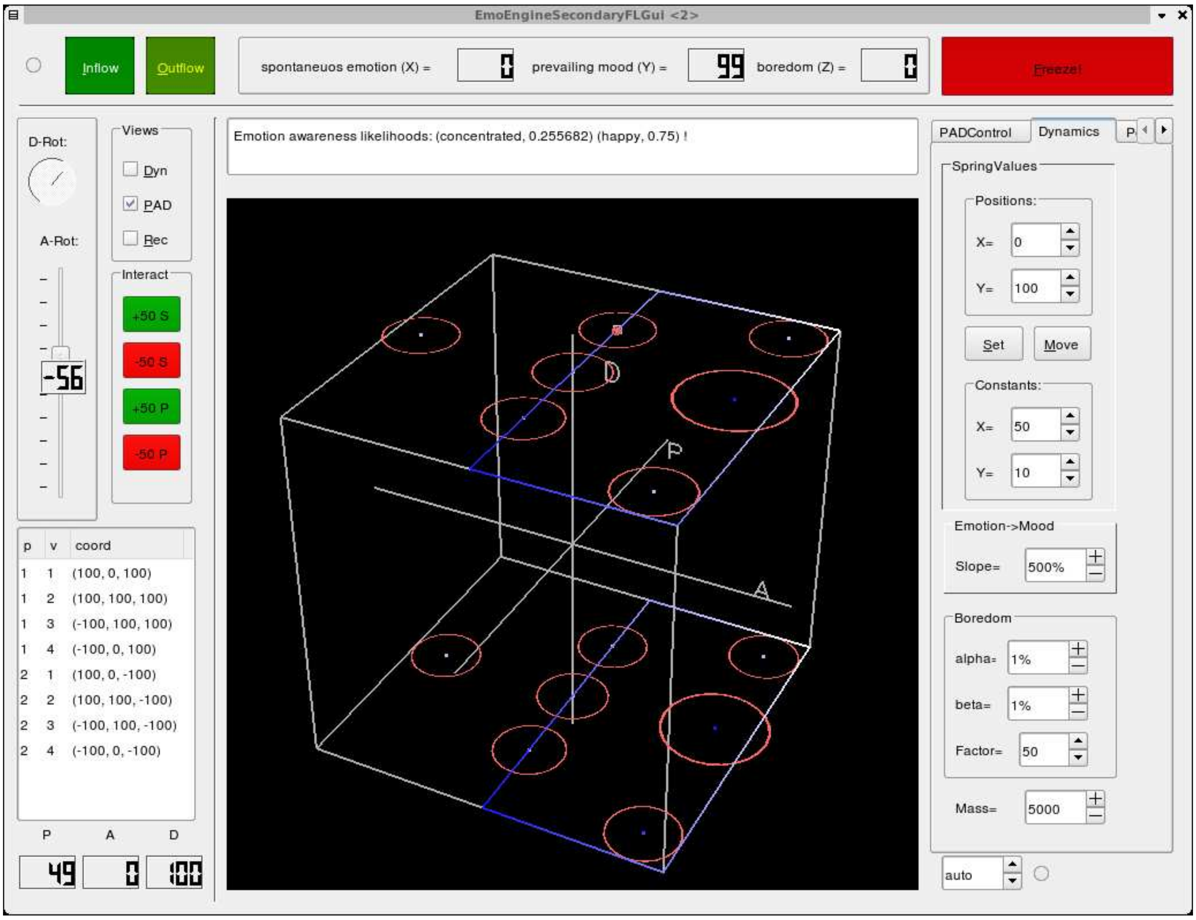
$$min(C1_x, C2_x) \leq P_x \leq max(C1_x, C2_x) \tag{6.6}$$

Of course, this condition also applies to the cases with the reference point outside the bounding rectangle, but as it involves more computation than the `inBoundingRectangle?`-algorithm, the overall performance is increased by first checking against the bounding rectangle.

The awareness likelihood of secondary emotions could already be determined by multiplying their overall intensity at time *t* after they were triggered by the *BDI-Agent* with the local intensity at the location of reference point `P`. Instead, however, in the current implementation the base intensity of each vertex is changed by premultiplying the polygon's overall intensity at time *t* before the reference point is taken into account. This way, the graphical representation of the polygon better reflects the secondary emotion's dynamically changing intensity distribution. The resulting intensity value at any given point within the polygon at time *t* does indeed not change by applying this algorithm.

6.2.3 The graphical user interface

The graphical user interface (GUI) of the emotion simulation system of Becker (2003) was modified to account for the additional visualization of secondary emotions and enable the online modification of all parameters of the emotion module during runtime. It comprises of



spontaneuos emotion (X) = 0 prevailing mood (Y) = 99 boredom (Z) = 0

Emotion awareness likelihoods: (concentrated, 0.255682) (happy, 0.75) !

p	v	coord
1	1	(100, 0, 100)
1	2	(100, 100, 100)
1	3	(-100, 100, 100)
1	4	(-100, 0, 100)
2	1	(100, 0, -100)
2	2	(100, 100, -100)
2	3	(-100, 100, -100)
2	4	(-100, 0, -100)

P 49 A 0 D 100

SpringValues
Positions: X= 0 Y= 100
Constants: X= 50 Y= 10
Emotion->Mood Slope= 500%
Boredom alpha= 1% beta= 1% Factor= 50
Mass= 5000
auto

Figure 6.6: In this screenshot of the graphical user interface (GUI) of the emotion module the blue frame indicates the graphical visualization of the secondary emotion *hope* in PAD space with an intensity of 0.0.

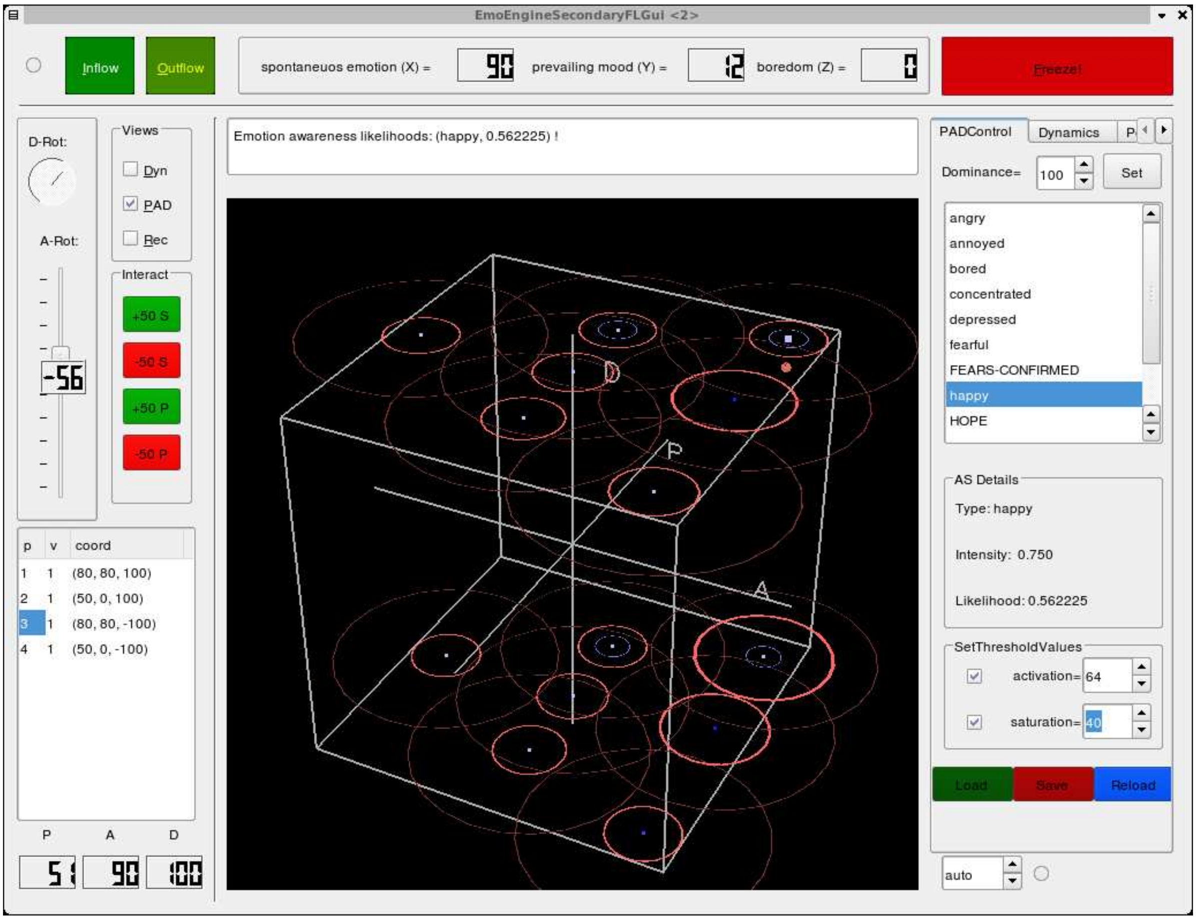
a three dimensional visualization of PAD space in its center surrounded by a standard user interface realized with Qt (Trolltech 2008) (cf. Figures 6.6 and 6.7).

In Figure 6.6 the secondary emotion *hope* is highlighted in PAD space, although it has an intensity of 0.0 at that moment. Its two areas in PAD space (in the *high dominance* and *low dominance* plane, respectively) are indicated by the two blue frames. To the lower left in Figure 6.6 the coordinates of the vertices of the two areas can be examined. This list is updated as soon as another emotion is selected on the right side of the GUI within the “PADControl” tab (see right side of Figure 6.7).

In the “Dynamics” tab presented in Figure 6.6 to the right all parameters of the emotion dynamics (cf. Section 6.2.1) can be adjusted if necessary. The actual values of the emotion dynamics “spontaneous emotion (X)”, “prevailing mood (Y)” and “boredom (Z)” are displayed at the top of Figure 6.6 and the corresponding “P” and “A” values together with the actual “D” value in the lower left corner.

The visualisation of PAD space can be rotated around the *Dominance* and *Arousal* axis independently by means of the “D-Rot” wheel and the “A-Rot” slider to the left of Figure 6.6. Thus, the course of the reference point through PAD space can be supervised intuitively.

Figure 6.7 presents another view of the emotion module’s GUI in which the primary emo-



spontaneuos emotion (X) = 90 prevailing mood (Y) = 12 boredom (Z) = 0

Emotion awareness likelihoods: (happy, 0.562225) !

☐ Dyn ☑ PAD ☐ Rec

A-Rot: -56

p	v	coord
1	1	(80, 80, 100)
2	1	(50, 0, 100)
3	1	(80, 80, -100)
4	1	(50, 0, -100)

P	A	D
51	90	100

Dominance= 100

AS Details

Type: happy

Intensity: 0.750

Likelihood: 0.562225

SetThresholdValues

☑ activation= 64

☑ saturation= 40

auto ○

Figure 6.7: This screenshot shows the primary emotion *happy* being highlighted and “activated” in the top right corner of PAD space.

tion *happy* is highlighted after a selection of its name in the list of emotions to the right. Below this list (in the “PADControl” tab) the details of the selected affective state (AS) are always updated (“AS Details”), i.e. the emotion’s “Type”, its current “intensity”, and its awareness “likelihood”; the last two values are continuously changing as a result of the implemented emotion dynamics.

Every primary emotion’s saturation and activation thresholds are visualized by red circles of different sizes in PAD space, if the corresponding check-boxes are ticked in the “SetThresholdValues” area of the “PADControl” tab. These thresholds can be adjusted at runtime—not only for every emotion independently, but also for every single vertex of a primary emotion, if it is located in PAD space more than once.

The primary emotion *happy*, for example, is located four times in PAD space and, accordingly, to the left in Figure 6.7 four entries, i.e. vertices, are given in the list of coordinates. Vertex number three is selected in this list and, consequently, the change of the “saturation” value (cf. Figure 6.7, right side) to 40 only applies to the one representation of *happy* located at $(80, 80, -100)$, i.e. the lower right corner of PAD space. This relatively high saturation threshold is visualized by a red circle with a bigger diameter than all other saturation circles.

The visualization of the activation and saturation thresholds also shows that most primary emotions overlap with at least one other emotion. Accordingly, in the white box at the top of Figure 6.6 two awareness likelihoods are reported simultaneously (i.e. *concentrated* with a likelihood of 0.255682 and *happy* with a likelihood of 0.75). So far, however, only the primary emotion with the highest likelihood is driving MAX's facial expressions, although the *Emotion-Agent* distributes all available information as a vector of emotion/likelihood pairs to the other agents by means of message communication.

6.3 The WASABI architecture's information flow

The sequence diagram in Figure 6.8 illustrates an example information flow within the WASABI architecture.

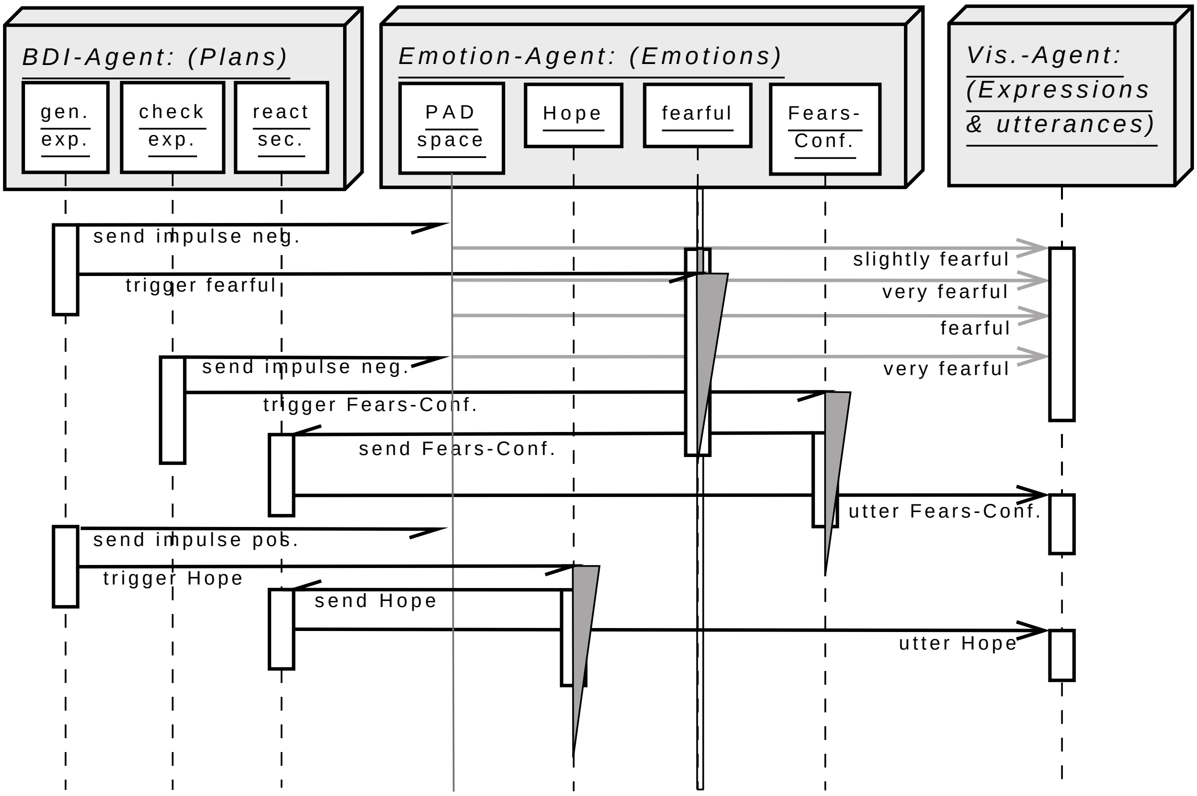


Figure 6.8: Sequence diagram of the information flow between the WBS-agents in Skip-Bo

The three agents *BDI-Agent*, *Emotion-Agent*, and *Visualization-Agent* ("Vis.-Agent") are represented as boxes in the top of Figure 6.8. In the top-left box *BDI-Agent* the three plans *generate-expectation* ("gen. exp.", Plan 6.11, p. 133), *check expectations* ("check exp.", Plan 6.12, p. 135), and *react-to-secondary-emotion* ("react sec.", Plan 6.13, p. 136) are rendered as three white rectangles to show their activity below. The same type of rectangle is used to depict the *PAD space* as well as the emotions *Hope*, *fearful*, and *Fears-Confirmed* ("Fears-Conf.") which all reside in the *Emotion-Agent*. The internal realization of the *Visualization-Agent* is not detailed here and in this example it only receives messages from the other agents, although in reality it also distributes information about the human player's interaction with the game interface by sending messages to the *BDI-Agent*.

An exemplary sequence of message communication is shown in Figure 6.8 with the timeline from top to bottom. In this example the *generate-expectation* plan is being called after MAX played his last card. This plan, first, *sends* a *negative impulse* (“`send impulse neg.`”) to the *Emotion-Agent* thereby indirectly changing MAX’s emotional state in *PAD space* (cf. Section 4.2, p. 87). Subsequently, while following the same plan, the primary emotion *fearful* is being *triggered* (“`trigger fearful`”) by the *BDI-Agent*—probably because MAX expects the human player to play an important card.

In the *Emotion-Agent*, however, the negative emotional impulse already pushed the reference point in PAD space close enough to the (not yet triggered) emotion *fearful* to let MAX experience *fear* with low intensity, because *fearful* has a slightly positive base intensity of 0.25. In Figure 6.8 this non-zero base intensity of *fearful* is indicated by the small double line along the dashed, vertical lifeline of *fearful*. Accordingly, “slightly fearful” is sent to the *Visualization-Agent* even before the *BDI-Agent* triggers the emotion *fearful*. As the intensity of *fearful* in the *Emotion-Agent* abruptly changes with the incoming *trigger fearful* message, MAX’s emotional state changes from *slightly* to *very fearful*. Such sudden changes in intensity are reproduced in Figure 6.8 by the three, gray triangles drawn along the emotion’s lifelines.

The intensity of *fearful* decreases within the next ten seconds and the reference point possibly changes its location in PAD space due to the implemented emotion dynamics. Thus, *very fearful* automatically changes to *fearful* (see right side of Figure 6.8) without any further *impulse* or *trigger* messages.

In the *BDI-Agent* the *check expectations* plan is activated next to check, whether a human player’s action meets the previously generated expectations. In the example the *BDI-Agent*, first, sends a *negative impulse* to the *Emotion-Agent* thereby indirectly changing the reference point’s location in PAD space such that MAX gets *very fearful* again. This sequence of different emotion intensities (*slightly fearful*, *very fearful*, *fearful*, *very fearful*) is possible for every primary or secondary emotion, although it is only exemplified for *fearful* in Figure 6.8. It follows from the dynamic interplay of lower-level emotional impulses and cognitively triggered changes in emotion intensity.

The *check expectations* plan *triggers* the secondary emotion *Fears-Confirmed* (“`trigger Fears-Conf.`”) in the *Emotion-Agent* thereby maximizing its intensity. Together with the negatively valenced mood *fears-confirmed* acquires a non-zero awareness likelihood, which is *sent* back to the *BDI-Agent* (“`send Fears-Conf.`”). The plan *react-to-secondary-emotion* is executed to process the incoming message and results in an “`utter Fears-Conf.`” message, which is *sent* to the *Visualisation-Agent* letting MAX produce an appropriate utterance (cf. Table 6.3, p. 137).

After the human player played a card on a center pile, MAX generates new expectations by means of the *generate expectations* plan. In the current example this plan, first, sends a positive impulse (“`send impulse pos.`”) to the *Emotion-Agent*, which influences MAX’s emotion dynamics in PAD space. Shortly afterwards the *BDI-Agent* triggers the secondary emotion *Hope* (“`trigger Hope`”) such that its intensity is maximized within the *Emotion-Agent* resulting in a non-zero awareness likelihood of *Hope*. Again, MAX’s awareness of *Hope* is realized by the *Emotion-Agent* sending an appropriate message back to the *BDI-Agent* (“`send Hope`”), which lets MAX utter an according sentence (cf. Table 6.5, p. 137).

The same mechanisms are realized for all other primary and secondary emotions leading to a continuous elicitation of mood congruent emotions, that are verbally and non-verbally expressed by MAX.

6.4 Evaluation of secondary emotion simulation

To evaluate the effect of secondary emotion simulation in concert with primary emotions hypothesis 6.1 is derived from the psychological findings discussed in Chapter 2.

Hypothesis 6.1 *MAX with primary and secondary emotions is judged older than MAX directly expressing only primary emotions.*

In Section 2.2.2 on page 54 the discussion of the ontogenetical background of emotion development suggests that secondary emotions are a product of ontogenetical development. Furthermore, children are less able to suppress their emotional expressions than adults. Accordingly, subjects playing Skip-Bo against a version of MAX with the new WASABI architecture (described in the previous sections) are believed to judge him older than subjects playing against a version of MAX with the older emotion dynamics system developed in the author's diploma thesis and empirically validated in the first empirical study (cf. Chapter 5).

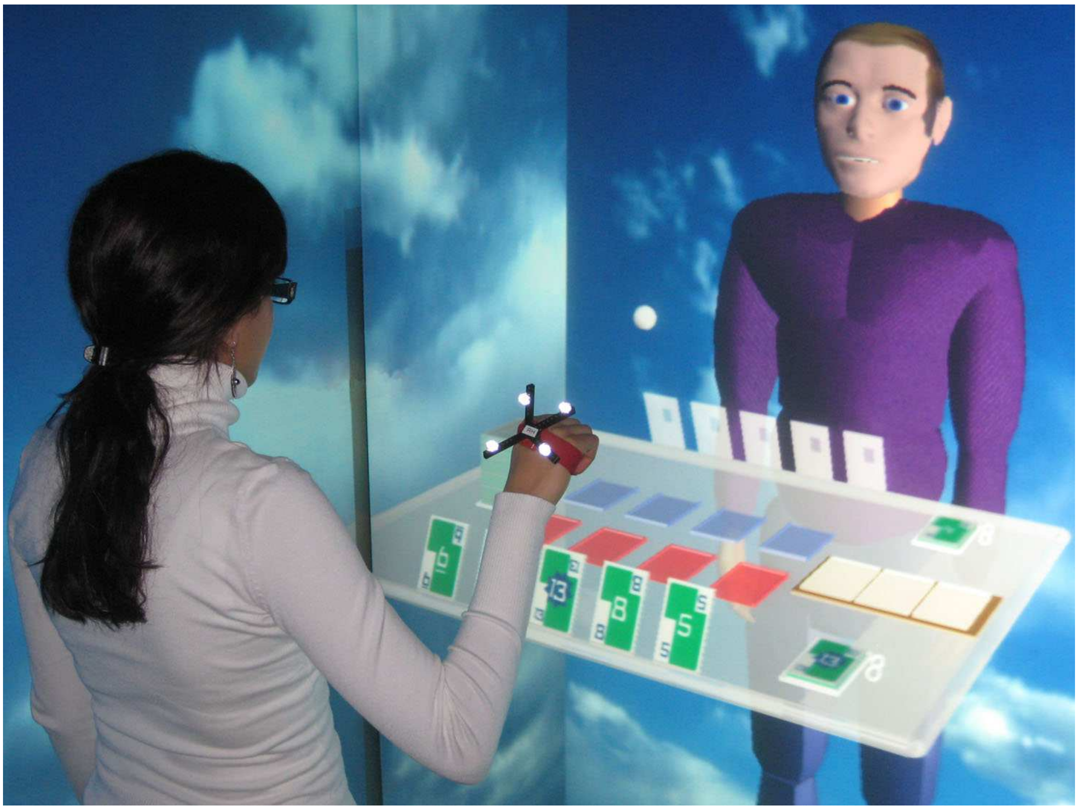


Figure 6.9: Skip-Bo against MAX in the three-sided large-screen projection system

6.4.1 Skip-Bo against MAX in the CAVE-like environment

The three-sided large-screen projection system allows for stereoscopic projection of interactive virtual environments. Together with marker-based motion tracking of the human player a high level of naturalness is achieved in human-computer interaction.

Setup

Figure 6.9 shows a human player just starting to play Skip-Bo against MAX in the CAVE-like environment. The game is projected between the human player and MAX in such a way that the human player gets the impression of a half-transparent white table that is slightly tilted toward him or her and on which his or her own hand cards are invisible to MAX. As the human player wears special glasses with polarization filters and markers he or she not only perceives the virtual world three-dimensional but is also able to inspect it by moving around within the physical boundaries of the installation. When doing so, MAX follows the human's movements with his eyes and head giving the impression of holding up eye contact.

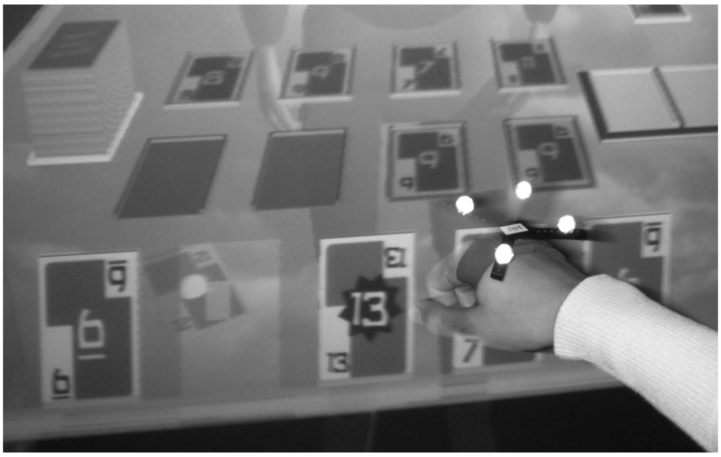
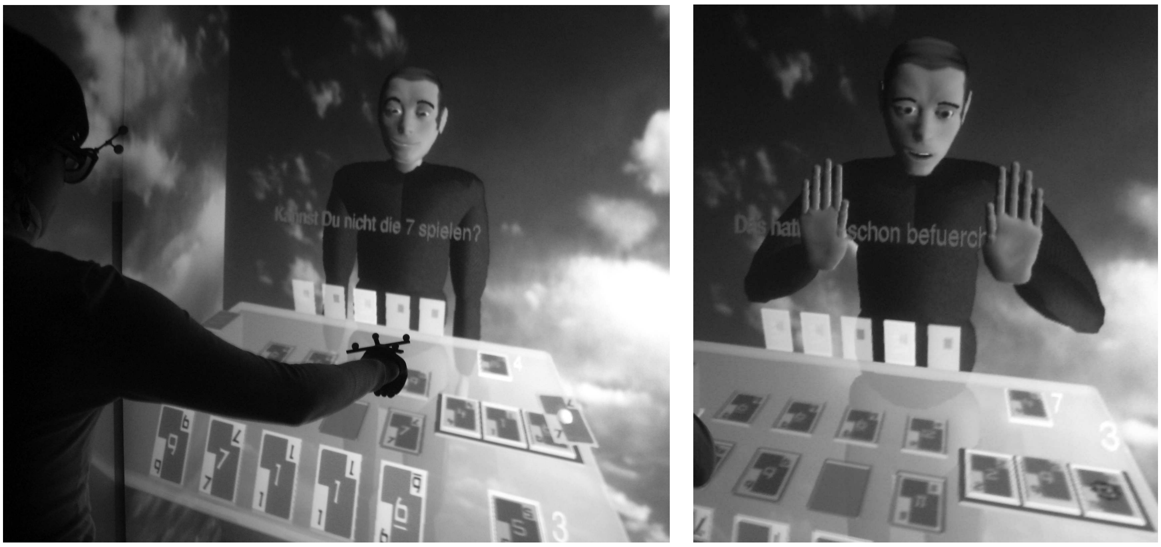


Figure 6.10: A card attached to the human player's white sphere

For the human to interact with the game he or she is equipped with a "rigid body" on the palm of his or her right hand (cf. Figure 6.9). Approximately ten centimeters in front of the rigid body a white sphere is visualized and its position as well as rotation is constantly updated with every movement of the human player's right hand. Thus, it is easy for the human player to use this virtual reference as a kind of three-dimensional pointer to select objects in front of him or her. As soon as the sphere touches one of the topmost virtual cards of the human player's stock piles or one of his or her hand cards, it is attached to the sphere (cf. Figure 6.10). By afterwards virtually touching one of his or her own stock piles or any of the three center piles the human player plays this card on one of these piles (cf. Figure 6.13(a)). MAX then controls the validity of this move and corrects it, if it was invalid, by moving the card back to the human's hand or stock pile.

During the human player's turn MAX performs the same gaze behavior as implemented for the first empirical study (cf. Figure 6.11(b)). When the human selects a card, he looks at the source of the card for two seconds before resuming to track the human player's head. After a valid move of the human player, however, MAX acknowledges this action as described in the context of Plan 6.2 on page 124. This is a difference to the first empirical study (cf. Section 5.2.2, p. 107), in which MAX gave no verbal feedback. These short acknowledgment sentences are added to this study, because MAX is only able to verbally report on the awareness of secondary emotions. If he were only producing utterances in the second experimental condition, but not in the first one, the conditions would be too different to each other.

MAX interacts with the game in the same way as described in Section 5.2.2 except for the



(a) MAX expresses his *hope* that the human player will play the card with the number seven next by saying “Kannst Du nicht die 7 spielen?” (Can’t you play the seven?)

(b) MAX realizes that his *fears* just got *confirmed* and utters “Das hatte ich schon befürchtet!” (I was already afraid of that!)

Figure 6.11: MAX expressing his *hope* and realizing that his *fears* got *confirmed*

additional utterances he now performs. He welcomes the human player in the beginning, says correcting sentences in case of a human player’s mistake and expresses his emotional state verbally in case of the awareness of any of the secondary emotions *hope*, *fears-confirmed*, or *relief*. In order to avoid misunderstandings every sentence uttered by MAX is displayed for twelve seconds in front of him as a “subtitle” (cf. Figure 6.11).

Subjects

Fourteen male and nine female subjects participated in the study and all but one subject were German. Their age ranged from 13 to 36 years and the average age was 23 years. The subjects were randomly assigned to the conditions resulting in the distribution given in Table 6.8.

	male	female	$\sum$
Condition (1)	6	5	11
Condition (2)	8	4	12
$\sum$	14	9	23

Table 6.8: The distribution of the subject’s gender on the two experimental conditions

Design

In order to assess the effect of secondary emotion simulation in addition to the simulation of primary emotions in the context of human-computer interaction and to validate Hypothesis 6.1, the following two conditions were designed:

(1) *Only primary emotions* condition: The emotion simulation is constrained to primary emotions and MAX expresses them directly by means of facial expressions and “affective sounds” such as grunts and moans. He appraises the actions of the human player negatively and his own progress in the game positively. He feels *dominant* whenever it is his turn and *submissive* (i.e. non-dominant) whenever it is the human player’s turn.

(2) *Primary and secondary emotions* condition: In addition to the setup of condition (1) secondary emotions are simulated in this condition and expressed verbally by MAX in case of positive awareness likelihood (cf. Section 6.3).

Notably, the number of verbal utterances performed by MAX is likely to be higher in condition (2) than in condition (1). This difference, however, adds to the impression of MAX as a less child-like interaction partner in condition (2), because young children are also less good at expressing their feelings verbally.

In order to model condition (1) the emotion module of the WASABI architecture is initialized according to Listing 6.4.

Listing 6.4: Initialization file `initPri.emo_pad` with only primary

```
# ONLY PRIMARY EMOTIONS
fearful -0.8 0.8 -1 MOOD_FEARFUL 0.2 0.64 1.0
concentrated 0 0 -1 MOOD_CONCENTRATED 0.2 0.64 1.0
concentrated 0 0 1 MOOD_CONCENTRATED 0.2 0.64 1.0
depressed 0 -0.80 -1 MOOD_SAD 0.2 0.64 1.0
happy 0.8 0.8 1 MOOD_FRIENDLY 0.2 0.64 1.0
happy 0.5 0 1 MOOD_FRIENDLY 0.2 0.64 1.0
happy 0.8 0.8 -1 MOOD_FRIENDLY 0.2 0.64 1.0
happy 0.5 0 -1 MOOD_FRIENDLY 0.2 0.64 1.0
bored 0 -0.85 1 MOOD_BORED 0.2 0.64 1.0
annoyed -0.5 0 1 MOOD_SAD 0.2 0.64 1.0
sad -0.5 0 -1 MOOD_SAD 0.2 0.64 1.0
surprised 0.1 0.8 1 MOOD_SURPRISED 0.2 0.64 1.0
surprised 0.1 0.8 -1 MOOD_SURPRISED 0.2 0.64 1.0
angry -0.8 0.8 1 MOOD_ANGRY 0.2 0.64 1.0
```

It is different to Listing 6.2 (p. 140) in the following aspects:

- The three secondary emotions *hope*, *fears-confirmed*, and *relief* are not included.

- Every primary emotion has the same *saturation* (0.2) and *activation* (0.64) threshold as well as *base intensity* (1.0).

In effect, by initializing the emotion module with the values of Listing 6.4 the simpler *emotion simulation system* of Becker (2003) is reproduced within the WASABI architecture.

To realize condition (2) the emotion module is initialized according to Listings 6.2, 6.3, D.1, and D.2 (cf. Section 6.2.1 and Appendix D).

Procedure

Subjects received written instructions of the card game (in German) with a screenshot of the starting condition and got the chance to ask clarifying questions about the gameplay before they entered the room with the three-sided large-screen projection system. Subjects entered the room individually and were equipped with the special glasses and the marker for the right hand. They were briefed about the experiment, in particular that they would play a competitive game. Then, subjects could play a short introductory game against a non-emotional MAX, which allowed them to get used to the interface, and also provided subjects the possibility to ask clarifying questions about the game. Each subject won this first game easily.

From now on, the experimenter remained visually separated from the subject only to supervise the experiment. After the game was reset manually, MAX welcomed the subject and asked him or her to play the first card. After the game was completed, the subjects were asked to fill in a questionnaire in German presented on the screen of another computer in the room next door. The questionnaire was the same as in the first empirical study except for one additional question (17b) asking for the presumed age of MAX (see Appendix B).

Results

The analysis of the questionnaires (cf. Figure 6.12) showed that all subjects liked to play the game, got sufficient instructions in advance, felt comfortable during the game, and wanted to play again with no significant differences due to the experimental condition.

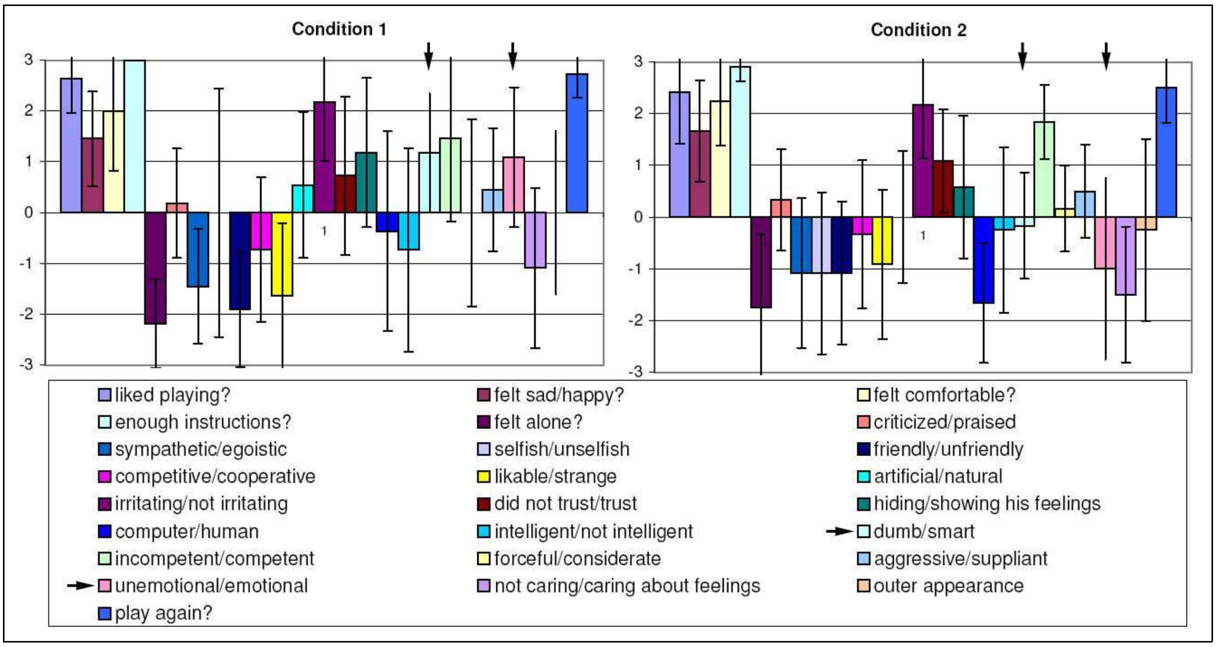
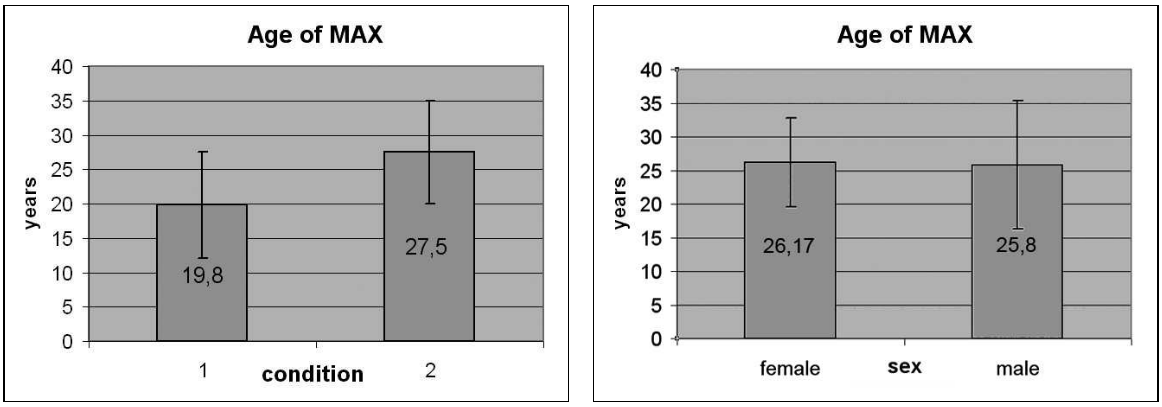


Figure 6.12: The mean values and standard deviations of the questionnaires for *primary emotions only* condition (1) and *primary and secondary emotions* condition (2) of the second empirical study (the highlighted results are discussed in the text)

Only the answers to three questions were significantly different between conditions. By answering question number 18 (“What kind of person has MAX been?”) subjects could judge MAX to be a “dumb” or “smart” person. The subjects participating in the *primary emotion* *only* condition (1) judged MAX to be smart (mean value 1.18) whereas the mean value of -0.16 in case of condition (2) indicates that the subjects playing against MAX with both primary and secondary emotions were quite undecided concerning this question (a two-tailed t-test results in p = 0.008). Although this result could be interpreted as MAX being judged less smart when additionally simulating secondary emotions, it is important to note that no significant difference between conditions occurred with respect to the very similar question number 17, which is concerned with MAX's level of intelligence.

The second statistically significant difference appeared in the answers to question number 22 ("How emotional did Max react?"). Participants of the *primary emotions only* condition (1) found MAX's reactions to be emotional (mean value 1.1) whereas subjects of the *primary and secondary emotions* condition (2) judged his reactions as unemotional (mean value -1.0, p = 0.004).



(a) Between *primary emotions only* condition (1) and *primary and secondary emotions* condition (2) a significant difference occurred

(b) No gender-related effects appeared with regard to question 17b

Figure 6.13: The mean values and standard deviations of the answers to question number 17b "If MAX were a real human, how old would you judge him to be?"

To verify Hypothesis 6.1 question number 17b ("If MAX were a real human, how old would you judge him to be?") was added to the questionnaire. The two mean values and standard deviations of the subject's answers to this question are presented in Figure 6.13(a). In the *primary emotions only* condition (1) MAX was assumed to be significantly younger (mean value 19.8 years, standard deviation 7.7) than in condition (2), in which secondary emotions were simulated as well (mean value 27.5, standard deviation 7.5). A two-tailed t-test assuming unequal variances results in p = 0.025.

As the distribution of male and female subjects varied between conditions (cf. Table 6.8) the answers to question number 17b of all nine female subjects were compared to the answers of the 14 male subjects regardless of the experimental condition. The mean values of these two groups did not differ significantly (cf. Figure 6.13(b)) such that no gender effects occurred. This result strengthens the supposition that the between conditions difference can be interpreted to confirm the initial Hypothesis 6.1.

6.4.2 Conclusion

The results of this second study are to some respect unexpected. Intuitively one could have expected that the perceived level of intelligence of MAX—being judged by the subjects—should be higher in case of the additional simulation of secondary emotions. The questionnaire showed, however, an opposite trend in the answers related to intelligence and smartness of MAX.

A possible explanation of this contra intuitive effect is connected with the different age that is attributed to MAX in both conditions. Although MAX's outer appearance stayed the same in both conditions, the ascribed age differed significantly. Thus, it is reasonable to assume that subjects judging MAX younger (condition 1) might have had less expectations concerning his level of intelligence and smartness than those who judged him older (condition 2). Accordingly, the mismatch between expected behavior based on ascribed age and actual behavior resulted in the subjects of condition (2) judging MAX to be less smart.

The difference in perceived emotionality of MAX might result from the lower base intensities of primary emotions in the *primary and secondary emotions* condition (2). In this condition MAX is less often surprised than in condition (1), because the base intensity of *surprise* is set to 0.0. Accordingly, surprise cannot be elicited before BDI-based reasoning processes appraised an event as unexpected.

Hypothesis 6.1 could be confirmed. When secondary emotions are added to the simulation of primary emotions MAX is judged significantly older. In other terms, the more complex Affect Simulation achieved by the WASABI architecture matches MAX's outer appearance better than the previously developed *emotion simulation system* of Becker (2003), when the findings of developmental psychology are taken into account.

6.5 Summary

In this chapter secondary emotions were integrated into the WASABI architecture by, first, extending the cognitive reasoning capabilities of the BDI-based cognition module in the Skip-Bo scenario letting it generate and process expectations.

Second, the emotion module was extended to combine primary emotions represented as single vertices with particular intensities with secondary emotions represented as four-sided polygons with variable intensities for each of its four vertices. The two different intensity functions were explained and combined in the dynamic calculation of an emotion's awareness likelihood resulting from its configuration in PAD space. An overview of the graphical user interface for supervising the emotion dynamics and changing its parameters at runtime was given and the information flow of the WASABI architecture was exemplified to provide an overview of the internal message communication between the WBS-agents in the Skip-Bo scenario.

Finally, a second empirical study was conceived to validate the hypothesis that MAX with the previous emotion simulation system—being limited to only primary emotions—would be judged significantly younger than an emotional MAX driven by the new WASABI architecture, in which primary and secondary emotions are simulated in combination. The results of the study confirm this hypothesis.

7 Résumé

The subject of this thesis is the development of a computational simulation of affect for embodied agents. The conceptualized WASABI architecture ([W]ASABI [A]ffect [S]imulation for [A]gents with [B]elievable [I]nteractivity) builds upon the author's previous implementation of an emotion dynamics for artificial humanoid agents, that was limited to the simulation and direct expression of primary emotions.

The author follows two motivations in proposing his Affect Simulation Architecture:

1. A suitable simulation of affect is assumed to increase the believability of embodied agents and, thus, to facilitate human-computer interaction. Therefore, a comprehensive simulation has to be conceptualized, computationally realized, its effects empirically investigated, the initial conception refined if necessary, reimplemented, and empirically investigated again, and so on. With the development of the WASABI architecture the author consequently followed this cycle of computational implementation and empirical investigation with the aim to increase the believability of the virtual human MAX.

2. Researchers coming from different fields outside the Computer Science community are interested in using the increasingly powerful computer simulations of humanoid agents to investigate the applicability and validity of their theoretical conceptions. With the development of the WASABI architecture the author takes an interdisciplinary approach by combining findings from psychology, neurobiology, and cognitive science based on computational methods of Artificial Intelligence.

Researchers in the field of Affective Computing mainly follow the rational reasoning approach to modeling emotions for their virtual or robotic embodied agents. Accordingly, most of them build upon the "Cognitive Structure of Emotions" as proposed by Ortony et al. (1988), which is commonly known as the OCC-theory of emotions. This semantics-based theoretical approach, however, is best suited to derive logical rules for agents that reason about emotions rather than have emotions of their own. Therefore, most OCC-based implementations extend this theory by integrating other affective phenomena such as personality or mood and some use fuzzy logics to integrate learning into their architectures. The resulting emotions are driving or at least modulating an animated agent's verbal and non-verbal expressions, may it be a virtual or robotic animal or humanoid agent.

To this respect the foremost motivation of Affective Computing researchers is to increase their agents believability—the first motivation above. Whether these computational affective states are somewhat comparable to their biological archetypes, is of minor interest and mostly regarded as an unsolvable "philosophical" question.

This question—how similar not only the results but also the underlying processes are to the biological prototype—is central to those researchers who are motivated to conduct interdisciplinary research (see the second motivation above). A review of the interdisciplinary background reveals that the findings of different disciplines are more than only compatible to each other. They can be interpreted to support the general idea of a dynamic interplay between an organisms cognitive, conscious and non-conscious processes in the brain and its evolutionary, older homeostatic regulation of the body. This dynamics is proposed to result in “hot”, consciously experienced emotions, which are—after successful implementation on a machine—not only more plausible to the human interlocutor, but also help to carefully validate the predictions derived from the underlying theoretical conceptions.

7.1 Results

The WASABI architecture follows the theoretical separation of “bodily” emotion dynamics and cognitive appraisal. The emotion dynamics is based on dimensional emotion theory and combined with the BDI-based reasoning capabilities of the virtual human MAX, that are used to model *prospect-based* emotions (Ortony et al. 1988). In contrast to other OCC-based emotion simulation architectures, however, the most often direct, rule-based, connection between appraisal outcome and emotion elicitation is broken up by modeling the influence of simulated bodily feedback.

The WASABI architecture

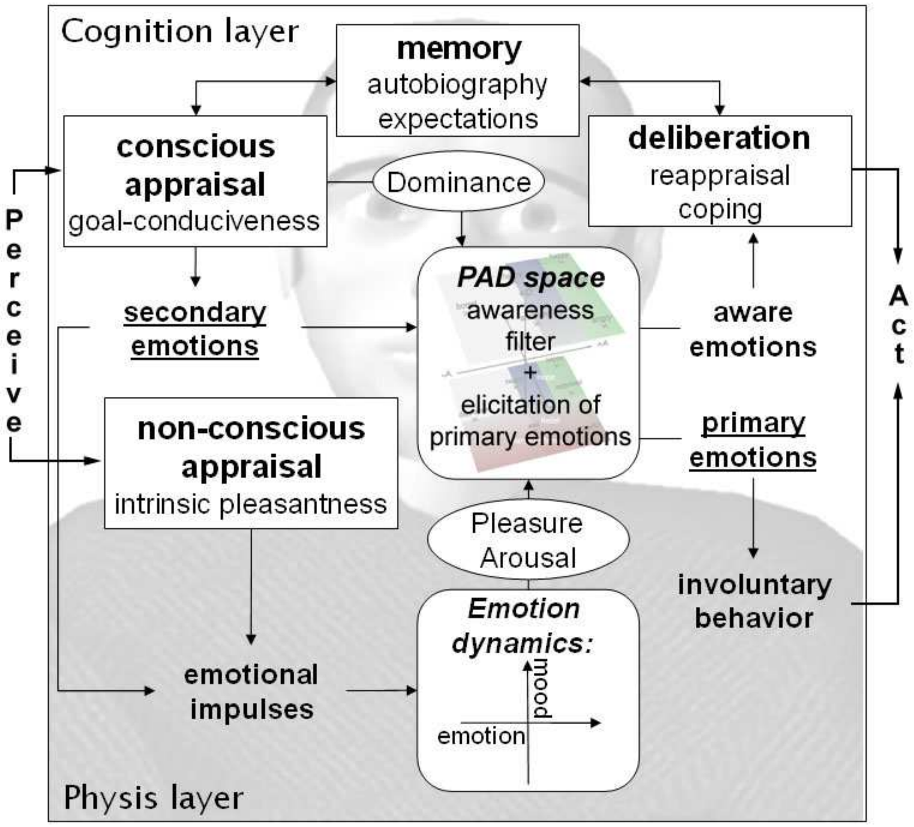


Figure 7.1: The conceptual distinction of cognition layer and physis layer in the WASABI architecture

In Figure 7.1 the conceptual distinction of an agent's simulated physis (i.e. body) and its cognition is presented and the different modules and components of the WASABI architecture are assigned to the corresponding layers.

To the left of Figure 7.1 the virtual human MAX perceives some (internal or external) stimulus. *Non-conscious appraisal* is realized by directly sending a small positive *emotional impulse* to the *Emotion dynamics* component of the WASABI architecture, e.g., when MAX in the museum guide scenario detects a skin colored region in the video stream. This establishes the "low road" (LeDoux 1996, cf. Figure 2.11(a)) of primary emotion elicitation. The presence of visitors in the museum is interpreted as *intrinsically pleasant* similar to Scherer (2001).

Another path resulting in *emotional impulses* begins with *conscious appraisal* of the perceived stimulus (cf. Figure 7.1, top left). This process resides in the *Cognition layer*, because it is based on the evaluation of goal-conduciveness of an event (Scherer 2001) and can be considered the "high road" of emotion elicitation (LeDoux 1996, cf. Figure 2.11(a)). Therefore, MAX exploits his BDI-based cognitive reasoning abilities to update his *memory* and generate *expectations*. These deliberative processes not only enable MAX to derive his subjective level of *Dominance* from the situational and social context, but also lead to the proposal[1] of cognitively plausible *secondary emotions*.

These *secondary emotions* are, however, first *filtered* in *PAD space*, before MAX might get *aware* of them (cf. Figure 7.1, middle). Independent of this filtering process, every cognitively plausible *secondary emotion* influences the *Emotion dynamics* component of the WASABI architecture, thereby modulating MAX's *Pleasure* and *Arousal* values, i.e. his simulated physis in the *Physis layer*. This influence is achieved by interpreting the valence component of any *secondary emotion* as an *emotional impulse* (cf. Figure 7.1, left). This way, *secondary emotions* "utilize the machinery of primary emotions" (Damasio 1994, cf. Figure 2.12(b)), because they might result in the elicitation of mood-congruent *primary emotions*, which—in the WASABI architecture—drive MAX's facial expressions *involuntarily*. Furthermore, as the *Pleasure* and *Arousal* values are incessantly modulating MAX's *involuntary behaviors* (i.e. breathing and eye blinking) as well, even "unaware" *secondary emotions* have an effect on MAX's bodily state and involuntary behavior.

In combination with the actual level of *Dominance*, *primary emotions* are elicited by means of a distance metric in *PAD space*. As mentioned before, these primary emotions are directly driving MAX's facial expressions. Although this automatism might be considered unnatural for an adult human, it has proven applicable and believable in the situational contexts in which MAX was integrated so far.

After the awareness filter has been applied, the resulting set of *aware emotions* consists of primary and secondary emotions together with their respective awareness likelihoods. They are finally subject to further deliberation and reappraisal resulting in different coping behaviors. A situation-focused coping behavior is implemented in the museum guide scenario by letting MAX leave the display, when he gets aware of being very angry. In the card game scenario the direct vocal and facial expression of negative emotions has proven sufficient to let the human players play in accordance with the rules.

[1] In technical terms this "proposal" is called *triggering* of secondary emotions.

Primary emotions—the first empirical study

After successful integration of the previous *emotion simulation system* into the cognitive architecture of MAX in the context of the museum guide scenario, it was reasonable to more carefully validate the simulation of an emotion dynamics which is rather independent of an agent's cognitive abilities. Therefore, the simulation of primary emotions was systematically tuned to realize positive and negative empathic behavior in a card game scenario that was taken to Japan and combined with bio-metrical emotion recognition based on skin conductance and electromyography.

In general, MAX was more perceived as a human being the more emotional reactions he showed, because human-likeness was rated higher in both empathic conditions than in the non-emotional or self-emotional condition. Even his outer appearance, albeit not changed between conditions, was rated more positive in the empathic than in the non-empathic conditions.

The statistical analysis of the questionnaires as well as the bio-metrical data confirmed the hypothesis that MAX's emotional reactions in this competitive scenario are less stressful and irritating for human players, if also negative emotions are simulated and expressed. Furthermore, a certain emotional contagion between MAX and the human player was detected in that the emotions expressed by MAX induced similarly valenced emotions in the human player.

Secondary emotions—the second empirical study

With the positive results of the first empirical study it was reasonable to further elaborate the idea of emotion dynamics in the attempt to integrate secondary emotions. To simulate secondary emotions MAX's cognitive reasoning abilities are extended enabling him to process expectations within the Skip-Bo card game scenario. Based on these expectations the mutual connection between cognition and emotion gives rise to cognitively plausible, prospect-based, secondary emotions, that are mood-congruent and "cognitively elaborated" (Ortony et al. 2005). MAX expresses his awareness of the secondary emotions *hope*, *fears-confirmed*, and *relief* verbally and they are accompanied by facial expressions of primary emotions.

A final empirical study was conducted to falsify the hypothesis derived from developmental psychology that MAX only expressing primary emotions would be judged younger than MAX additionally expressing secondary emotions within the card game scenario. The three sided large screen projection system and the sophisticated sensor technology provided the opportunity to realize very natural and realistic human-computer interaction for this study.

The questionnaire-based results of this study confirmed the initial hypothesis that MAX with secondary emotion simulation is judged significantly older than without.

7.2 Discussion and future perspectives

In summary, the results of both empirical studies support the assumption that the simulation of affect achieved by the WASABI architecture increases the believability of the virtual human MAX. As the WASABI architecture combines findings and theoretical conceptions of different disciplines in a novel and creative way, also the second objective of this thesis is fulfilled.

Of course, the proposed Affect Simulation Architecture can still be refined and some limitations are still waiting to be solved. The most important ones are discussed next.

Direct expression of primary emotions

As already critically observed in the end of Chapter 5, letting primary emotions directly drive MAX's facial expressions might be considered inappropriate for MAX outer appearance resembling an adult human. As discussed in Chapter 2, adults acquire the ability to short-cut their bodily feedback during ontogenesis. This ability to hide emotional expressions could easily be achieved in MAX's cognitive architecture, but as it adds another layer of complexity it was deliberately not done so far.

The simultaneous experience of opposing emotions

The mood congruency of all elicited emotions is always assured by the emotion dynamics component of the WASABI architecture. This entails, however, that some plausible mixtures of emotions, such as fear and joy occurring at the same time, e.g., when taking a joy-ride in a roller coaster, are impossible within the architecture. Although these mixed feelings might occur much less frequently in everyday life (Larsen, McGraw & Cacioppo 2001) the WASABI architecture might need to be refined to also cover these special emotional episodes.

Integrating further emotions

The simulation of secondary emotions is exemplified in this thesis by integrating three prospect-based emotions into the WASABI architecture and the integration of two more prospect-based emotions is outlined. The simulation of other secondary or even tertiary (or social) emotions could be achieved as well and provides a challenging goal for future work.

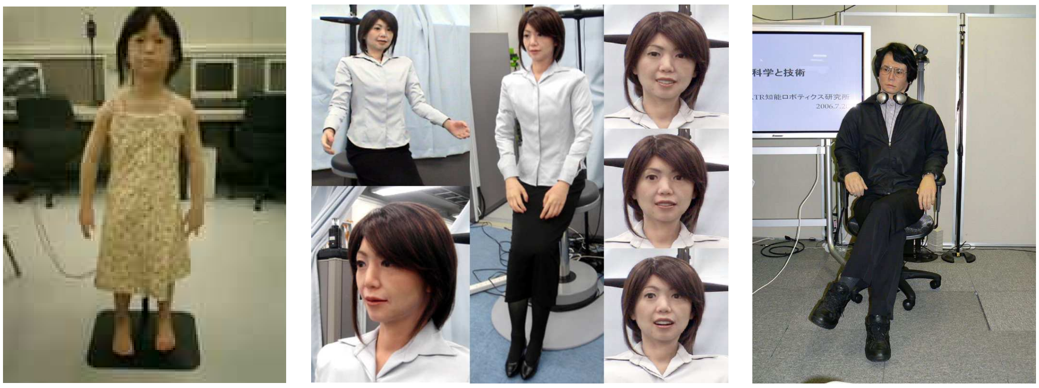
Simulation of further effects of emotions on cognitive processes

So far, the cognition module of the architecture only reappraises the aware emotions letting MAX perform different coping behaviors. Research in psychology, however, suggests to also model a lower level influence of affective states on cognition. As described in (Becker et al. 2006) emotions could also function as modulators of cognitive processes by constraining the action selection of the BDI-interpreter or systematically changing the problem-solving process: negative emotions seem to lead to a narrowed problem-solving, while positive emotions lead to broader problem-solving attempts to achieve multiple goals simultaneously (Sloman 1987).

From virtual to physical agents

The virtual human MAX enables us to study a form of human-computer interaction that is already very similar to human-human interaction—with one important difference: MAX is not able to manipulate the physical world. Accordingly, a human interlocutor needs not fear to be physically harmed by an angry MAX. Reconsidering the results of the first empirical study the relatively lower stress levels of those human players, that played against a negatively empathic MAX, might as well result from their amusement about such funny animations like MAX being afraid to lose the game or him expressing his anger.

In the aim to find an answer to the last question the author plans to apply the WASABI architecture to physical robots next. As it would be challenging to compare the results of the second empirical study, in which MAX was presented three dimensionally and in life-size, with results attained from experiments with physical robots, these robots should posses a comparable level of anthropomorphism.



(a) Repliee R1

(b) Repliee Q2

(c) Geminoid HI-1

Figure 7.2: The different humanoid robots of ATR

The sophisticated androids (cf. Figure 7.2) under development at the Advanced Telecommunication Research Institute International (ATR) in Japan provide the necessary similarity and Ishiguro (2005) suggests to use the term “Android science” to refer to a new “cross-interdisciplinary framework” that tries to “tackle the problem of appearance and behavior”. In short, the more human-like a robot’s outer appearance is designed the higher are the human’s expectations concerning the naturalness of the robot’s behavior.

The high level of anthropomorphism achieved by such androids as Geminoid HI-1 (Sakamoto, Kanda, Ono, Ishiguro & Hagita 2007, cf. Figure 7.2(c)) and Repliee Q2 (Minato, Shimada, Itakura, Lee & Ishiguro 2005, cf. Figure 7.2(b)) affords an increase of autonomy in the generation of social and emotional cues. The WASABI architecture presented in this thesis might help to achieve this higher level of autonomy.

The case of “love”

Before we can reasonably answer the question, whether humans will ever fall in love with virtual or robotic agents, more theoretical and applied research on affective phenomena has to be undertaken to better understand the social emotion “love”. The class of social emotions is the most complex one and can hardly be explained in logical terms alone, because its experience involves a variety of bodily and mental fluctuations and heavily relies on an individuals social context and personal experiences.

The WASABI architecture is well-suited to help in understanding, how the dynamic interplay of a human’s body and mind together with his past experiences and future expectations sometimes turns “cold” cognitions into “hot” emotions.

Bibliography

André, E., Klesen, M., Gebhard, P., Allen, S., & Rist, T. (1999). Integrating models of personality and emotions into lifelike characters. In *Proceedings International Workshop on Affect in Interactions - Towards a New Generation of Interfaces*, (pp. 136–149).

Asimov, I. (1942). Runaround. Street & Smith. Science fiction short story.

Asimov, I., Silverberg, R., & Kazan, N. (1999). Bicentennial man (movie). Columbia Pictures. (based on a short story by Isaac Asimov).

Bates, J. & Reilly, W. S. (1992). Building emotional agents. Technical report, School of Computer Science.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. (2005). The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers. *TRENDS in Cognitive Sciences*, *9*, 159–162.

Becker, C. (2003). Simulation der Emotionsdynamik eines künstlichen humanoiden Agenten. Master's thesis, University of Bielefeld.

Becker, C., Kopp, S., & Wachsmuth, I. (2004). Simulating the emotion dynamics of a multimodal conversational agent. In *Workshop on Affective Dialogue Systems*, LNAI 3068, (pp. 154–165). Springer.

Becker, C., Kopp, S., & Wachsmuth, I. (2007). Why emotions should be integrated into conversational agents. In T. Nishida (Ed.), *Conversational Informatics: An Engineering Approach* chapter 3, (pp. 49–68). Wiley.

Becker, C., Leßmann, N., Kopp, S., & Wachsmuth, I. (2006). Connecting feelings and thoughts - modeling the interaction of emotion and cognition in embodied agents. In *Proceedings of the Seventh International Conference on Cognitive Modeling (ICCM-06)*, (pp. 32–37). Edizioni Goliardiche.

Becker, C., Nakasone, A., Prendinger, H., Ishizuka, M., & Wachsmuth, I. (2005). Physiologically interactive gaming with the 3D agent Max. In *International Workshop on Conversational Informatics, in conj. with JSAI-05*, (pp. 37–42)., Kitakyushu, Japan.

Becker, C., Prendinger, H., Ishizuka, M., & Wachsmuth, I. (2005a). Empathy for Max (Preliminary project report). In *The 2005 International Conference on Active Media Technology (AMT-05)*, (pp. 541–545).

Becker, C., Prendinger, H., Ishizuka, M., & Wachsmuth, I. (2005b). Evaluating Affective Feedback of the 3D Agent Max in a Competitive Cards Game. In *Affective Computing and Intelligent Interaction*, LNCS 3784, (pp. 466–473). Springer.

Becker, C. & Wachsmuth, I. (2006a). Modeling primary and secondary emotions for a believable communication agent. In Reichardt, D., Levi, P., & Meyer, J.-J. C. (Eds.), *Proceedings of the 1st Workshop on Emotion and Computing*, (pp. 31–34)., Bremen.

Becker, C. & Wachsmuth, I. (2006b). Playing the Cards Game SkipBo against an Emotional Max. In Reichardt, D., Levi, P., & Meyer, J.-J. C. (Eds.), *Proceedings of the 1st Workshop on Emotion and Computing*, (pp.6̃5)., Bremen.

Becker-Asano, C., Kopp, S., Pfeiffer-Leßmann, N., & Wachsmuth, I. (2008). Virtual Humans Growing up: From Primary Toward Secondary Emotions. *KI Zeitschift (German Journal of Artificial Intelligence)*, *1*, 23–27.

Beckermann, A. (2001). *Analytische Einführung in die Philosophie des Geistes*. de Gruyter.

Berry, D., Butler, L., de Rosis, F., Laaksolahti, J., Pelachaud, C., & Steedman, M. (2004). MagiCster—Final Evaluation Report. Technical report, DFKI.

Berry, D. C., Butler, L. T., & de Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human-Computer Studies*, *63*, 304–327.

Bickmore, T. & Cassell, J. (2005). Social dialogue with embodied conversational agents. In J. van Kuppevel, L. Dybkjaer, & N. Bernsen (Eds.), *Advances in Natural, Multimodal Dialogue Systems*. New York: Kluwer Academic.

Bouchard, T. J. & Loehlin, J. C. (2001). Genes, evolution, and personality. *Behavior Genetics*, *31*(3), 243–273.

Boukricha, H., Becker, C., & Wachsmuth, I. (2007). Simulating empathy for the virtual human max. In *2nd International Workshop on Emotion and Computing, in conj. with the German Conference on Artificial Intelligence (KI2007)*, (pp. 22–27).

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Havard University Press.

Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of human-computer studies*, *62*, 162–178.

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*, 119–155.

Breazeal, C. & Velásquez, J. (1998). Toward teaching a robot ”infant” using emotive communication acts. In *Proceedings of Workshop on Socially Situated Intelligence*, (pp. 25–40).

Breazeal, C. L. (2002). *Designing Sociable Robots*. Cambridge, Massachusetts: The MIT Press.

Buchanan, B. G. (2005). A (very) brief history of Artificial Intelligence. *AI Magazine*, *26*(4), 53–60.

Burghouts, G., op den Akker, R., Heylen, D., Poel, M., & Nijholt, A. (2003). An action selection architecture for an emotional agent. In Russell, I. & Haller, S. (Eds.), *Recent Advances in Artificial Intelligence*, (pp. 293–297). AAAI Press.

Cannon, W. B. (1927). The James-Lange theory of emotion: A critical examination and an alternative theory. *American Journal of Psychology*, *39*, 106–124.

Cassell, J. (2000a). More than just another pretty face: Embodied conversational interface agents. *Communications of the ACM*, *43*(4), 70–78.

Cassell, J. (2000b). Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied Conversational Agents* (pp. 1–27). Cambridge, MA: The MIT Press.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjailmsson, H., & Yan, H. (1999). Embodiment in conversational interfaces: Rea. In *CHI 99*.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). *Embodied Conversational Agents*. Cambridge, MA: The MIT Press.

Damasio, A. (1994). *Descartes' Error, Emotion Reason and the Human Brain.* Grosset/Putnam.

Damasio, A. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Harcourt.

Darwin, C. (1898). *The expression of the emotions in man and animals*. New York: D. Appleton and company. (received electronically from Electronic Text Center, University of Virginia Library, first publihsed in 1872).

Dautenhahn, K., Nourbakhsh, I., & Fong, T. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*, 143–166.

de Rosis, F., Matheson, C., Pelachaud, C., & Rist, T. (2003). MagiCster: Believable Agents and Dialogue. *Künstliche Intelligenz*, *4*, 24–29.

de Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., & de Carolis, B. (2003). From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies. Special Issue on "Applications of Affective Computing in HCI"*, *59*, 81–118.

Dehn, D. M. & van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, *52*(52), 1–22.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*, 177–190.

Duffy, B. R., Dragone, M., & O'Hare, G. M. P. (2005). The Social Robot Architecture: A Framework for Explicit Social Interaction. In *Toward Social Mechanisms of Android Science*, (pp. 18–28).

Dunn, B. D., Dalgleish, T., & Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience and Biobehavioral Reviews*, *30*, 239–271.

Egges, A. (2006). *Real-time Animation of Interactive Virtual Humans*. PhD thesis, Université de Genéve.

Egges, A., Kshirsagar, S., & Magnenat-Thalmann, N. (2003). A model for personality and emotion simulation. In *Knowledge-Based Intelligent Information & Engineering Systems (KES2003)*, (pp. 453–461).

Egges, A., Kshirsagar, S., & Magnenat-Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents. In *Computer Animation and Virtual Worlds*, (pp. 1–13).

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200.

Ekman, P. (1994). Strong evidence for universals in facial expression: A reply to Russell's mistaken critique. *Psychological Bulletin*, *115*(2), 268–287.

Ekman, P. (1999a). Basic emotions. In *Handbook of Cognition and Emotion* chapter 3, (pp. 45–60). John Wiley & Sons.

Ekman, P. (1999b). Facial expressions. In *Handbook of Cognition and Emotion* chapter 16, (pp. 301–320). John Wiley & Sons.

Ekman, P., Friesen, W., & Ancoli, S. (1980). Facial sings of emotional experience. *Journal of Personality and Social Psychology*, *29*, 1125–1134.

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

El-Nasr, M. S., Ioerger, T. R., & Yen, J. (1999). PETEEI: A PET with evolving emotional intelligence. In *Autonomous Agents*.

El-Nasr, M. S., Yen, J., & Ioerger, T. R. (2000). FLAME - Fuzzy Logic Adaptive Model of Emotions. *Autonomous Agents and Multi-Agent Systems*, *3*(3), 219–257.

Elliott, C. (1992). *The Affective Reasoner. A process model of emotions in a multi-agent system*. PhD thesis, Northwestern University. Institute for the Learning Sciences.

Elliott, C. (1994). Research problems in the use of a shallow artificial intelligence model of personality and emotion. In *Proceedings of the twelfth national conference on Artificial intelligence*, volume 1, (pp. 9–15)., Seattle, Washington, United States.

Elliott, C., Rickel, J., & Lester, J. C. (1997). Integrating affective computing into animated tutoring agents. In *IJCAI Workshop on Animated Interface Agents*.

Ellsworth, P. & Scherer, K. R. (2003). Appraisal processes in emotion. In *Handbook of affective sciences* (pp. 572–595). New York: Oxford University Press.

Ernst, G. & Newell, A. (1969). *GPS: A Case Study in Generality and Problem Solving*. Academic Press.

Fellous, J.-M. (2004). From human emotions to robot emotions. In Hudlicka, E. & Canamero, L. (Eds.), *AAAI Spring 2004 symposium on Architectures for Modeling Emotion: Cross-Disciplinary Foundations*.

Freeman, P. A. (1995). Effective computing science. *ACM Computing Surveys*, *27*, 27–29.

Gebhard, P. (2005). ALMA - A Layered Model of Affect. In *Autonomous Agents & Multi Agent Systems*, (pp. 29–36).

Gebhard, P. & Kipp, K. H. (2006). Are computer-generated emotions and moods plausible to humans? In *IVA*, (pp. 343–356).

Gebhard, P., Klesen, M., & Rist, T. (2004). Coloring multi-character conversations through the expression of emotions. In *Proceedings of the Tutorial and Research Workshop on Affective Dialogue Systems (ADS'04)*, (pp. 128–141).

Gehm, T. L. & Scherer, K. R. (1988). Factors determining the dimensions of subjective emotional space. In K. R. Scherer (Ed.), *Facets of Emotion* chapter 5. Lawrence Erlbaum Associates.

Gesellensetter, L. (2004). Planbasiertes Dialogsystem für einen multimodalen Agenten mit Präsentationsfähigkeit. Master's thesis, University of Bielefeld.

Giarratano, J. & Riley, G. (2005). *Expert systems: principles and programming* (4th ed.). Thomson course technology.

Gratch, J. (1999). Why you should buy an emotional planner. In *Proceedings of the Agents'99 Workshop and Emotion-based Agent Architectures (EBAA'99)*.

Gratch, J. (2000). Émile: Marshalling passions in training and education. In *Proceedings 4th International Conference on Autonomous Agents (Agents'2000)*, (pp. 325–332)., New York. ACM Press.

Gratch, J. & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Science Research*, *5*, 269–306.

Gratch, J., Rickel, J., André, E., Cassell, J., Petajan, E., & Badler, N. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, *17*, 54–63.

Griffiths, P. (2002). Basic emotions, complex emotions, machiavellian emotions. Available at philsci-archive.pitt.edu/archive/00000604/.

Hodges, A. (2000). *Alan Turing: The Enigma*. Walker & Company.

Hollnagel, E. (2003). Is affective computing an oxymoron? *International Journal of Human-Computer Studies*, *59*, 65–70.

Holodynski, M. & Friedlmeier, W. (2005). *Development of Emotions and Emotion Regulation*. Springer.

Huber, M. J. (1999). JAM: A BDI-theoretic mobile agent architecture. In *Proc. Int. Conf. on Autonomous Agents*, (pp. 236–243)., Seattle, WA.

Hudlicka, E. (2003a). Response: Is affective computing an oxymoron? *International Journal of Human-Computer Studies*, *59*, 71–75.

Hudlicka, E. (2003b). To feel or not to feel: The role of affect in human-computer interaction. *International Journal of Human-Computer Studies*, *59*, 1–32.

Ishiguro, H. (2005). Android science: Toward a new cross-interdisciplinary framework. In *Proceedings of the CogSci 2005 Workshop "Toward Social Mechanisms of Android Science"*.

Itoh, K., Miwa, H., Nukariya, Y., Zecca, M., Takanobu, H., Roccella, S., Carrozza, M., Dario, P., & Takanishi, A. (2006). Behavior generation of humanoid robots depending on mood. In *9th International Conference on Intelligent Autonomous Systems (IAS-9)*, (pp. 965–972).

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

Johnson, W. L. & Rickel, J. (1997). Steve: An animated pedagogical agent for procedural training in virtual environments. *ACM SIGART Bulletin*, *8*, 16–21.

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, *11*, 47–78.

Kopp, S. (2003). *Synthese und Koordination von Sprache und Gestik für Virtuelle Multimodale Agenten*. PhD thesis, University of Bielefeld.

Kopp, S., Becker, C., & Wachsmuth, I. (2006). The Virtual Human Max - Modeling Embodied Conversation. In *KI 2006 - Demo Presentation, Extended Abstracts*, (pp. 21–24).

Kopp, S., Gesellensetter, L., Krämer, N., & Wachsmuth, I. (2005). A conversational agent as museum guide—design and evaluation of a real-world application. In *Intelligent Virtual Agents*, LNAI 3661, (pp. 329–343). Springer.

Kopp, S., Jung, B., Leßmann, N., & Wachsmuth, I. (2003). Max - A Multimodal Assistant in Virtual Reality Construction. *KI-Künstliche Intelligenz Special Issue on Embodied Conversational Agents*, *4/03*, 11–17.

Kopp, S. & Wachsmuth, I. (2000). A knowledge-based approach for lifelike gesture animation. In *Proceedings of the 14th ECAI 2000*.

Kopp, S. & Wachsmuth, I. (2002). Model-based Animation of Coverbal Gesture. In *Proceedings of Computer Animation 2002*, (pp. 252–257). IEEE Press.

Kopp, S. & Wachsmuth, I. (2004). Synthesizing multimodal utterances for conversational agents. *Computer Animation and Virtual Worlds*, *15*(1), 39–52.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, *50*(5), 372–385.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International affective picture system (IAPS): Instruction manual and affective ratings. Technical report, University of Florida, Center for Research in Psychophysiology, Gainesville, FL.

Lange, C. G. (1885). *Om Sindsbevoegelser: Et psykofysiologiske Studie*. Kopenhagen: Kronar. (deutsch 1887: Ueber Gemüthsbewegungen. Leipzig: Theodor Thomas).

Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology*, *81*(4), 684–696.

Latoschik, M. E., Biermann, P., & Wachsmuth, I. (2005). Knowledge in the loop: Semantics representation for multimodal simulative environments. In *Proceedings of the 5th International Symposium on Smart Graphics*, (pp. 25–39).

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

LeDoux, J. E. (1995). EMOTION: Clues from the Brain. *Annual Reviews of Psychology*, *46*, 209–235.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155–184.

Leßmann, N. (2002). Eine kognitiv motivierte Architektur für einen situierten Agenten. Master's thesis, University of Bielefeld.

Leßmann, N., Kopp, S., & Wachsmuth, I. (2006). Situated interaction with a virtual human - perception, action, and cognition. In G. Rickheit & I. Wachsmuth (Eds.), *Situated Communication* (pp. 287–323). Berlin: Mouton de Gruyter.

Leßmann, N., Kranstedt, A., & Wachsmuth, I. (2004). Toward a Cognitively Motivated Processing of Turn-Taking Signals for the Embodied Conversational Agent Max. In *AAMAS 2004 Workshop Proceedings*, (pp. 57–64).

Levenson, R. W. (1988). Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity. In H. L. Wagner (Ed.), *Social Psychophysiology and Emotion: Theory and Clinical Applications* (pp. 17–42). Hoboken, NJ: John Wiley & Sons.

Leventhal, H. & Scherer, K. R. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion*, *1*, 3–28.

Levy, D. (2007). *Intimate relationships with Artificial Partners*. PhD thesis, Maastricht University.

Lindblom, J. & Ziemke, T. (2003). Social situatedness of natural and artificial intelligence: Vygotsky and beyond. *Adaptive Behavior*, *11*, 79–96.

MacLean (1949). Psychosomatic disease and the “visceral brain”: recent developments bearing on the Papez theory of emotion. *Psychosomatic Medicine, 11,* 338–53.

MacLean (1970). The triune brain, emotion and scientific bias. In *The neurosciences; Second study program* (pp. 336–49). New York: Rockefeller University Press.

Maia, T. V. & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task. *PNAS, 101*(45), 16075–16080.

Marinier, R. & Laird, J. (2004). Toward a comprehensive computational model of emotions and feelings. In *International Conference on Cognitive Modeling*.

Marinier, R. P. & Laird, J. E. (2006). A cognitive architecture theory of comprehension and appraisal. In *Agent Construction and Emotion*.

Marinier, R. P. & Laird, J. E. (2007). Computational modeling of mood and feeling from emotion. In *CogSci*, (pp. 461–466).

Marsella, S. & Gratch, J. (2006). EMA: A computational model of appraisal dynamics. In *European Meeting on Cybernetics and Systems Research*.

McCrae, R. R. & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*, 175–215.

McDougall, W. (2001, 1919). *An Introduction to Social Psychology* (Online ed.). Kitchener: Batoche Books.

McIntosh, D. N. (1996). Facial feedback hypotheses: Evidence, implications, and directions. *Motivation and Emotion*, *20*(2), 121–147.

Meyer, W.-U., Reisenzein, R., & Schützwohl, A. (2001). *Einführung in die Emotionspsychologie, Band I: Die Emotionstheorien von Watson, James und Schachter* (2nd ed.). Verlag Hans Huber.

Meyer, W.-U., Schützwohl, A., & Reisenzein, R. (2003). *Einführung in die Emotionspsychologie, Band II: Evolutionspsychologische Emotionstheorien* (3rd ed.). Verlag Hans Huber.

Mikulas, W. L. & Vodanovich, S. J. (1993). The essence of boredom. *The Psychological Record*, *43*, 3–12.

Minato, T., Shimada, M., Itakura, S., Lee, K., & Ishiguro, H. (2005). Does gaze reveal the human likeness of an android? In *Proceedings of the 4th International Conference on Development and Learning*, (pp. 106–111).

Miwa, H., Itoh, K., Takanobu, H., & Takanishi, A. (2004). Development of mental model for humanoid robots. In *Proceedings of the 15th CISM-IFToMM Symposium on Robot Design, Dynamics and Control*.

Netica (2003). Norsys Software Corp. http://www.norsys.com.

Neumann, R., Seibt, B., & Strack, F. (2001). The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. *Cognition & Emotion*, *15*, 725–747.

Newell, A. (1982). The knowledge level. *Artificial Intelligence*, *18*, 87–127.

Newell, A. & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, *9*(3), 184–211.

Oatley, K. & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, *1*, 29–50.

Ochs, M., Devooght, K., Sadek, D., & Pelachaud, C. (2006). A computational model of capability-based emotion elicitation for rational agent. In *Workshop Emotion and Computing - German Conference on Artificial Intelligence (KI)*.

Ochs, M., Niewiadomski, R., Pelachaud, C., & Sadek, D. (2005). Intelligent expressions of emotions. In *1st International Conference on Affective Computing and Intelligent Interaction ACII*.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

Ortony, A., Norman, D., & Revelle, W. (2005). Affect and proto-affect in effective functioning. In J. Fellous & M. Arbib (Eds.), *Who needs emotions: The brain meets the machine* (pp. 173–202). Oxford University Press.

Ortony, A. & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315–331.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. University of Illinois Press.

Paiva, A., Dias, J., Sobral, D., & Aylett, R. (2004). Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-03)*, New York. ACM Press.

Parkinson, B., Fischer, A. H., & Manstead, A. S. (2005). *Emotion in Social Relations*. New York: Psychology Press.

Pelachaud, C. & Bilvi, M. (2003). Computational model of believable conversational agents. In M.-P. Huget (Ed.), *Communication in MAS: background, current trends and future*. Springer-Verlag. To appear.

Pelachaud, C., Peters, C., Bevacqua, E., & Chafaï, N. E. (2008). Webpage of the GRETA Group: Embodied Conversational Agent Research at the IUT Montreuil. online.

Pelachaud, C. & Poggi, I. (2002). Multimodal embodied agents. *The Knowledge Engineering Review, 17*, 181–196.

Pfeifer, R. (2001). Embodied artificial intelligence 10 years back, 10 years forward. *Informatics. 10 Years Back. 10 Years Ahead, LNCS 2000*, 294–310.

Picard, R. W. (1997). *Affective Computing*. Cambridge, MA: The MIT Press.

Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(10), 1175–1191.

Plutchik, R. (1980). *Emotion. A psychoevolutionay synthesis*. New York: Harper & Row.

Prendinger, H., Becker, C., & Ishizuka, M. (2006). A study in users' physiological response to an empathic interface agent. *International Journal of Humanoid Robotics, 3*(3), 371–391.

Prendinger, H., Descamps, S., & Ishizuka, M. (2002). Scripting affective communication with life-like characters in web-based interaction systems. *Applied Artificial Intelligence, 16*(7–8), 519–553.

Prendinger, H., Dohi, H., Wang, H., Mayer, S., & Ishizuka, M. (2004). Empathic embodied interfaces: Addressing users' affective state. In *Workshop on Affective Dialogue Systems*, (pp. 53–64).

Prendinger, H. & Ishizuka, M. (2001a). Let's talk! Socially intelligent agents for language conversation training. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 31*(5), 465–471.

Prendinger, H. & Ishizuka, M. (2001b). Social role awareness in animated agents. In *Proceedings 5th International Conference on Autonomous Agents (Agents-01)*, (pp. 270–277)., New York. ACM Press.

Prendinger, H. & Ishizuka, M. (2002). SCREAM: SCRipting Emotion-based Agent Minds. In *Proceedings First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-02)*, (pp. 350–351)., New York. ACM Press.

Prendinger, H. & Ishizuka, M. (2005). The Empathic Companion: A character-based interface that addresses users' affective states. *International Journal of Applied Artificial Intelligence, 19*(3), 267–285.

Prendinger, H., Mori, J., & Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human-Computer Studies, 62*(2), 231–245.

Prendinger, H., Saeyor, S., & Ishizuka, M. (2003). MPML and SCREAM: Scripting the bodies and minds of life-like characters. In H. Prendinger & M. Ishizuka (Eds.), *Life-Like Characters. Tools, Affective Functions and Applications*. Springer. This volume.

Rao, A. & Georgeff, M. (1991). Modeling Rational Agents within a BDI-architecture. In Allen, J., Fikes, R., & Sandewall, E. (Eds.), *Proc. of the Intl. Conference on Principles of Knowledge Representation and Planning*, (pp. 473–484). Morgan Kaufmann publishers Inc.: San Mateo, CA, USA.

Reeves, B. & Nass, C. (1998). *The Media Equation. How People Treat Computers, Television and New Media Like Real People and Places*. CSLI Publications, Center for the Study of Language and Information. Cambridge University Press.

Rehm, M. & André, E. (2005). Catch me if you can – exploring lying agents in social settings. In *Autonomous Agents and Multiagent Systems*, (pp. 937–944).

Reilly, W. S. N. (1996). *Believable Social and Emotional Agents*. PhD thesis, Carnegie Mellon University. CMU-CS-96-138.

Reisenzein, R. (1992). A structuralist reconstruction of wundt's threedimensional theory of emotion. In *The structuralist program in psychology: Foundations and applications* (pp. 141–189). Toronto, Ontario, Canada: Hopgrefe & Huber.

Reisenzein, R. (1994). Pleasure-arousal theory and the Intensity of Emotions. *Journal of Personality and Social Psychology, 67*, 525–39.

Reisenzein, R. (2000a). Worum geht es in der Debatte um Basisemotionen? In *Kognitive und motivationale Aspekte der Motivation*. Göttingen: Hogrefe.

Reisenzein, R. (2000b). Wundt's three-dimensional theory of emotion. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, *75*, 219–250.

Reithinger, N., Gebhard, P., L"ockelt, M., Ndiaye, A., Pfleger, N., & Klesen, M. (2006). Virtualhuman—dialogic and affective interaction with virtual characters. In *ICMI'06*, (pp. 51–58)., Banff, Alberta, Canada.

Russell, J. (1980). A cirsumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Russell, J. & Mehrabian, A. (1974). Distinguishing anger and anxiety in terms of emotional response factors. *Journal of Consulting and Clinical Psychology, 42*(1), 79–83.

Russell, J. & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*(11), 273–294.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.

Russell, J. A. & Feldmann Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819.

Russell, S. & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach* (2nd ed.). Prentice Hall.

Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., & Hagita, N. (2007). Android as a telecommunication medium with a human-like presence. In *HRI*, (pp. 193–200).

Scherer, K. R. (1984). On the Nature and Function of Emotion: A Component Process Approach. In K. Scherer & P. Ekman (Eds.), *Approaches to Emotion* (pp. 293–317). N.J.: Lawrence Erlbaum.

Scherer, K. R. (1999). Appraisal theory. In *Handbook of Cognition and Emotion*. John Wiley & Sons.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal Processes in Emotion* chapter 5. Oxford University Press.

Scherer, K. R. (2005). Unconscious processes in emotion: The bulk of the iceberg. In P. Niedenthal, L. Feldman Barrett, & P. Winkielman (Eds.), *The unconscious in emotion*. New York: Guilford Press.

Scherer, K. R., Dan, E., & Flykt, A. (2006). What determines a feeling's position in affective space? A case for appraisal. *Cognition and Emotion*, *20*, 92–113.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review*, *61*(2), 81–88.

Schwarz, M. W., Cowan, W. B., & Beatty, J. C. (1987). An Experimental Comparison of RGB, YIQ, LAB, HSV, and Opponent Color Models. *ACM Transactions on Graphics*, *6*(2), 123–158.

Shortliffe, E. H., Rhame, F. S., Axline, S. G., Cohen, S. N., Buchanan, B. G., Davis, R., Scott, A. C., Chavez-Pardo, R., & van Melle, W. J. (1975). Mycin: A computer program providing antimicrobial therapy recommendations. *Clinical Medicine*, *34*.

Sietz, H. (2004). I, robot (movie). 20th Century Fox. (based on the short story collection by Isaac Asimov).

Sloman, A. (1987). Motives, mechanisms and emotions. *Cognition and Emotion*, *1*, 217–234.

Sloman, A. (1992). Prolegomena to a theory of communication and affect. In *Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues*. Springer.

Sloman, A. (1998). Damasio, Descartes, Alarms and Meta-management. *Systems, Man, and Cybernetics*, *3*, 2652–2657.

Sloman, A. (2000). Architectural Requirements for Human-like Agents Both Natural and Artificial. (What sorts of machines can love?). In K. Dautenhahn (Ed.), *Human Cognition and Social Agent Technology* chapter 7, (pp. 163–195). Amsterdam: John Benjamins.

Sloman, A., Chrisley, R., & Scheutz, M. (2005). The architectural basis of affective states and processes. In *Who needs emotions?* Oxford University Press.

Sloman, A. & Croucher, M. (1981). Why robots will have emotions. In *In Proceedings IJCAI*, Vancouver.

Spielberg, S. (2001). A.I. (movie). Warner Bros. Pictures. (partly based on a short story by Brian Aldiss).

Staller, A. & Petta, P. (1998). Towards a tractable appraisal-based architecture for situated cognizers. In *Grounding Emotions in Adaptive Systems,Workshop Notes, Fifth Int. Conf. of the Society for Adaptive Behaviour (SAB98)*.

Strack, R., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of facial expressions: A non-obtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, *54*, 768–777.

Svennevig, J. (1999). *Getting Acquainted in Conversation*. John Benjamins.

Svennevig, J. & Xie, C. (2002). Book Reviews - Getting Acquainted in Conversation. A Study of Initial Interactions. *Studies in Language*, *26*, 194–201.

Tanguy, E. (2006). *Emotions: the Art of Communication Applied to Virtual Actors*. PhD thesis, University of Bath.

Tanguy, E., Willis, P., & Bryson, J. J. (2003). A layered dynamic emotion representation for the creation of complex facial animation. In Rist, T., Aylett, R., Ballin, D., & Rickel, J. (Eds.), *Intelligent Virtual Agents*, (pp. 101–105). Springer.

Tanguy, E., Willis, P. J., & Bryson, J. J. (2006). A dynamic emotion representation model within a facial animation system. *International Journal of Humanoid Robotics*, *3*(3), 293–300.

Thayer, R. E. (1996). *The origin of everyday moods*. Oxford University Press.

ThoughtTechnology (2003). Thought Technology Ltd. http://www.thoughttechnology.com.

Trolltech (2008). Qt: Cross-Platform Rich Client Development Framework. http://trolltech.com/products/qt/.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460.

Velásquez, J. D. (1997). Modeling emotions and other motivations in synthetic agents. In *Proceedings 14th National Conference on Artificial Intelligence (AAAI-97)*, (pp. 10–15).

Velásquez, J. D. (1998). When robots weep: Emotional memories and decision-making. In *Proceedings 15th National Conference on Artificial Intelligence (AAAI-98)*, (pp. 70–75).

Velásquez, J. D. & Maes, P. (1997). Cathexis: A computational model of emotions. In *Proceedings of the first International Conference on Autonomous agents*, (pp. 518–519).

Vinayagamoorthy, V., Gillies, M., Steed, A., Tanguy, E., Pan, X., Loscos, C., & Slater, M. (2006). Building expression into virtual characters: State of the art report. Technical report, EUROGRAPHICS 2006.

Wachsmuth, I. (2000). The Concept of Intelligence in AI. *Prerational Intelligence - Adaptive Behavior and Intelligent Systems without Symbols and Logic*, *1*, 43–55.

Wachsmuth, I. & Cao, Y. (1995). Interactive graphics design with situated agents. In *Graphics and Robotics* (pp. 73–85). Springer.

Wachsmuth, I., Lenzmann, B., Jörding, T., Jung, B., Latoschik, M., & Fröhlich, M. (1997). A Virtual Interface Agent und its Agency. In *Proceedings of the First International Conference on Autonomous Agents*.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548–573.

Weizenbaum, J. (1976). *Computer power and human reason: From judgement to calculation*. Freeman.

Wikipedia (2008). HSV-Farbraum. online.

Winston, P. (1992). *Artificial Intelligence* (3rd ed.). Addison-Wesley.

Wundt, W. (1922/1863). *Vorlesung über die Menschen- und Tierseele*. Leipzig: Voss Verlag.

Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175.

Zecca, M., Roccella, S., Carrozza, M., Miwa, H., Itoh, K., Cappiello, G., Cabibihan, J.-J., Matsumoto, M., Takanobu, H., Dario, P., & Takanishi, A. (2004). On the development of the emotion expression humanoid robot WE-4RII with RCH-1. *4th IEEE/RAS International Conference on Humanoid Robots*, *1*, 235–252.

Zinck, A. & Newen, A. (2007). Classifying emotion: A developmental account. *Synthese*, *161*(1), 1–25.

A Instructions for the card game Skip-Bo

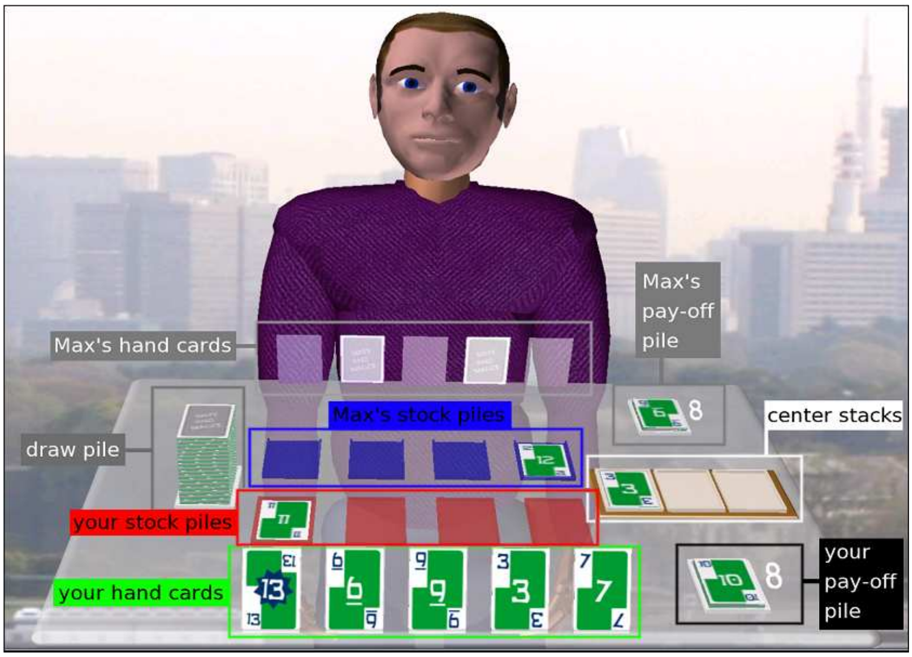


Figure A.1: The card game “Skip-Bo” as an interaction scenario for an Empathic Max.

Purpose of the game

Both players try to be the first to get rid of a pile of “pay-off cards” by playing them to center stacks which are begun with a one and continue in upward sequence to a twelve. The players take alternate turns and they play with nine series of cards each ranging from 1 to 13, which makes a total of 117 cards. **Thirteens are wild cards (jokers) and may be played in place of any card you desire.**

At the beginning you will be dealt five cards to your hand (so called hand cards) which are placed at the bottom of the screen facing to you (see above in the screenshot). These cards are only visible to you. Then eight cards are dealt to make up the pay-off piles which are placed to the right side of the table. Only the top cards of the two piles are face up and therefore visible to you as well as your opponent Max. You will go first.

The Play

The object of the game is to be the first to **get rid of all the cards in your pay-off pile** by playing them to the three white center stacks. The first card in each center stack must be a 1 (or a 13), then 2, 3, and so on in sequence up to 12, each card played being one higher than the card it covers.

It is also possible to play a card from your hand to a center stack or to one of your four red stock piles in front of you, or to move a card from one of your stock piles to a center stack. There is no restriction on the ranks of cards which can be played on the stock pile.

You may play as many cards to center stacks as you want in any order, but as soon as you move a card from your hand to one of your stock piles your turn ends, and Max takes a turn. You must play a card to a stock pile at the end of each turn.

If during your turn you manage to play all five cards from your hand, without playing to a stock pile, you immediately draw five more cards from the draw pile and continue playing. If in the beginning of your turn you have fewer than five cards in your hand, the required number of cards will be drawn from the draw pile to bring your hand up to five cards again.

If you complete a center stack by playing a twelve (or a thirteen representing a twelve) to the center, Max shuffles the completed stack into the draw pile, creating a space for a new center stack, and you can continue playing.

Summary

- Who is first to get rid of all cards in his pay-off pile wins the game.
- In the beginning of your turn the required number of cards will be drawn from the draw pile to bring your hand up to five cards automatically.
- You may play as many cards as you want from either your pay-off pile, your hand or one of your four red stock piles to any of the three white center piles as long as you follow the order of cards.
- Whenever you run out of hand cards without having played a card to one of your red stock piles you are immediately dealt five new hand cards.
- You finish your turn by playing one of your hand cards to one of your red stock piles.

Useful Strategies and Hints

(i) Always keep in mind the number of your current pay-off card!

(ii) You may try to keep an eye on the current pay-off card of Max. Sometimes it might be better not to play a card if this lets Max play his pay-off card afterwards.

(iii) You may try to get rid of your hand cards first.

(iv) When playing your last card to one of your red stock piles you may try to keep the following strategy in mind:

  - Always play high cards on empty stock piles.

- If there are cards on some stock piles already, you may try to play cards on top of them in descending order, e.g. an 11 on top of a 12 or a 7 on top of an 8.

Good luck!

B The questionnaire

Thank you for having played Skip-Bo against Max!

Please complete the following form:

(All personal data will only be used for statistical analysis)

How old are you?	_____		
Are you male or female?	○ male	○ female	
Have you been born in Japan?	○ yes	○ no	
How often do you play card games? (In real life or on the computer)	○ I never played a card game before	○ I casually play card games	○ I regularly play card games
1. Did you like to play this game?	No, I did not like it at all!	○ ○ ○ ○ ○ ○ ○	Yes, I really liked it!
2. How did you feel during the game?	Very sad!	○ ○ ○ ○ ○ ○ ○	Very happy!
3.Did you feel comfortable in this situation?	No, not at all!	○ ○ ○ ○ ○ ○ ○	Yes, completely!
4. Did you get enough instructions to play the game?	No, I would have needed more instructions!	○ ○ ○ ○ ○ ○ ○	Yes, I got absolutely sufficient instructions!
5. Did you feel alone during the game?	No, I did not feel alone!	○ ○ ○ ○ ○ ○ ○	Yes, I was feeling alone!
6. Did you feel criticized or praised during the game?	I felt like being criticized!	○ ○ ○ ○ ○ ○ ○	I felt like being praised!
7. MAX was..	very sympathetic.	○ ○ ○ ○ ○ ○ ○	very egoistic.
8. MAX behaved..	selfish.	○ ○ ○ ○ ○ ○ ○	unselfish.
9. MAX was..	friendly.	○ ○ ○ ○ ○ ○ ○	unfriendly.
10. MAX played..	competitive.	○ ○ ○ ○ ○ ○ ○	cooperative.
11. How do you think about MAX in general?	He is likable!	○ ○ ○ ○ ○ ○ ○	He is strange!
12. Did MAX behave naturally?	His behavior was very artificial!	○ ○ ○ ○ ○ ○ ○	His behavior was very natural!

13. Was MAX irritating you?	He was very irritating!	○ ○ ○ ○ ○ ○ ○	He was not irritating me at all!
14. Has MAX been trustworthy to you?	No, I did not trust him!	○ ○ ○ ○ ○ ○ ○	Yes, I always trusted him!
15. How honest was MAX to you?	He was hiding his true feelings!	○ ○ ○ ○ ○ ○ ○	He was showing his true feelings!
16. How much did you think of MAX as a human being during the game?	I was always aware that I just played against a computer program!	○ ○ ○ ○ ○ ○ ○	I always had the feeling of playing against another human being!
17. How intelligent did MAX play?	Very intelligent!	○ ○ ○ ○ ○ ○ ○	Very unintelligent!
17b. If MAX were a real human, how old would you judge him to be?		_____ (Only asked in the final empirical study described in Chapter 6.)	
18. What kind of person has MAX been?	MAX was a very dumb person!	○ ○ ○ ○ ○ ○ ○	MAX was a very smart person!
19. Was MAX capable of playing the game?	MAX was incapable of playing the game	○ ○ ○ ○ ○ ○ ○	MAX was very capable of playing the game!
20. How did MAX react within the game?	He was reacting very forceful!	○ ○ ○ ○ ○ ○ ○	He was reacting very considerate!
21. How did you judge the personality of MAX?	He is a very aggressive person!	○ ○ ○ ○ ○ ○ ○	He is a very suppliant person!
22. How emotional did MAX react?	He was very unemotional!	○ ○ ○ ○ ○ ○ ○	He was very emotional!
23. Did MAX care about your feelings?	He did not care about my feelings!	○ ○ ○ ○ ○ ○ ○	He really cared about my feelings!
24. Did you like the outward appearance of MAX?	No, you have you change that!	○ ○ ○ ○ ○ ○ ○	Yes, he looks good!
25. Would you like to play again?	No, get me outa here!	○ ○ ○ ○ ○ ○ ○	Yes, with 20 cards on the pay-off pile, please!

C Additional SkipBo Plans

Plan C.1 init SkipBo

```
Goal: PERFORM REACT-TO-INIT-GAME($numberOfSpecialcards$)
Body
    do initialization
    send *impulse 100*
    utter welcome-message
```

MAX performs **Plan C.1** whenever the *init-game* command is given by the interface. As the JAM-interpreter is integrated into a software agent that runs concurrently within our group's software framework (cf. Leßmann (2002)) this command might be given be message communication from any other software agent. Therefore, the command line interface of the visualization process is used here to give this command manually.

After the necessary initializations are done an emotional impulse of $+100$ is sent by the cognition module to the emotion module most probably resulting in a positive mood and happiness of MAX. At last MAX utters a welcome message greeting his opponent and encouraging him to play the first card[1].

Plan C.2 let max take a hand card

```
Goal: PERFORM TAKE-CARD
Body
    send *takeCard*
```

Plan C.3 react to new hand card

```
Conclude: PERFORM REACT-TO-HAND-CARD($cardID$)
Body
    if *max has five hand cards* then
        call think-skip-bo
    else
        call take-card
    end if
```

Plan C.2 lets MAX simply send the request to take a new hand card to the visualization agent. As soon as the visualization agent finishes with the necessary updates, it informs the JAM agent (i.e. the cognition module) of the new hand card automatically (see Plan C.3).

[1] "Willkommen in der AG Wissensbasierte Systeme! Bitte spielen Sie eine Karte."
(Welcome to the AI and VR lab. Please play a card.)

When new information about a hand card for MAX arrives in the cognition module, **Plan C.3** is triggered to either let MAX start *thinking* about how to play SkipBo (line 4) or take another card (line 6) if MAX still does not have enough cards on his hand.

Plan C.4 let MAX expect some proposition

```
Goal: PERFORM EXPECT($prop, value, valence$)
Body
    ASSERT $expect\ prop\ value\ valence$
```

Plan C.5 let MAX check a given proposition

```
Goal: PERFORM EXPECTED($prop, value$)
Body
    $answer \leftarrow false$
    if FACT $expect\ prop\ value\ valence$ then
        $answer \leftarrow true$
    end if
    return $(answer, valence)$
```

D Further initialization files for secondary emotions

Listing D.1: Initialization file `fears-confirmed.se`

```
polygon_begin QUAD
vertex -100 100 -100 1.0
vertex 0 100 -100 0
vertex 0 -100 -100 0
vertex -100 -100 -100 1.0
polygon_end
decayFunction LINEAR
lifetime 10.0
standardIntensity 0.0
type FEARS-CONFIRMED
tokens_begin OCC
fears-confirmed
worst_fears_realized
tokens_end
```

Listing D.2: Initialization file `relief.se`

```
polygon_begin QUAD
vertex 100 0 100 1.0
vertex 100 50 100 1.0
vertex -100 50 100 0.2
vertex -100 0 100 0.2
polygon_end
polygon_begin QUAD
vertex 100 0 -100 1.0
vertex 100 50 -100 1.0
vertex -100 50 -100 0.2
vertex -100 0 -100 0.2
polygon_end
decayFunction LINEAR
lifetime 10.0
standardIntensity 0.0
type RELIEF
tokens_begin OCC
```

```
relief
tokens_end
```