TECHNISCHE
UNIVERSITÄT
WIEN

VIENNA
UNIVERSITY OF
TECHNOLOGY

D I S S E R T A T I O N

Behaviour Coordination
for Models of Affective Behaviour

ausgeführt zum Zwecke der Erlangung
des akademischen Grades eines
Doktors der technischen Wissenschaften unter der Leitung von

o. Univ.-Prof. Ing. Dr. Robert Trappl
und
Dipl.-Ing. Dr. Paolo Petta

eingereicht an der Technischen Universität Wien
bei der Fakultät für Informatik

von

Dipl.-Ing. Stefan Rank
Matrikelnummer: 9725431
Grienauergasse 1
A-2380 Perchtoldsdorf

Wien, im Jänner 2009

Abstract

Affective agent architectures are an important field of research for a wide range of disciplines. Software or robotic agents that can reproduce some of the (human) phenomena that are labelled as emotional have a range of applications in entertainment, pedagogy and human computer interaction in general. In addition, the topic is relevant to cognitive science, and the processes underlying affective phenomena are important, if not essential, for the problems of action selection and behaviour coordination in control architectures of entities situated in open environments.

In this thesis, based on previous experience in modelling emotion, the method of scenario-based analysis for the comparison and design of affective agent architectures as well as a new approach towards incremental modelling of emotional phenomena are introduced. The approach uses concurrent processes, resources, and explicitly modelled related limitations as building blocks for affective agent architectures. Rather than presupposing a complex framework, an incremental methodology is proposed. As a key element, scenario-based analysis is presented as a suitable new method to capture the relevant aspects of an affective scenario of use for the comparison and the design of affective agent architectures.

This thesis focuses on behaviour coordination as a central problem of an autonomous intelligent agent. Based on a survey of current agent architectures and their scenarios of use, as well as an analysis of recent developments in theories of emotion, the proposed implementation approach is motivated for the computational modelling of emotion in virtual worlds. Importantly, the incremental method of building different executable computational models of emotional phenomena based on reusable building blocks avoids the reification of emotion types and static appraisal frames. The focus of the thesis, thus, is on coordination mechanisms in a concurrent model of affective competences.

Zusammenfassung

Affektive Agentenarchitekturen stellen ein wichtiges Forschungsfeld für verschiedene Disziplinen dar. Software-Agenten oder robotische Agenten, die die (menschlichen) Phänomene, die als emotional bezeichnet werden, teilweise reproduzieren können, haben breite Anwendungsmöglichkeiten in der Unterhaltung, der Pädagogik und in der Human-Computer-Interaction allgemein. Darüberhinaus ist das Thema auch für die Kognitionswissenschaften relevant und die Prozesse, die affektiven Phänomenen zugrunde liegen, sind wichtig, wenn nicht sogar essentiell, für die Probleme der Handlungsauswahl und Verhaltenskoordination in Kontrollarchitekturen für Entitäten, die in offenen Umgebungen situiert sind.

Basierend auf früheren Erfahrungen in der Modellierung von Emotion, führt diese Arbeit die Methode der Szenario-basierten Analyse für den Vergleich und das Design von affektiven Agentenarchitekturen sowie einen neuen Ansatz für das inkrementelle Modellieren von emotionalen Phänomenen ein. Der Ansatz verwendet parallele Prozesse, Ressourcen und dazugehörende, explizit modellierte Beschränkungen als Grundbausteine affektiver Agentenarchitekturen. Statt ein komplexes Framework vorauszusetzen, wird eine inkrementelle Methodik vorgestellt. Als Schlüsselelement wird die Szenario-basierte Analyse als passende neue Methode präsentiert, um die relevanten Aspekte eines affektiven Szenarios für den Vergleich und das Design von affektiven Agentenarchitekturen einzufangen.

Die Verhaltenskoordination wird in dieser Arbeit als zentrales Problem für autonome intelligenten Agenten betrachtet. Aufgrund eines Überblicks über derzeitige Agentenarchitekturen und ihrer Anwendungsszenarien, sowie einer Analyse neuerer Entwicklungen in den Emotionstheorien, wird der vorgeschlagene Implementierungsansatz als passend für die komputationale Modellierung von Emotionen in virtuellen Welten gezeigt. Die inkrementelle Methode, verschiedene ausführbare Modelle von emotionalen Phänomenen mit wiederverwendbaren Grundbausteinen zu implementieren, dient zur Vermeidung der Reifikation von Emotionsklassen und statischen Appraisalframes. Der Fokus dieser Arbeit sind daher die Koordinationsmechanismen in einem parallelen Model affektiver Kompetenzen.

Acknowledgements

I would like to express my gratitude to my advisors, Paolo Petta and Robert Trappl. Without their encouragement and the intellectual freedom I enjoyed while working with them, this research would not have been possible.

I must thank OFAI (the Austrian Research Institute for Artificial Intelligence), the institute that gave me the opportunity to work in different projects* catering to my diverse interests. Further, I want to thank my colleagues at OFAI for a great office environment, quiet when necessary and bustling when appropriate, but always enjoyable.

I would also like to thank my family who provided me with a wonderful start into life and supported me during my whole education, in many different ways.

Last, but definitely not least, I offer all my love to Marlene. Thank you for putting up with me, for supporting me all this time, and for being an invaluable source of inspiration.

An updated version of this thesis, along with notes about any changes as well as links to related software and other resources, should be available for download at
http://strank.info/phd/

Contents

Contents

List of Figures

List of Tables

Part I

Emotions, Agents, Stories

1   Introduction

Where the contributions of this thesis to the field of affective agent architectures are laid out, the research approach based on the incremental and iterative extension of executable models is explained, and the structure of the thesis is explained.

The human mind is the most complex and still the most mysterious kind of matter that we know of. Emotions are, in turn, one of the most mysterious phenomena that the human mind, in contact with other human minds, can exhibit. One of the practical approaches towards solving the mysteries that lie within emotional phenomena is to build models, in particular computational models that are executable and that can be experimented with [Fum et al. 2007].

Finding a suitable approach towards building computational models of different emotional phenomena that enhance our understanding is the main topic of this thesis. It is still an unsolved problem how to implement intelligence or emotion in a machine. While there have been big advances since the effort of the Artificial Intelligence field began, both in the modelling technologies available and in the methods employed, and although many areas of application have been opened up by insights and new techniques, many potential application areas are still untapped due to the lack of suitable research results.

I argue that the research approach proposed in this thesis can be used to cover phenomena as described by current theories of emotion, while making explicit the symbolic and computational shortcuts necessary to achieve a certain depth of modelling. This furthers the line of research of Tabasco [Petta 2003] (see section 2.5): in addition to doing away with reification of emotions, I aim at deconstructing the notion of static appraisal frames, to be replaced by coordination mechanisms in a concurrent model.

On a more general level, this research addresses the question of when emotional mechanisms are warranted in artificial agents. For which constellations of environment and agents are mechanisms modelled after human emotion useful or necessary? What 'emotions' can even occur in the limited virtual environments that synthetic characters inhabit?

Table 1.1:   Main research contributions.

Scenario-based analysis method (Chapter 2)	The detailed characterisation of the scenario of use for a specific agent architecture is introduced as a new and essential means to capture the *emotional potential* of a scenario as part of the interaction potential: what emotional phenomena can actually occur in the world of the architecture and which cannot, and therefore can only be simulated or portrayed relying on designer knowledge. The scenario description not only meets a requirement for design[1] but provides also an effective tool for the comparison of (affective) agent architectures [Rank Petta 2006].
Analysis of recent theories of emotion and their relation to implemented models (Chapters 3 & 4)	Based on an overview of recent advances in psychological theories of emotion [Frijda 2007; Scherer 2000; Scherer et al. 2001], those aspects that are most relevant to the implementation of affective agent architectures and those that have been taken up in different implementations are presented and analysed.
Building blocks for computational modelling of emotion in a virtual world (Chapter 5)	Executable computational process models of emotion are based on specific sets of modelling primitives. Motivated by the requirements and the potential of a specific scenario and concepts used by emotion theories, I propose as building blocks explicitly bounded resources and concurrent processes acquiring and using them[2]. The computational model thus focuses on the identification and management of parallel processes and the resources they access. Although the target domain is a virtual world, resources are intended for 'physically' anchoring the agent in the world. This approach aims at an incremental modelling of a growing collection of emotional episodes, with a clear documentation and delineation of technically necessary simplifications of the natural phenomena.
An implementation that accounts for modelled emotions based on the building blocks (Chapter 6)	An episode of disgust is used to exemplify and discuss the approach, which is realised using real-time cooperative microthreading technology. The construction of such an affective architecture poses the question of what configurations of processes and resources are relevant for specific emotional phenomena (from social emotion episodes such as anger, shame, and gratitude to individual dispositions) and how they are coordinated. It is intended as a principled step towards answering this question for different areas of application, such as interactive stories.

1.1 Research Contributions

One of the main contributions of this thesis (see table 1.1) is the proposal to use explicitly bounded resources and concurrent processes as building blocks for computational modelling of emotion in a virtual world. The proposal is intentionally minimalist in that it refrains from imposing any specific architectural framework with fixed parts and partitions. Complementing the short overview in this section, conclusions are offered in chapter 7 for a summary evaluation of my work. The role of emotion for believability, as a means of conflict resolution in the control structure of agents, and as a driving force in drama connects all of these contributions.

1.2 Open Challenges from Prior Work

The approach taken in this thesis is based on experience gained in creating synthetic actors. In earlier work—ActAffAct [Rank 2004, 2005], see also section 2.5—I constructed affective agents to be used in story generation. They were based on an extended belief-desire-intention (BDI) model by integrating an appraisal process and static reified emotion types as described in the OCC theory [Ortony 2003; Ortony et al. 1988]. This approach can be successful for a given target scenario if the details of the interaction between agents and their environment are suitably restricted. Compared to the richness of emotional life in humans, however, and also considering the possibilities of human-computer interaction (HCI) in recent virtual worlds, several limitations became apparent.

This section presents details on the limitations of earlier approaches and the objectives that result from addressing them. I propose to re-think the building blocks of affective agent architectures to provide for a reasonably complete and integrated agent architecture [Barsalou et al. 2007; Thórisson 2007] while delineating shortcuts in modelling necessary for technical reasons or reasons rooted in the targeted scenario.

My approach acknowledges that ongoing 'always on' interaction of the agent with its environment[3] is fundamental, and that physical and material constraints are important. These tenets can be put to use in a

---

[1] This is similar to the distinction of design space and niche space by Sloman [Sloman 2000, 2004; Sloman Scheutz 2002], see also [Arbib Fellous 2004; Dolan 2002].

[2] Please note that the proposal of using the building blocks described in this thesis as abstractions concerns the runtime of models. At design time, there may well be more suitable, higher-level abstractions.

[3] Terminology used in this thesis: *agents* (virtual bodies plus control architectures) are embedded in *their environment*; agents and environment form the virtual *world*.

virtual agent situated in a virtual world without sacrificing the benefits of dealing with an environment under full control of the modeller.

1.2.1 Scenario Definition or Pitfalls to Avoid

As stated above, a suitably restricted scenario makes the use of a limited model of emotion viable. Models of artificial emotion are a crucial element for building intelligent agents in social virtual worlds [Gratch Marsella 2007], and can also provide valuable benefits for theorising in emotion psychology [Cleeremans French 1996; Gratch et al. 2006, 2009]. Human emotions, however, are extremely complex phenomena involving cognitive, perceptual, expressive and social competences. Attempts at operationalising mechanisms of emotion for an artificial system in a virtual world need to choose from a range of approaches differing in scope, detail, and fidelity of simulation—both of the system itself and the environment it is embedded in. The structure of current computer systems and the abstractions that are easily available in current programming languages often bias the structure of computational models: A single locus of control, widespread use of symbols for internal state and communication, absence of timing constraints, and conceptually unbounded use of resources are typical symptoms.

In early cognitive architectures, such features of implementations have posed difficulties for their application in robotic scenarios, and the concept of embodiment is often used to argue for the need of a different approach for agents in the real world [Chrisley Ziemke 2003]. More generally, it is the particular scenario of use, including the motivation for modelling emotion, that determines the requirements that a particular implementation needs to meet [Rank Petta 2006].

A key point of the research approach pursued in this thesis is therefore to start with an analysis of the limitations of earlier approaches and to use them to formulate requirements and objectives for manageable scenarios that can help to overcome these limitations. Details on the scenario-based approach can be found in section 2.3, a more complete presentation of the specifics of related agent architectures will be presented in chapter 4. The following list gives an overview of the limitations I aim to address:

- *Symbolic* sensing and acting: The simulation of sensing and acting in a virtual world by exchanging pre-structured symbolic information does not match the intricate relation between perception and appraisal in real life. The disambiguation, structuring, and valuation of information acquired in sensorimotor interaction is an integral part of appraisal as proposed by emotion models in cognitive psychology.

- *Reified* emotions: An advantage of the use of a small set of distinct and fixed emotion types as a mechanism rather than a means of analysis/reflection is that it circumvents the need for a process-based implementation of finer granularity [Petta 2003]. However, emotions have temporal extent, they unfold over time, vary over time, and include constituent processes at different time scales [Frijda 2007]. Inevitably, abandoning the use of reified emotions entails the issue of what alternative approach to adopt to define the specific emotional phenomena to be reproduced and evaluated.

- *Rigid behaviour structures*: Classical BDI architectures use planners that are based on action sequences of relatively coarse granularity. The necessary realisation of higher level actions in adaptive lower level structures is often neglected and this often leads to easily recognisable repetitiveness. An agent that is situated in a world often can exploit the structure of the world itself to limit the complexity of these coordination tasks and to produce varied behaviours.

These identified limitations led to adopting the following objectives and goals for the design of a revised agent architecture and matching target scenarios.

- *Sensing and acting over time*: In a virtual world, the designer can choose from a range of mechanisms for the implementation of sensorimotor interaction with the environment. To match the inherently temporal nature of emotional phenomena, suitable mechanisms have to be chosen. The dynamicity introduced thereby goes some way towards overcoming limitations rooted in the shortcuts of symbolic approaches.

- *Embodiment for virtual worlds*: More generally, the simulation of the virtual world needs to take issues of embodied grounding into account [Anderson 2003; Chrisley Ziemke 2003]. In particular, the (bodily) facilities available to an agent are necessarily bounded. This includes the processing capacity, leading to the necessity of tractable mechanisms. Further, the agent is physically located in a world—if virtual—and thus needs to obey its physical laws, but it can also exploit the structure this embedding provides in the form of its lifeworld [Agre Horswill 1997]. The proposal to use parallel processes that are by default independent is directly motivated by the need to overcome the limitation of rigid behaviour structures. Such a collection of processes however then needs flexible coordination mechanisms, regulatory components,

and means to actively direct and focus attention (internally and externally) to actually engender the variety sought for.

- *Multimodality*: To ensure a virtual counterpart to sensing and acting that at least approaches the breadth of human activity, I target a scenario that involves four different modalities: seeing, hearing, smelling, and basic forms of physical movement and manipulation. The complexity of these modalities (and other objectives such as real-time interaction) limits the extent to which they are simulated. Therefore the mechanisms are simplified using for example false-colour rendering to model vision and simple spheres of influence for smelling and hearing [Thalmann et al. 1997]. The consideration of multiple modalities avoids simulating a single modality in detail at the cost of reduced breadth of coverage of agent-environment interaction.

- *Real-time human-computer interaction*: Synthetic characters are intended for interaction with humans in real-time. Targeting interaction with humans from the start constrains the allowed pace of the model: the pace of timely interaction must not be neither too slow nor too fast.

- *Social setting*: A social setting, i.e. interaction between two (or more) agents and a human user, is fundamental in order not to lose core aspects of emotion from sight [Parkinson 2001]. Many emotional phenomena can occur in social settings only.

- *Status of situational meaning structures*: One possible solution to the problem of reified emotions, apart from limiting oneself to the recognition and reproduction of expressive behaviour and coping behaviour, may be to focus on the framing of individual appraisals as they occur inside an architecture. 'Situational meaning structures' (or appraisal frames) [Frijda 2007] are theoretical constructs that aggregate information of individual appraisal checks (or 'stimulus evaluation checks' [Scherer 2005]). When process-based models of emotion are implemented, a decision needs to be taken on how, when, and whether to aggregate individual implemented appraisal checks into such a larger structure [Marsella Gratch 2009; Reisenzein 2001]. Pre-existing and static appraisal frames (such as reified emotion types) result in predominantly linear dynamics in the effects on behaviour. To overcome this limitation, coordination mechanisms of more fine-grained appraisal processes are needed, but it is an open research question whether the effects of independent appraisal checks are by themselves sufficient to model the role of situational meaning structures in emo-

tional phenomena. Ideally, the observed behaviour that is usually explained with the use of situational meaning structures should be an emergent effect of such coordination (and any other) mechanisms.

- *Parsimony*: From the perspective of designing an emotionally competent agent, it is preferable to employ a parsimonious model that is readily understandable. However, such tendency might conflict with the objective of grasping the complexity of emotional phenomena.

1.2.2 Incremental Modelling, One Phenomenon at a Time

Taken together, the objectives pose a significant challenge: a sufficiently complete and valid infrastructure is needed in which elements relevant for emotion are present and functional. For the purpose of this thesis, a sufficiently complex three-dimensional world that can be manipulated by a human user has been developed as such a test-bed. Rather than creating one general model, I aim to incrementally model a growing collection of specific emotional phenomena in a grounded fashion. As part of this approach, I fully expected to introduce several shortcuts as necessary to model any specific phenomenon. In addition to the objectives introduced above, these shortcuts, and the shortcuts that any implementation of a virtual world needs to take compared to real-world interaction, would be made explicit. Focusing on one specific phenomenon at a time also helps to avoid introducing unwarranted postulates needed in models claiming generality: e.g., a set of a priori personality parameters and very general competences, such as a global planning process.

The development of a process-based model followed structuring principles. Based on earlier work [Rank et al. 2005], I use an initial typology of processes (i.e.: activities, behaviours, tasks, regulation processes, and action tendencies) and based on the component process model of emotion [Scherer 2005], I assume five principal domains of resources corresponding to the five organismic subsystems involved in emotion (i.e.: information processing, support, executive, action, and monitoring).

To exemplify and validate the approach, I will use the scenario of modelling a specific emotional phenomenon, disgust, in a virtual world setting similar to a computer game. The application of the approach is illustrated through the modelling of a specific episode of disgust; after which the subsequent extension to further phenomena and scenarios is addressed.

1.2.3 Evaluation

The evaluation of the proposed architectural design principles poses another challenge. The underlying motivation for the concrete implementation work is the creation of synthetic characters to be used in effective human-computer interaction. The design of the architecture draws on psychological theories of emotion. Its implementation is a research endeavour in computer science.

These three different (though related) disciplines suggest three different evaluation criteria. In the case of HCI, usability; for psychology, a test framework modelled after controlled laboratory experiments; for computer science, tests of feasibility and performance. In the scope of this work, relatively simple performance measures defined on the specific target phenomenon seem to be the most suitable evaluation criterion. An example is a higher level description of a possible interaction sequence between the agents (and potentially a user) that can be matched against the trace of actual runs of the system. This can be used to test whether intended phenomena actually do emerge for specific sufficient starting conditions and compliant user interactions.

The fundamental evaluation criterion, however, is the feasibility of recreating features of the particular emotional phenomenon to be modelled, solely through the interaction of relatively unspecific coordination mechanisms, rather than by introducing computational mechanisms modelled after the features themselves[4].

A further possible aim for designing a model of emotion could include considerations of the ontogenetic development of such an architecture and the correspondence with current knowledge about brain anatomy [Granger 2006]. I currently exclude such considerations, to constrain the scope of the effort. Another candidate aspect for future work is extending the coverage of 'higher level' capabilities such as complex memory formation, planning, and deliberation, on top of the proposed architectural design principles.

1.3 Thesis Structure

As far as possible, each part of this thesis is written as a self-contained module. Even so, earlier parts do provide background for the later ones. Therefore, I recommend to read at least some of the earlier sections before skipping to the back of the text. As a guide for the experienced reader, each chapter starts with a short precis of the contents to allow to judge the relevance to the reader's particular interests. As

---

[4] This could also be described with the, unfortunately heavily overloaded, term 'emergence'.

a further aid for navigation and perusal, I provide a compiled summary of all parts here.

Part I
Emotions, Agents, Stories

This part introduces the research topic, the research approach, and the research questions of the field of affective agent architectures.

Chapter 1   Introduction
Where the contributions of this thesis to the field of affective agent architectures are laid out, the research approach based on the incremental and iterative extension of executable models is explained, and the structure of the thesis is explained.

Chapter 2   Affective Agent Architectures
Where the terminology and research questions surrounding the concept of affective agent architectures are introduced; where the application areas they are useful for are reviewed with a focus on Interactive Storytelling and the requirements of interactive storyworlds; where the concept of a scenario is defined to support the comparison of different architectures.

Part II
Behaviour Coordination

This part focuses on the modelling of emotion, both in theoretical approaches of different disciplines and in the implementations of computational models of emotion, and its relation to the topic of behaviour coordination.

Chapter 3   Theoretical Foundations
Where the relevant concepts of cognitive appraisal theories of emotion are explained; where the importance of emotion in behaviour coordination is motivated; where the descriptive concepts of levels, modelling primitives, and embodiment are examined.

Chapter 4   Survey of Related Work
Where the uptake of emotion theories in computational modelling is examined; where different agent architectures and their application scenarios are introduced; where implemented mechanisms that involve emotion are pointed out.

Part III
Affect, Processes, and Resource Management

This part presents the specific approach of this thesis based on architectural building blocks and describes its application to emotional phenomena.

Chapter 5 Building Blocks for Computational Modelling
Where the procedural primitives and the details of the chosen implementation are described. Where the modelling of the limitations of an agent is explained.

Chapter 6 Disgust as First Emotional Phenomenon
Where the instantiations of the modelling primitives in an executable environment for reproducing a specific emotional phenomenon are described. Where the extension of the implemented methods to further areas of emotional life is described. Where the relevance of the concepts developed in previous chapters is tested in the analysis of further scenarios.

Part IV
Closure?

This part summarises the ideas presented in this thesis and points out potential areas for future research.

Chapter 7 Conclusions
Where the main contributions of the thesis are summarised and a plan for potential future work is laid out.

Appendices

The *appendices* and the *bibliography* provide background information on the implementation described in the thesis and the sources of literature used within this research. Further, an index and a glossary are provided for quick reference.

App. A Microprocesses Package Based on Python and Twisted
Where the interface of the microprocesses package, a Python package that implements cooperative microthreading based on Python generators and the network library Twisted, is described.

App. B Resources and Interprocess Communication
Where the implementation of resources and interprocess communication in the implemented system, BehBehBeh, is described.

App. C   Physics and 3D Engine Integration in BehBehBeh
                Where the implementation of the environment simulation for the
        scenario of BehBehBeh based on a graphics engine and physics engines
        as well as several other open-source libraries is described.

1.4 Summary

In this chapter, I have introduced the research contributions, the re-
search approach, its objectives and potential strategies for evaluation.
The chapter concluded with a survey of the structure of the thesis text.

2 Affective Agent Architectures

Where the terminology and research questions surrounding the concept of affective agent architectures are introduced; where the application areas they are useful for are reviewed with a focus on Interactive Storytelling and the requirements of interactive storyworlds; where the concept of a scenario is defined to support the comparison of different architectures.

Intelligent software agents are a research area in artificial intelligence that works on developing entities that are autonomous, persist over time, and can act in and adapt to their environment. Agent architectures are used to implement these software agents in a principled fashion. Affective agent architectures target domains in which emotional and related phenomena are important or investigate the potential of contributions from emotion theory to improve performance in other domains.

This chapter first introduces the agent concept and presents the different areas that this concept has been used in, as a tool for implementations or even only as a metaphor for aspects of models. Due to the broad applicability, there are many application areas that can benefit from an agent-oriented perspective. This chapter gives an overview of these areas and takes a more in-depth view at interactive storytelling in particular. So far a well-structured and clear-cut way of comparing different agent architectures in particular in the areas that are relevant to the present thesis was missing. I therefore introduce the method of scenario-based analysis and comparison. Finally, the idea of interactive storyworlds is used to concretise the research goals already presented in section 1.2.

2.1 The Intelligent Agent Concept

[Russell Norvig 1995, 2003] have proposed the idea of an intelligent agent as a unifying theme to present all of the diverse subfields of Artificial Intelligence (AI). In this context, agents are regarded as the main topic of AI: entities that receive percepts from their environment and perform actions in it, see Fig.2.1. For the purpose of this chapter, it shall suffice to consider two aspects of the intelligent agent concept:

software as an agent, and the *human as primary example* of an agent. Chapter 3.3 will expand on more detailed aspects, such as different ideas about the meaning of autonomy in affective agents and why a strict conceptual separation of percepts and actions can be detrimental in the design of agents.
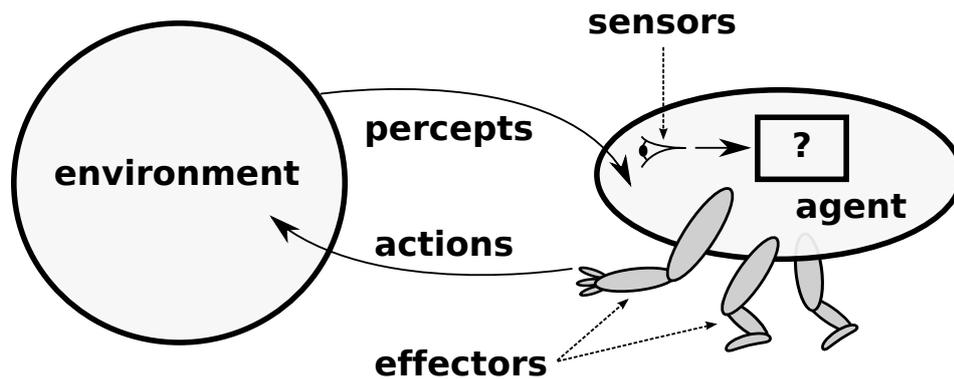
Figure 2.1: The intelligent agent concept as illustrated by [Russell Norvig 1995], p.32.

2.1.1 Software Agents

There is no undisputed definition of the term 'software agent' in the scientific field. The definition that is probably most directly accessible is derived from the common usage of 'agent' in everyday language, e.g. the agent of an artist. Such an agent should be someone with expertise in a specific field who is entrusted to act on your behalf. The most important aspects of this definition are therefore *trust* and *personalisation*: the agent can be trusted to do the right thing and it is *your* agent who respects *your* wishes and preferences.

Agent Oriented Software Engineering (AOSE) (and to some degree also the related concept of a Service Oriented Architecture (SOA)) advocate the use of agent concepts in the design and construction of software systems. Among the intended benefits, especially for distributed systems, are an increase of *flexibility* and *robustness*. The abilities to flexibly use multiple ways of doing something and to robustly detect failures and respond appropriately are needed to deal with unexpected events in often unpredictable environments.

In more general terms, [Wooldridge Jennings 1995] name four characteristics that are relevant for software agents: *reactivity*, *social abil-*

ity, *autonomy* and *proactiveness*. An agent is thus an autonomous entity that can balance proactive and reactive behaviour as appropriate. It pursues its own agenda, interacts with other agents, exists over an extended period of time and does not require external control. Beyond these technical definitions of agency, the relation of intelligent agents to humans is of particular relevance for the modelling approach in this thesis.

2.1.2 Humans As Exemplary Agents

The notion of an agent is foremost derived by taking humans as example. Just as humans are capable of dealing successfully with their highly dynamic environment, agents should be designed to capture the most important aspects of intelligence for their specific environments. The environment will often contain other agents, thus requiring social abilities including interaction that is meaningful to third parties. Furthermore, software agents that interact also with humans, need to be understandable for them as well.

Several functions of such intelligent agents therefore borrow from psychological concepts: perception and situation awareness, action tendencies, coping, learning, communication, and signalling. On the other hand, software agents are a valuable tool for models of natural systems, including humans. The field of Agent-Based Social Modelling (ABSM) is dedicated to these methods including cognitive modelling.

Emotional phenomena in humans and in interactions between humans are prime examples of what on the one hand can be modelled using agents and on the other hand can improve the functionality of software agents in many respects. Human emotions are part of the subjective interpretation of dynamic and unpredictable situations instigating specific kinds of reactions, see chapter 3.1. Software agents in general could benefit from the properties that an emotional system promises: indeed, it may be the decisive factor in a dynamic environment. Not only the interaction with humans is inherently dynamic and unpredictable, also the environment of autonomous software agents in computer networks, up to a global scale like the Internet, might be easier to master with such a functionality.

The incorporation of emotion promises to contribute to all of the agent characteristics proposed. The emotion process as a whole is directed at selecting and steering action. A response that is both fast and adapted to the situation as interpreted by the agent would provide improved reactivity, and emotional expressions also add to what information is communicated between agents and how. This for humans mostly involuntary communication of emotional states would also contribute to the social abilities of agents, as well as the consideration of

standards (or social norms) in the elicitation of emotions and the possibility of long-term relationships through the longevity of preference values for other agents. As a final example, motivations, a fundamental element in the modelling of emotion, provide a basis for autonomy.

2.2 Application Areas of Affective Agent Architectures

In this section, I will characterise three different, but related, application areas of affective agent architectures: Cognitive Science, Embodied Conversational Agents (ECA), and Autonomous Personality Agents (APA).

In Cognitive Science and related areas such as Psychology, intelligent agents provide an ideal platform to test theories and models computationally. An executable model of cognitive faculties enables different research methodologies that were impractical before the advent of modern computing [Cleeremans French 1996]. Further, executable models require building a reasonably complete model where no parts can be washed over by hand-waving arguments (rather, at the minimum, simplified mock versions need to be built). Finally, a computational model is more conducive to provide explanatory models that indicate mechanisms rather than models that are only descriptive [Fum et al. 2007].

The term ECA has been popularized by [Cassell et al. 2000], a succinct definition is given in [Cassell 2001]: 'an interface in which the system is represented as a person, in which information is conveyed to human users via multiple modalities such as voice and hand gestures' (p.67). While the effectiveness of such a person-like interface for any specific application scenario can be debated, in any case such interfaces need to integrate an understanding of human emotions in interaction.

I use the term APA to describe agents that have their own agenda 'built in' and that exhibit some specific personality. For sustained interaction with humans, it is important for an agent to act coherently and consistently over time, an ability that psychology describes as personality. While an advanced ECA likely also falls under that definition, an APA is useful in many different application areas, ranging from pedagogical applications to healthcare (cf. the idea of a virtual butler, that does not necessarily require an ECA interface) and to entertainment and art in new media [Manovich 2001; Murray 1997; Wardrip-Fruin Montfort 2003]. Precursors of personality agents that are truly autonomous can be found in present day gaming, where such an agent is typically called a Non-Player Character (NPC).

This short list already gives an impression of the diversity of appli-

cations that affective agent architectures can be used in. While the architectures used in different areas sometimes are similar, the application scenarios differ significantly. In order to capture the relevance of differences and similarities, the next section introduces scenario characterisations as a tool for comparing different applications of affective agent architectures.

2.3 Scenarios for Comparability

In the EU-project Humaine[5], I developed the notion of scenario descriptions as a tool for comparison of affective agent architectures by making criteria applicable across different agent applications explicit [Rank Petta 2006]. To be able to relate different approaches that meet specific sets of requirements, I used a context-dependent analysis based on scenario descriptions: *scenario-based analysis*. Such descriptions explicate purpose of and motivation for building a system, detailing the desired interactions while maintaining a clear separation from implementation aspects.

Aaron Sloman [Sloman 2005] makes a similar argument concerning planning and evaluation of research[6], arguing for a focus on robotic scenarios. Scenarios are points in the niche space[7] for agents: the possible purposes and environments of use. However, scenarios are also valuable for the design phase of a new agent application and they suggest suitable evaluation criteria.

As scenario-based design and evaluation is an established concept in usability [Cooper 1999; Nielsen 1993], I need to point out the differences to the present concept: in usability, specific user stories allow for iterative refinement of the actual requirements of a system. A scenario in usability is often a prototype that handles one user achieving one specific goal and thereby focuses on specific functionalities and a certain depth of the system[8]. The analogy I use for (affective) agent architectures consists in regarding human behaviour as the system—a scenario then is the part that an agent architecture should be able to reproduce. Scenarios are thus used to capture the *emotional potential* of an envisioned use of a system: those characteristics that help constrain which kinds of emotional phenomena *can* actually occur *in* the virtual world, and which can*not*—and therefore can only be 'faked' or portrayed. Note that in most of today's applications of character-based

---

[5] http://emotion-research.net/projects/humaine/

[6] See [Sloman Wyatt 2006] for a proposed scenario template.

[7] In evolutionary biology the role of an organism in an ecological system to which it is adapted is called its ecological niche.

[8] Note that there are also other uses of the term scenario, e.g. to capture the social situations surrounding the use of a system [Crowley 2006].

interfaces, the expressed emotions are only portrayed in this sense: e.g., the simplified interaction in many finite-choice-based dialog systems cannot support a deeper, grounded, model of emotion.

2.3.1 Scenario Descriptions

The basic characteristics of an application scenario are the **motivation** for building the system, its **purpose**, and the **details of a possible deployment**. Motivations may range from specific data about humans or animals that the architecture should model (e.g. [Gratch Marsella 2004c]) to explicit hypotheses and open empirical questions that need to be tested, and might also include specific engineering goals, e.g. the improvement of behaviour selection in robots [Moshkina Arkin 2003]. A characterisation of the purpose of the system positions it along a spectrum between real-world applications and the creation of virtual entities that can be used in controlled experiments for the scientific validation of (e.g. psychological) theories. The latter empirical and scientific context is closely connected to the envisioned mode of evaluation, possibly including explicit performance functions but also less concrete design criteria that the system should meet at a social level.

As to the details of deployment, a crucial point is the characterisation of the system's *interaction qualities*: this includes the *user interface* as well as the interactions between the agents and their environment. In fact, the user interface can be regarded as a special case on the spectrum of *agent-environment interaction* that ranges from sequenced binary decisions and discrete sensations (e.g., in the Prisoner's Dilemma) to the complexity of human interaction in the real world. User interfaces can be as reduced as in choice-based menu interaction or as complex as, potentially, in robotic applications. In the case of affective interactions, ontologies derived from folk psychology or a particular emotion theory are needed to describe the interactions possible in a scenario. I want to emphasise that such characterisations are mostly formulated from an external perspective: the objects, properties, relations, and processes mentioned in these descriptions do *not* automatically imply or justify their actual use or reification in architectures tackling the scenario (symbolic models that directly operationalise folk psychological terms are one possibility, but usually lack grounding [Norling 2004]).

Interactions can be described informally as typical scenario scripts that illustrate the possible activities, including tool use and social relations, as well as the utilisation of second-order resources, complemented by negative scripts that explicate interactions that fall outside a given scenario[9]. A more formal description of interaction qualities could also include all agent tasks possible; agent-local performance

measures (e.g., the amount of resources collected over time or per timestep in a simulation); the average number of conflicting long-term or short-term tasks; and further qualitative behavioural criteria such as coherence, variedness, or believability in virtual character applications. Even though hard to quantify, the latter often form an essential part of scenario descriptions.

Another part of the details of deployment is the characterisation of the environment as presented to the agent. This comprises the intrinsic limitations, dynamics, and regularities of the interactions. For simulations, this includes properties such as being time-stepped or asynchronous, with the resulting differences for the possible interactions and mechanisms (cf. the 'PEAS' characterisation in [Russell Norvig 2003]). It is apparent that the interface to its world differs substantially from a robotic agent to a virtual one, but even in simulated environments a range of sensorimotor interactions is possible, including simple choices, artificial life simulations, and simulated physics. The scenario should also specify the numbers of agents and agent types (including interacting humans) in terms of typical and hard or practical limits[10]. For practical reasons, references to related scenarios are also helpful.

My overall tenet is that architectures targeting related scenarios will benefit from analyses framed by scenario characteristics. This is especially true for target scenarios calling for several emotional competencies, such as interactive storyworlds [Rank Petta 2005b] addressed in section 2.4.

To summarise, a scenario that is suitably defined and restricted clarifies what kind of (artificial) emotions are warranted in it, which are frequent, and which are possible at all.

2.3.2 What Architectures Aim For

The questions that scenario descriptions are to frame are: What are different affective architectures aimed at in their implementation of emotional mechanisms? What is the intended role of mechanisms designed to capture (parts of) the phenomena called 'emotion' for a system in a given scenario? In this section, I give an overview of my conceptualisation of the possible roles emotion can play in an affective architecture. Examples of architectures and their respective scenarios can be found

---

[9] A negative script for a robotic competitive two-resource problem could for example include recognising that the other robot is currently consuming, or maybe even any sensing of the other robot at all. A cooperative solution would then fall outside this specific scenario, but could be part of an extended version.

[10] Consider e.g. the (arguable) inappropriateness of shame (as opposed to remorse) in a simulation with only two agents.

in chapter 4. Issues that are related to problems of shallow vs. deep modelling of emotion need to be considered: When is it justifiable to attempt to model only a single mechanism purported to be part of emotional phenomena? The roles of emotion as a whole as well as the roles single affective mechanisms can play in scenarios need to be examined from a computer science and AI perspective, striving to distinguish ultimate functions from proximate effects [Ketelaar Clore 1997].

The functional roles attributed to emotion include not only informing an entity's decision making system about the significance of situation changes [Frijda 1986; Sander et al. 2003], but also integrating disparate architectural mechanisms, as well as a strategic role in structuring situated behaviour—especially in socially situated contexts. [Scheutz 2004] presents a list of potential roles of emotion in artificial agents and [Arzi-Gonczarowski 2002] proposes upgrades to basic reactive mechanisms that are related to emotional functionality. Such different functional roles are called for in different subsets of all possible scenarios. The generation and recognition of expressive behaviour are prominent examples of emotional functionality, as they are the focus of most work in the realm of ECAs, but may be less important for research on behaviour arbitration in robots. To my mind, an overarching characterisation of emotion can be reached by considering the relation to an agent's autonomy. Autonomy is here understood as a matter of degree (cf. the important role of automaticity in human behaviour [Gollwitzer Bargh 2005], see also chapter 3.3), and an agent needs to determine when and how it can (or needs to) exert its autonomy with respect to its environment. Autonomy, thus, refers to the ability to actively limit or allow influences of the environment on the agent[11].

A basic need of an autonomous agent is therefore the active construal of the subjective meaning of changes (including the passing of time) in relation to its concerns, where a concern indicates the disposition to desire occurrence or non-occurrence of a given kind of situation [Frijda 1986], and the reconfiguration of the agent to meet the identified challenge or opportunity. This is increasingly relevant in scenarios that call for the repeated re-evaluation of construals because of ambiguities and the lack of clear-cut distinctions, such as in sustained real-time interaction with humans. Functionalities of single mechanisms can be regarded as subserving the appraising and coping processes. One behavioural functionality associated with emotional processes (and an agent's personality) is the long-term persistence and coherence of behaviour [Ortony 2003]. Process component models of

---

[11] Note that this conception of autonomy differs from the definition in [Russell Norvig 2003], p.37: 'To the extent that an agent relies on the prior knowledge of its designer rather than on its own percepts, we say that the agent lacks autonomy.'

emotion elaborate in detail on perceptual functions of emotional processes, i.e., the detection of constituents of emotional experience, such as novelty, pleasantness, and goal relevance. The more the interaction with an agent's environment is unconstrained, the more complex are principled assessments of these components. It then becomes worthwhile to provide operationalisations even of single affective mechanisms.

The social lifeworld of an agent [Rank Petta 2005a] needs to be continually re-enacted, with support from emotional competences such as the detection of norm violations and from shared activities, i.e., routines of structured agent interactions. The knowledge about regularities and their active maintenance is as crucial in the social context as at the physical and behavioural levels. A social aspect of emotional competence therefore is the ability to exploit these regularities as a special case of (external and internal) cognitive artifacts [Hutchins 1999]. A different perspective on the social functionality of emotion is provided by transactional accounts that describe emotional behaviour as skillful engagement with the world for relationship reconfiguration [Griffiths Scarantino 2005]. The strategic role of emotional behaviour, on these accounts, is to influence the behaviour of other organisms. A similar conceptualisation is at the heart of Affect Control Theory [Heise 2002, 2004], which highlights the confirmation of expectations related to one's social role. Evidence sustaining these views includes the importance of social context, such as the effects of audiences on the likelihood of display of emotion. Examples of scenarios that warrant several of these functionalities can be found in robotic applications integrating human-robot interaction, and in virtual worlds that allow relatively unconstrained real-time interaction with multiple agents and human users. While designers of robotic scenarios often need to be preoccupied with the details of real-world interaction, virtual worlds allow to create scenarios of differing degrees of complexity.

A more constrained example of the latter type of scenario covered in the next section are virtual storyworlds. Here, for an interesting amount of detailed interaction to be possible without preventing dramatically interesting coherent activities, different functionalities of emotion are needed: The coherent expression of emotion and its long-term influences, as well as the embedding in a social context are crucial for dramatic experiences. In relatively unconstrained environments, the problem arises of relating the rich and direct interactions on a small time-scale to the variety of activities on a larger time-scale. In current agent architectures these different aspects are separated as they involve different representations and levels of abstraction: from direct coupling to symbol manipulation on the other end. Comparing scenarios and identifying and investigating those that warrant multiple levels

should help to better understand and integrate this current gap. The following section elaborates on interactive storyworlds as one class of scenarios with suitable characteristics.

2.4 Requirements of Interactive Storyworlds

This section presents interactive storyworlds as a class of scenarios that allow for investigation of the aims laid out in the previous sections. The particular scenario for the implementation described in this thesis cannot be termed a real storyworld: the motivation and purpose of the implementation relate more immediately to the scientific requirements of a PhD thesis rather than an entertainment application. Nevertheless, the requirements that I adopted are motivated by the long-term vision of autonomous characters in interactive storyworlds, even if the specific scenario (a disgust episode, see section 6) is inevitably and intentionally restricted in scope and depth.

Emotions have been characterised as the essence of a story [Elliott et al. 1998] and play a central role in engaging drama. Drama is the art that deals with a refined version of emotional interaction between individuals [Egri 1946; Vogler 1996]. Drama theories describe plays, film scripts, and stories in general as changing constellations of emotional characters. These ideas provide a starting point and success criteria for the creation of interactive storyworlds. The goal of such a storyworld is to simulate characters whose behaviour is coherent in the context of drama. Explicit directions, however, given by a 'director' component, should not be necessary, the dramatic structure should rather emerge from the very interaction with and of the characters. Simplified formalisations of theories of drama can be used to evaluate the results: The stories that arise should be motivated by emotional constellations, not by plot directions. Based on the observation that emotions are a prominent part of analysing and writing successful drama, a storyworld system can use a psychological understanding of the functions and functioning of emotion to facilitate interactions that resemble drama.

2.4.1 Interactive Storyworlds

Storyworlds are virtual worlds inhabited by synthetic characters. They provide an environment in which users participate actively in the creation of a narrative. Several recent projects that include simulated worlds target the area of interactive narratives as a medium [Cavazza et al. 2002; Louchart Aylett 2004; Loyall et al. 2004; Magerko et al. 2004; Mateas Stern 2006], others pursue pedagogical aims [Machado 2004; Marsella et al. 2000; Martinez-Miranda et al. 2008]. Synthetic

characters are also put to use in entertainment and edutainment software. Synthetic tutors or life-like pedagogical agents are used to facilitate playful learning; the AI component in computer games controls the actions of NPCs. Both could be seen as synthetic actors that follow a script but nevertheless have to react to a changing environment in order to show appropriate and believable behaviour. A truly adaptive engine for these applications could reduce the frequent issue that the interactions are perceived as predictable and possibly boring.

Emotion in humans influences their behaviour. The expression of emotion, be it consciously or unwillingly, conveys important information to others. Emotional processes are important for reactive and adaptive actions in a dynamic environment. Synthetic actors with the ability to express emotions and to understand emotions expressed by others would be able to use this information in the planning and coordination of their actions. Using an imitation of the human emotion system to influence their actions could provide them with flexibility and adaptivity in an environment that is complex and unpredictable: e.g., in an environment in which autonomous actors strive for possibly conflicting goals, leading to a struggle that needs to be resolved—i.e. in a dramatic environment.

There are several different approaches to adapt narrative to the possibilities of new media. The notion of the narrative paradox characterises interactive narrative as a compromise between authorial control over the story flow and the freedom of interaction allowed for the user of the system [Aylett 2000]. Approaches to address the narrative paradox range from plot-driven to character-based [Avradinis Aylett 2003; Iurgel 2003; Louchart Aylett 2003]. At least the approaches that are described as character-based require synthetic agents with autonomy and personality. Affective agent architectures are used to construct such autonomous personality agents [Gratch Marsella 2007; Trappl Petta 1997; Trappl et al. 2003], and computational models of emotion are seen as a prerequisite for the required emotional and social competences (see also chapter 4).

Unconstrained real-time interaction requires agents that are situated. These should be agents that adapt to changes caused by the user and to changes in the environment, while preserving the consistency and coherence of their own actions according to their role and personality [Ortony 2003]. To my mind, the decisive factor in this type of synthetic characters is the operationalisation of a model of emotion. Confer also the analysis of emotion at the 'molar' level in [Smith Lazarus 1993] as core relational themes that resemble distilled versions of plot elements in drama theory.

To the individual agent, storyworlds present themselves as inherently social domains, as social interaction is crucial for the solving of

dramatic conflicts. In the ideal case, the author of such a storyworld would be able to shift from today's straitjacket of having to specify exact sequences of actions to the authoring of possible behaviours and activities, regularities in the environment, and the setting up of an initial constellation of characters endowed with particular personality traits. While the latter in all likelihood would not be an easier process of authoring, it could open the door to a more flexible, and user-driven, experience of dramatic structures.

The importance of emotion for (interactive) narrative is widely recognised [Cavazza et al. 2003; Loyall et al. 2004; Martinho et al. 2003], the focus varying from the role of expressive behaviour to convey character state to the motivational influence of emotional evaluations. My own approach [Rank 2005] (see next section) positions emotional processes as the central element for adaptive behaviour in a dynamic storyworld that allows for emergent plots.

The context of virtual worlds and real-time interaction restricts the set of usable techniques. Furthermore, there are many approaches, based on abstractions readily available in mainstream programming languages, that fail to meet the requirements of the specific scenario presented in this thesis, as already pointed out in section 1.2.1 and elaborated on in this section. The type of scenario I have in mind include (the modelling of) two or more virtual characters in a three-dimensional world that can be manipulated by a human user. For the purpose of this thesis, such a world is to be a test-bed for the incremental modelling of a growing collection of specific emotion phenomena in a grounded fashion, explicating all simplifications considered as technically necessary. This section details the requirements that impinge on my choice of basic building blocks for such simulations.

Cognitive architectures often use a single control structure and a small set of general representation schemes. An example is Soar [Laird et al. 1987] and emotion models based on it [Gratch Marsella 2004a; Marinier Laird 2006] which use, at the core, a single execution cycle and a central hierarchical data structure modified by operators. Similarly, logic-based agents employ propositions and a single inference mechanism. Such mechanisms and representations have their uses, and the present proposal is not aimed at replacing them with a 'better' alternative. Rather, I try to identify building blocks for my approach that allow the use of diverse resourceful [Minsky 2006] methods (i.e., having several ways to perform a task and the ability to switch between them as appropriate in a situation), while providing the features that I deem fundamental: fidelity of simulation, compatibility with theories of emotion, tractability, and the explication of modelling shortcuts.

2.4.2 Fidelity of Simulation

For my conception of storyworlds, an agent is to be simulated as a physical system with resource bounds in both space and time. The simulation of an 'emotional mind' as a physical system in a digital computer that is inherently serial calls for mechanisms that approximate the continuous and massively parallel nature of emotional agents acting in the real world, i.e., living beings. Further, all the resources in the model are part not only of the agent but also of the (simulated) world and thus *not* under the exclusive control of the agent. As an example of an external influence, consider a physics simulation that is part of the environment: Given its continuous effect on the agent, it warrants inclusion of adaptive action monitoring processes in the agent. Similarly, simulated visual perception has to be actively directed to focus on specific parts of the environment, and processes internal to the agent are likewise bounded, e.g. in terms of available processing time. Fidelity of simulation also implies that robotic architectures, where physical and material constraints cannot be abstracted, are a relevant source of insights. [Arzi-Gonczarowski 2004] argues for transferring the issues of embodied grounding that are relevant in creating particular but approximate models of intelligence for robots to the theoretical foundations for general accounts of intelligence. Similarly, [Aylett 1999; Petta 1999] present examples of applying the striving for applicability in the real world, as in robotic architectures, to virtual worlds.

2.4.3 Emotion Processes in Psychology

The building blocks chosen should allow assembly of models that are compatible with the rich descriptions of emotion processes in the psychological literature (see section 3.1). Two specific accounts of cognitive appraisal theories of emotion [Ellsworth Scherer 2003] shall be mentioned here: Scherer's component process model (CPM) [Scherer 2001] and Frijda's description of the laws of emotion [Frijda 1986, 2007] are both well suited for a process-oriented modelling approach. The following two aspects of these theories had a main impact on the choice of modelling primitives undertaken.

First, emotions are an inherently temporal phenomenon. They are episodes over time, vary over time, and include constituent processes at different time scales [Frijda 2007]. Descriptive accounts of emotion that disregard the temporal development of emotion episodes are of limited use in real-time environments. The chosen modelling primitives should support the temporal characteristics of emotional mechanisms.

Second, both theories assume that the normal operation of a (human) agent involves many parallel processes, mostly unsynchronised

and not reflected in overt behaviour. This can be inferred from the descriptions of the internal changes making up emotion: Scherer characterises emotion as an episode of interrelated, synchronised changes in several organismic subsystems; Frijda highlights the control precedence of the action readiness associated with an emotion, i.e. the tendency to override other concurrent activities. Multiple action readinesses can be aroused in response to parallel formation of appraisal patterns, and while only one may initially instigate overt action, others can surface later. Further, the primary effect of control precedence is counteracted by simultaneously active regulation processes.

2.4.4 Emotion and Sociality

Emotional phenomena combine aspects internal to the organism, such as the synchronisation of subsystems, with external aspects of social behaviour, including dispositions and roles [Parkinson 2001]. The social aspect of emotion is another consideration explicitly included from the start. It has no direct impact on my choice of building blocks, but it motivates starting points for models of specific emotional phenomena to be refined incrementally. As detailed in chapter 6, disgust was chosen as example phenomenon, as the notion encompasses both, an emotion with comparatively straightforward elicitors—core disgust—as well as a distinctly social phenomenon in the form of sociomoral disgust [Rozin et al. 2000].

2.4.5 Engineering Requirements

Viability

The requirement of the model of emotion to perform on mainstream hardware entails simplifications and further alterations of the modelled phenomena and the use of standard engineering methods to create a modular and parsimonious architecture. These design properties are valid modelling goals in themselves that do not necessarily need to correspond to features of the modelled system.

Explication of Shortcuts

Modelling activities may introduce further simplifications in addition to those due to viability concerns. One of my explicit aims is the clear delineation and characterisation of any such shortcuts deemed necessary. Two prominent examples unavoidable in current designs of broad virtual agents regard the interfaces of the agent to its environment in terms of sensing and acting and of common-sense reasoning. Subse-

quent modelling efforts may refine the model by selectively reducing the scope of such shortcuts.

2.5 TABASCO and ActAffAct

This section elaborates on my previous work as well as the TABASCO architecture that are the sources for many of the ideas put forward in this thesis.

2.5.1 TABASCO

TABASCO [Petta 2003; Staller Petta 1998] is a Tractable Appraisal-Based Architecture for Situated Cognizers, an architecture for situated agents that aims at integrating the emotion process, introducing benefits that emotion offers to agent engineering. TABASCO is of special importance to this thesis as it was a main inspiration for my previous and also the current work. It uses a layered approach and considers several variants of appraisal theories of emotion. The architecture has been implemented in an interactive exhibit, The Invisible Person [Petta 1999; Petta et al. 1999; Psik et al. 2003], and as part of a dramatic environment test-bed, ActAffAct [Rank 2004]. The interactive exhibit featured a single child-like agent on a virtual stage and was running continuously in a museum in Vienna, open to visitors all day, for nearly ten years until the end of 2008 (see section 6.5 for more details on the Invisible Person). The starting point of this line of work is the integration of insights from psychological theories on emotion with results from areas including cognitive robotics, ethology, and intelligent agents for a principled generation and understanding of expressive behaviour. The following description of ActAffAct serves as a reference for the origins of the current work as the insights gained during this work led to most of the objectives described in section 2.4.

2.5.2 ActAffAct: Appraisal Made Easy and Coherence of Behaviour

ActAffAct is mainly a test-bed for the generation of dramatically interesting stories. It consists of a simple 2-dimensional virtual space that is populated by four agents with differing personalities and several objects that carry the potential for conflicts. The four characters are modelled after dramatic archetypes, specifically: a hero/protagonist, a villain/antagonist, a victim, and a mentor. Planning capabilities in ActAffAct are derived from JAM (the Java Agent Model) [Huber 1999]—a descendent of the Procedural Reasoning System—that interleaves plan execution and plan refinement. An appraisal register, a

notion taken from TABASCO, adds the functionality of relevance detection by exploiting information used in the planning process. The resulting emotion instances in turn influence the planning process to cope with the triggered emotion. As a BDI architecture, JAM provides a plan representation language, goal- and event-driven (i.e., proactive and reactive) behaviour, a hierarchical intention structure, and utility-based action selection. This architecture was extended to fit the needs of the ActAffAct scenario.

For a viable and tractable implementation of appraisal in the toy-world of ActAffAct, a simple way to relate current behaviours to an agent's changing concerns was needed: Behaviours are categorised as either trying to achieve something, helping others to achieve, or hindering them from achieving something. This reduces a part of the runtime task of cognitive appraisal—namely assessing the relevance of a percept to one's own current goals—to simple pattern matching. The same holds for evaluation of the conformance of an action to the standards of an agent (e.g., the social norms), as these are expressed in terms of desired and undesired behaviours as well. This categorisation of behaviours encodes the kinds of social commitment they introduce, and allows the matching of sensory data, i.e., these categorised behaviours provide situational meaning structure frames.

Filling of these situational meaning structures is realised by the interpretation of current sensory data via plan-like structures that translate the outside world data into inside subjective interpretations (i.e. beliefs), taking the agent's current context into account. An example at a low level is asserting that an object sensed near the agent is reachable by it, but this may range all the way to interpreting the picking up of a flower by the agent next to itself as the anticipated possibility of being offered a present, thereby forming a positive expectation. These structures parallel what [Vosinakis Panayiotopoulos 2003] call programmable perception, in the case of ActAffAct using information about an agent's current activities. The implementation circumvents the social credit assignment problem by using custom-built interpretations that identify a responsible agent for every perceived action in the world. These use only minimal knowledge of causal structures [Mao Gratch 2003], and sometimes rely directly on high-level percepts from the simulated world.

In order to ensure overall coherence, execution of behaviours has often been split into three phases, of which the first and last are harder to interrupt, simulating commitment to one's intentions (in contrast to utility-driven control that might drop just-started behaviours as well as ones near successful completion). A timed pattern influences execution depending on the level of completion. Further, a simple approach of aggregating a mood level from the history of recent appraisals is

used to regulate the selection and execution of behaviours to enhance coherence.

The intensity of the outcome of an appraisal of a perceived and interpreted fact is expressed numerically and compared against a threshold: in case of exceeding that value, it instigates coping with this situation, and the agent's appraisal is expressed. Such expressive actions that indicate an agent's emotional state are in turn perceived and interpreted by other agents triggering further appraisals.

Coping may introduce new activities. It is the main source of variation in generated plots. Coping activities lead to action that, by way of the emotion process, is related to motivating percepts and concerns of the agent. They use the information made available by the appraisal of an event to decide on a suitable course of action to tackle this subjective interpretation. Overall, this provides for the causal relations needed for the dramatic plot of a story.

ActAffAct's (simulated) domain was tested with different setups of the four characters, one of which excluded the antagonist and thereby the main source of conflict. In the no-conflict case the resulting interactions of the characters, not surprisingly, do not deserve to be called dramatic. A qualitative evaluation of the scenario with the full cast, however, showed that minimal storylines can indeed be generated using this approach. A quantitative evaluation was not pursued as this would have required a measure of 'storyness' for the comparison of generated sequences of action, a complex research problem in its own right. Even so, it could be demonstrated how a rather simple appraisal-based architecture can address challenges in the distinctively social domain of dramatic interaction successfully.

2.6 Summary

In this chapter, the concept of an intelligent agent and the specialisation of affective agents were explained. An overview of the application areas for affective agent architectures was presented and the scenario-based method of comparing different approaches was introduced. This provides the background for part II where the construction of agents and the modelling of emotion will be examined in more detail. In addition to looking at storytelling and interactive storyworlds as a specific example of a class of scenarios with requirements that are well suited as test-beds for affective agent architectures, this chapter also presented my previous work on the ActAffAct system.

Part II

Behaviour Coordination

3 Theoretical Foundations

Where the relevant concepts of cognitive appraisal theories of emotion are explained; where the importance of emotion in behaviour coordination is motivated; where the descriptive concepts of levels, modelling primitives, and embodiment are examined.

Processes involving emotion are now considered to be an important, if not downright essential, part of the control structures of boundedly rational situated beings. They allow them to react adaptively in a dynamic, changing, and highly complex environment. They are highly flexible (compared to reflexes e.g.) and still do not necessarily require conscious or voluntary reasoning.

Emotion is about the subjective significance (including the imputed causes) of events and actions we observe. It is possible to make sense of others by using simplified models of them to explain their actions (cf. the notion of a Theory of Mind [Premack Woodruff 1978]). This is also the way in which emotions are used as explanatory parts of drama theory and as inspiration for components of different agent architectures. But what processes in an individual are responsible for emotion and the actions that are attributed to emotions?

The sections of this chapter look at the development of different theories that try to capture phenomena involving such processes. This chapter is about the uptake of different theories in computational modelling as well as about the different perspectives in AI regarding modelling of behaviour in general. At the end of the chapter, I collect criteria and dimensions from theory and from the scenario-based method described in section 2.3 to be used for the survey of related work in chapter 4.

3.1 Emotion Theories

Most words that are used in emotion research have a vast range of connotations stemming from 'folk psychology' (or commonsense psychology), a term used in psychology and philosophy of mind to denote the common understanding of mental states. The usage of words like 'belief', 'desire', 'pain', or 'anger' in natural language points to the con-

cepts and possibly a whole theory used by us to explain behaviour, of other people and of our own. See [Davies Stone 1998; Goldman 1993] for a glimpse of the controversy about what form this theory might have. In any case, the typical use of such words probably only scratches the surface of the underlying processes responsible for the experience and display of emotions that, in turn, allow us to use these terms successfully for reasoning. One of the motivations of emotions research is to clarify the nature of the responsible structures and processes.

There is no consensus in the cognitive sciences, not even in psychology, about *the* single emotion theory. Many different and sometimes contradictory theories exist. To understand the current situation, it is advisable to first take a brief look at the history of emotion research.

3.1.1 History

As is the case for many scientific endeavours, one can find a starting point for emotion theory in ancient Greece. Plato talked about a tripartite structure of the soul, distinguishing cognition, affect and conation. This separation has found its way into literature and resurfaces with different names in several theories. Aristotle on the other hand claimed that such a distinction is impossible. This conflict is mirrored in much more recent debates in the psychology of emotion about the relation between cognition and emotion [Leventhal Scherer 1987], such as whether emotion is primary and independent of cognition or secondary and dependent on it. Modern componential theories see this dispute as a largely semantic one, that vanishes if emotions are considered as processes at several levels.

One of the first theories of emotion was formulated by Charles Darwin [Darwin 1872]. Darwin's work is notable as it claims that animal emotions are homologues for human emotions, i.e. they are similar because of their shared ancestry and thus suitable objects of scientific study. He further claims the universality of 'basic' emotions across cultures and species, and cites as evidence the similarity of expressions of anger or disgust in animals and humans in the different cultures he visited.

The *James-Lange theory* [James 1884; Lange 1885] also dates to the end of the but last century. It is notable as it posits bodily changes as the primary effects and holds that emotions are only the feeling of these changes as they occur (i.e. we feel afraid because our heart races). In contrast, Wilhelm Wundt, the father of psychology in Germany, claimed based on introspection that emotion comes first. His work can also be seen as the beginning of efforts to capture emotional states with dimensional models [Wundt 1905]. The Cannon-Bard theory [Bard Rioch 1937; Cannon 1927] is another counterpoint to James-Lange. It

also puts emotions before physiological reactions and proposes the hypothalamus as the brain mechanism that is responsible for emotional responses, as well as the neocortex as the source of inhibition if appropriate. Since then, an increasing number of areas of the brain have been shown to be involved in emotions. The 'Papez circuit' proposed in the 1930s [Papez 1937] is not used as a reference anymore, while the notion of the limbic system, proposed around 1950 [MacLean 1949] as one of the three main parts of the brain apart from the reptilian brain and the neocortex, if heavily criticised (e.g. [LeDoux 2000]), is still in use today [Dalgleish 2004].

In the 1960s, the first cognitive appraisal theories arose [Arnold 1960]. Current appraisal theories will be covered in the next section in more detail. One of the early, more 'cognitive' approaches is the so-called Schachter-Singer theory [Schachter 1970; Schachter Singer 1962]. It holds that the experience of emotion depends on two factors: the physiological arousal as well as the cognitive appraisal of the arousal. An emotional experience only occurs if this interpretation of the event is emotive.

3.1.2 Cognitive Appraisal Theories of Emotion

Psychological theories of emotion have developed sophisticated models that try to explain human emotional behaviour. Although these models are formulated in the terminology of psychological research that often lacks a detailed formalisation, architectures for software agents can try to operationalise them and provide preliminary validation of the feasibility of the proposed model. Psychology has put forward several theoretical ideas about the processes of emotion in humans (see [Dalgleish Power 1999; Davidson et al. 2003; Lewis Haviland 1993] and with a different focus [Forgas 2000; Martin Clore 2001]). The one family I will elaborate upon are the cognitive appraisal theories of emotion.

Out of the various existing theories, the cognitive appraisal theories of emotion have experienced a substantial consolidation and empirical validation in the recent past. In this model, emotional experiences result from evaluations of the subjective significance of construed situations and events, according to specific dimensions or criteria [Arnold 1960; Ellsworth Scherer 2003; Frijda 1986, 1993; Ortony et al. 1988; Roseman Smith 2001; Roseman et al. 1996; Scherer et al. 2001]. A basic function of the emotional in a system is seen as the permanent assessment of the surroundings according to the entities' goals, intentions and standards: *appraisal*. An individual's situation in the environment is checked with every perception for subjective relevance. Several criteria are evaluated as detailed in the next paragraph. When a perception is found to be of (e.g., negative or positive) importance, this

does, however, not lead to a response directly; rather, another level of indirection is posited. Appraisal triggers first a *motivation* for possible further action, a so-called *action tendency*. Changes in the environment that are deemed to be of subjective relevance are addressed by proposing lines of reactions. Rather than the descriptive characterisations of emotions in dimensional or categorical models, emotion is delineated as a process which I adopt as crucial element for the survival of agents situated in complex social environments.

Implementation

For an implementation very detailed information about this process is needed. There is some consensus about what criteria are needed for the appraisal process (e.g. Scherer's 'stimulation evaluation checks' organised in terms of four objectives: novelty, intrinsic pleasantness, and goal relevance for the objective relevance; causal attribution, outcome probability, discrepancy from expectation, goal/need conduciveness, and urgency for the objective implications; control, power, and adjustment for the objective coping potential; internal standards and external standards for the objective normative significance [Scherer 2001]), but less on how such checks are exactly executed. Opinions differ also on the influence of action tendencies.

Psychological models of appraisal describe human faculties and are as such of course formally underspecified for the purpose of implementation. The design of appraisal-based agent architectures—targeting much simpler creatures—incorporates and has to refine various aspects of the psychological models considered for implementation. A prominent feature of different appraisal theories are the specific criteria used in the process of evaluating environmental change. According to [Ellsworth Scherer 2003], major subjective dimensions of stimulus appraisal include: novelty and pleasantness (gatekeepers of the emotion system, deciding whether a stimulus is noteworthy); the conduciveness of a stimulus to the agent's concerns (i.e., motivational constructs, subsuming needs and goals, see below); the potential to cope with the situation; and the social implications (an event's relation to social norms and values upheld by the agent). According to [Frijda 1986], appraisal criteria are coding categories, i.e., possible components of situational meaning structures that characterise the subjective interpretation of a situational change. At least the simpler coding categories for a current stimulus can be implicit in the perception process and therefore immediate: this would e.g. apply for the gatekeeper 'novelty'. Social implications and the identification of an agent deemed responsible for a change of the environment on the other hand can involve substantial processing.

OCC and Ortony 2003

Out of the several works on appraisal, a very influential one for implementation and application has been [Ortony et al. 1988], often referred to as the OCC model, although it is a theory. Examples of deployed systems that reference OCC as a theoretical basis include [André et al. 2000; Bartneck 2002; Bates et al. 1993; Elliott 1994; Elliott et al. 1999; Krenn 2003; Martinho Paiva 1999]. What I will call the Ortony 2003 model is a specialisation of the original theory for the purposes of building believable agents [Ortony 2003]. In a sense it tries to alleviate the complexity of emotion and the problems of misuse of the OCC model to cater for scenarios that 'only' need coherent and consistent behaviour of characters.

Ortony relates types of emotional reactions to types of emotional responses. The first scheme includes what can be subject to appraisal and according to what it may be appraised. A basic distinction is made between positive and negative reactions to a situation. Apart from that, the object of the appraisal can either be an event, an action (by somebody), or an actual object. Events might be relevant to an individual's goals, to the standards it tries to uphold, or to its tastes. Table 3.1 shows the types of reactions an appraisal can result in.

To be able to process its situation in such a manner, the individual needs to have goals, standards and tastes. To make its reactions believable, i.e. consistent, these properties have to be relatively stable and coherent. Teh goals of an individual directly affect its actions. Humans are said to have a hierarchy of goals dependent on each other, with long-term higher-level goals on top, and transient subgoals beneath. Standards and norms are responsible for moral value in the individual's world-view, and tastes or preferences give certain values to single important objects of perception. All these perceptions can also be directed inward, perceiving bodily changes or cognitive processes.

The emotion instances given in parentheses in table 3.1 are only examples that fall under the respective category, they are by no means the definitive members of their group: they are just emotion-words used in everyday language to explain reactions that might belong to this category.

Each of the emotion types is associated with a varied group of reactions or behaviours. The appraisal of an event, action, or object as belonging to one of these types triggers an inclination to use one of these behaviours, a response tendency, as Ortony calls it. There are mostly involuntary expressive reactions on the one side of the spectrum, changes in the style of internal information processing and attention, and coping activities that include complex planned actions on the other end.

Table 3.1:   Valenced reactions according to [Ortony 2003], p.194.

to...	**Positive Reactions**	**Negative Reactions**	against...
Event	...because something good happened (joy, happy-for, gloating) ...about the possibility of something good happening (hope) ...because a feared bad thing did not happen (relief)	...because something bad happened (distress, sorry-for, envy) ...about the possibility of something bad happening (fear) ...because a hoped-for good thing did not happen (disappointment, sadness)	Goal
Action	...about a self-initiated praiseworthy act (pride, gratification = pride + joy) ...about an other-initiated praiseworthy act (admiration, gratitude = admiration + joy)	...about a self-initiated blameworthy act (shame, remorse = shame + distress) ...about an other-initiated blameworthy act (reproach, anger = reproach + distress)	Standard
Object	...because one finds something appealing or attractive (love, liking)	...because one finds someone/-thing unappealing (hate, dislike)	Taste

Frijda defines an emotion as a change of action tendency (e.g. 'approach' or 'avoidance'), a change in the readiness to alter one's relation to external or internal objects. Ortony's response tendency can therefore be seen as realising a change of action tendency on the one hand, when it triggers plans and alter internal states, and on the other hand as the result of a change of action tendency, when it deploys expressive reactions. One theorised purpose of expressive response tendencies is communicative, to reveal the elicitation of an emotion directly to others that are watching.

Effects of information processing can either result in redirecting attention or changing general evaluations about other agents or values that the individual has. In Ortony 2003, *coping* is a very wide category (and a prominent term in emotion research): Problem-oriented coping is the most direct approach to solving the cause of an emotional

reaction, trying to use a plan that brings the situation under control, thus improving a bad situation or realizing or extending a good one. Emotion-oriented coping is more involved, it is directed either at oneself or another participating individual in trying to change or stabilize the situation indirectly.

Ortony 2003 also elaborates on models of personality as an important part of modelling a consistent and therefore believable character. Several models of personality exist, most of them rely on the characterisation of people along a small number of dimensions, e.g. the big five traits [Pervin 1994].

A proposal for an approach to implement a mechanism in that way is to start with only few parameters as for example the distinction of promotion and prevention focus. This could influence the choice of coping strategies directly. Promotion focus would tend to choose plans that strive for pleasure, whereas prevention would first try to prevent harm.

Concern Processing

Another central term in theories of emotion is the notion of a *concern*. It is defined as a disposition to desire occurrence or nonoccurence of a given kind of situation. This definition, taken from [Frijda 1986] p.335, is related to but distinct from goals and motives as the latter terms induce connotations of activity control. Concerns range from very concrete considerations—i.e., relating to an agent's immediate tasks—to abstract ones such as feeling competent that can lie dormant until an emotionally pertinent event takes place. The relevant characteristic of concerns is that they 'can be defined as the dispositions that turn given kinds of events into satisfiers or annoyers, into positive or negative reinforcers, for the subject or for the species as a whole' [Frijda 1986], p.335.

In the advent of a sufficiently relevant change in the perceived environment, the emotion process instigates a pervasive synchronised effort of the agent's resources, including the planning subsystem, to adapt to and cope with the perceived subjective significance of the eliciting event: to attend to or avoid, support or obstruct, pay attention to or ignore, etc., the perceived likely meaning (implications) of the stimulus event. This action tendency change, the essence of an emotional episode, may be subjectively experienced, and overtly disclosed in expressive behaviour. Frijda uses the term *relational action tendency* (RAT) to denote a state of readiness to achieve, maintain, avoid, or end a specific subjective relationship with the environment [Frijda 1986], p.70f. The term coping denotes the ensuing activities to map this actual motivating impulse to subsequent action: action ten-

dencies may be inhibited or reinforced; planning parameters updated; current plans-as-commitments revised; or new goals established; so as to adapt to the change subjectively assessed. Concerning the potential to cope with a situation as a dimension of appraisal, there is a distinction of the possibility of control of the situation in general (i.e., by any agent) and one's own power to change the situation.

Appraisal criteria constitute a simplified characterisation of the process of interpreting stimuli, not all those evaluations have to take place to warrant an effect, they interact amongst themselves, and outcomes can be refined at a later point in time. The order of these validations, whether they happen sequentially at all, or whether there are dominant combinations of criteria for certain stimuli remains a research question.

3.1.3 Aspects Most Relevant to Behaviour Coordination

When summarising the interface of the emotional process when construed as a module of an agent architecture, it has to be considered that it involves instigating a synchronised effort of different parts of an agent. Accordingly, main outputs of emotions are action control precedence signals (e.g., interrupts) that try to influence different parts of the agent, as well as the emotion process itself. Further outputs include internal awareness and overt manifestations (e.g., facial expressions, gestures, exclamations: the very expressiveness of behaviour) of the action tendency change proposed. The interface of the emotion process thus provides a structured subjective interpretation of situational changes along the appraisal criteria; preliminaries for adaptation in terms of suggested action tendency changes (i.e., action control influences); subjective awareness; and overt signalling under limited voluntary control.

Situational parameters of the emotion process consist of the agent's motivational constructs; expectations; and situational changes sensed. Dispositional factors—side-conditions of the emotion process—include: the coding categories to appraise aspects of situational change and the readiness to use these categories (as they also involve resource costs); response thresholds and response strategies (e.g., favouring antagonistic or protective behaviour); and previous experiences with different types of events. Finally, regulatory influences can be seen to implement management functionalities by constraining and shaping: the appraisal process (e.g., altering the readiness to use certain coding categories for some time); action tendency changes and admissible ways to cope with and implement them; and allowable ways of expression—such as display rules, culturally dependent rules that specify when and how it is appropriate to display certain emotions. In the context of emo-

tional processes the concept of separable modules that interact using a clearly defined interface is difficult to maintain as an emotional subsystem is characterised by its pervasive and synchronised influences on other parts of an agent architecture. Nevertheless, subcomponents might be sensibly identified as modules, as exemplified by neuroscientific research that characterises the human amygdala as containing modules for relevance detection [Sander et al. 2003].

As already mentioned, from an engineering standpoint, appraisal is under-specified, a lot of the steps from a perception to an action are unclear. There also must be an underlying model of the perceiving entity, it must be capable of remembering its goals, standards and preferences, must be able to act and perceive, to expect and anticipate. Many models use a vast simplification of these processes for the purpose of modelling; the processes that supposedly happen in a human agent do not justify a clear-cut distinction between perception and response. It could be described more aptly as a sensorimotor circuit where sensations and actions happen in a close feedback loop. Sensations are actively sought for, caused by the stimulus of ongoing action that needs input. The roles of cause and effect are constantly shifting; sensations are attained only by action and one's own actions are sensed as well. They get their specific interpretation only through the integration into a larger context [Dewey 1896; Pfeifer Scheier 1994].

I will now try to flesh out the steps of appraisal with a simple specification. The first issue will be the appraisal criteria, the Stimulus Evaluation Checks that each perception passes. Several theories exist, but there seems to be a consensus on at least the criteria in the following list:

- relevance, goal significance, focus

- standards compliance, blameworthiness

- intrinsic pleasantness, valence, appealingness

- novelty, unexpectedness, suddenness, familiarity

- responsibility

- coping potential

The first three out of the above list of evaluation criteria can be seen as corresponding directly to the three necessary value types in an agent: goals, standards, and preferences. The latter three are applicable to every appraisal. Novelty modifies the intensity of an appraisal: if the perception it deals with has not been expected, the intensity of the appraisal is higher, as opposed to for example one's own actions.

Responsibility of an action attributes causality to an agent, which is directly usable in directing coping plans or expressive actions, and for changing evaluative processing about this agent.

Coping potential represents the possible degree of control about the situation. It could be higher for perceptions that can easily be answered with several possible coping plans or for those where there is already a suitable plan in execution; when there is no plan available, the coping potential would be low. The value representing the degree of control could also modulate the intensity of the appraisal, as with novelty. But first an implementation needs a basic intensity value for an appraisal. The first three evaluation criteria can be seen as corresponding to the three possible objects of appraisal as differentiated by [Ortony 2003] (event, action or object). They may be implemented as quantitative evaluations that provide a basis for an intensity value. This, however, only provides an initial account of an appraisal intensity neglecting the dynamics of intensity. The laws of hedonic asymmetry in [Frijda 1986] point to a fundamental difference in the changes of positive and negative intensities, respectively; furthermore repeated sensation of the same type of perception should lead to a decreasing intensity.

The individual evaluation checks fill in situational meaning structures, i.e., subjective modes of appearance of a situation [Frijda 1986], p.190, by applying a mapping which normally is not unique (and thus e.g. open to reappraisal). Some theories assert a specific temporal, even sequential, organisation of these evaluations of stimuli in humans [Scherer 2001]. This sequence is hypothesised to have direct and observable effects in (e.g. facial) behaviour.

[Lazarus 1991; Smith Lazarus 1993] analyse the elicitation of emotions on two levels: on the 'molecular' level, appraisals are divided into components corresponding to the dimensions of stimulus evaluation checks. When they are combined on the 'molar' level, they are characterized by so called *core relational themes*, which incidentally bear a striking resemblance to distilled versions of plot elements in the theory of narrative. The core relational theme for jealousy for example is 'resenting a third party for loss or threat to another's affection', the one for compassion reads 'being moved by another's suffering and wanting to help'.

3.2 Modelling Primitives and Architectural Levels

In this section, to characterise the approach in this thesis, I provide a working definition of emotion based on the definitions in section 3.1 and the defining characteristics of 'higher and lower levels' in architectures for situated agents, as well as the requirements for application

scenarios that warrant the use of different levels in control mechanisms and the need for emotional processes as part of the architectures. Further terms to be analysed include decision-making; action selection; behaviour coordination; and behaviour arbitration; and their relation to other notions prominent in agent architectures. The following section will present a survey of different terminologies used to describe agent architectures, and their relevance to behaviour coordination and related processes.

Emotion, as described in cognitive appraisal theories, is characterised as interplay of processes (rather than the descriptive characterisations of emotions in dimensional or categorical models). See [Scherer et al. 2004] for a definition of an emotion episode taken from Scherer's appraisal theory. For the purposes of computational modelling, the following is a concise working description of my approach to emotion:

> Emotions involve continuous processes that evaluate stimulus events according to subjective criteria, and, for events of subjective relevance, influence continuous activity in a synchronised fashion.

A stimulus event, in this context, is understood as an interaction between the agent and its world, possibly comprising environmental and internal changes.

A computational model needs to provide all of the infrastructure of a complete agent [Pfeifer Scheier 1999] that is explicitly and implicitly implied by descriptions of the appraisal process. The involvement of multiple continuous processes points to this embedding in a complete agent architecture: it has to allow the synchronised recruitment of resources. Further, the definition of emotion in [Scherer et al. 2004] delineates it from other affective phenomena, like moods and affective dispositions, that interact through said continuous processes in a complete agent.

In computational models of emotion, such processes are relevant in 'higher and lower levels' of an architecture and especially to link them. The separation of higher and lower levels is a helpful simplification that needs to be re-evaluated regularly; cf. that some theories of human competencies distinguish two separate systems (planning and control) even for seemingly low-level motor processes [Glover 2004]. Characteristics that can be used to differentiate levels in an agent architecture are the time-scale of operations at different levels [Gat 1997], the temporal regularities of representations used, and the degree of independence from the outside world (as reflected in the reification of the notion of duration at higher levels). Operations at a lower or reactive level involve more direct coupling to an agent's environment, the time-scale of operations corresponds directly to the temporal resolution of

the agent's interface to its world, and no symbolic representations need to be involved. Examples of processes on higher levels include any planning, that uses a representation of possible actions, or counterfactual reasoning, that can detach representations from the current state of the world. The notion of representation has stirred many controversies in behaviour-based AI [Brooks 1991] and cognitive science [Clark 1997]. Relevant properties of higher- and lower-level representations are the rate of change and whether they refer to internal or environmental conditions.

Having distinguished higher and lower levels, the different influences of emotional processes on these interacting levels, with different scopes of competence, involve mechanisms for evaluating interactions on lower levels to provide meaningful information for higher levels and vice-versa. Further capabilities are the regulation of the information flow between the levels and the recruitment of already existing abilities on the different levels to effect internal and external changes.

Distinctions between higher and lower levels only make sense in the context of a specific agent and its relation to its environment. I try to capture this context for the description of affective agent architectures as scenario descriptions, see section 2.3. Scenarios for the use of agent control architectures can range from simple virtual worlds with one task and a very restricted set of interactions to very complex virtual worlds and robot scenarios that target human-like competencies. I argue that the complexity of a scenario is an indicator for the existence of higher and lower levels on the one hand and emotional processes on the other. For the purpose of this thesis, I want to point out some requirements for scenarios that warrant different architectural levels and emotional processes and are relevant to behaviour coordination: *situatedness* (including temporal criteria for task achievement), possibly *ambiguous and unreliable sensing and acting*, multiple *conflicting tasks*, *variability of resources* and resource usage. The next section focuses on terminology for the concepts involved in behaviour coordination.

3.2.1 Architectural Terminology

In the following, action and interaction are used as non-specific terms for any execution done by an agent, interaction stressing the coordinative nature of agent-world interaction [Petta 2003]. [Clancey 2002] emphasises the role of activities in situated behaviour, as the analysis of behaviour that is focussed only on tasks neglects (or abstracts away from) relevant parts of real-world behaviour. Different sets of research have touched upon this very topic and many ideas have already been formulated in Activity Theory [Leontev 1978] and also at least as early

as in [Dewey 1896].

Situatedness denotes the predicament of being in a world. This implies that an agent has to obey the spatial and temporal conditions of the (simulated) world, but it also entails that it can exploit what the environment does for it. Further, it means that such an agent cannot choose not to act, so there is no need to continuously control the agent on a low level of single operations where its actions are already structured by the needs of situated activity and routine functioning [Bargh Chartrand 1999; Chartrand Bargh 2002].

For the notion of *concern* and its relation to goals and motivations, see section 3.1.2. The term *motivation* is used to describe the force that starts, controls, and sustains internal and external interactions. The relation between concerns and motivations can be understood as the passive and the active part, respectively, of the same component. When a motivation to do something ceases, a relevant concern would continue to exist.

The autotelic principle [Steels 2004] points to a form of motivation different from reward for autonomous agents that posits the balance between challenges and skills as the motivating force inside the agent. This can be regarded as an elaboration of the concern to feel competent. Another general principle useful in thinking about concerns is homeostasis.

An *operation* is the basic unit of an agent's executive. Such a single interaction is always embedded in a structured *behaviour*, a compound interaction that can run in parallel with others and in an unsupervised fashion. A behaviour can, but does not need to, be used in order to achieve a goal, where *goal* is a state and the achievement can take the form of bringing about the state, but also avoiding or maintaining it. I use the term *task* as the more general term for the specification of what to do. This can take the form of goals but includes also qualitative criteria, specific activities, or generally patterns of behaviour.

The number of tasks, possible conflicts, and interactions between them are fundamental factors of the complexity of a scenario. The relation between the time-scales of tasks on the one hand and primitive operations on the other is an indicator for the need for high-level behaviours such as planning and the use of hierarchy in the organisation of behaviours [Aylett Jones 1996].

I think of all behaviours as embedded in *activities*, the broader context of what an agent is currently doing. The activities of an agent are generally not goal-directed but only provide usual behaviours or behaviour sequences. Examples of such an activity could be to 'lead a conversation' or 'to go shopping' with typical more goal-directed behaviours such as 'acquiring an information' or 'acquiring a specific item'. Other typical behaviours sustain the current activity: 'turn-

taking nods', 'small-talk', 'scanning a shelf', or 'checking the shopping list' during shopping. The concept of a *lifeworld* [Agre Horswill 1997] that captures the regularities of an agent's activities is covered in section 3.3.3.

Appraisal (see section 3.1.2) denotes the subjective evaluation of stimulus events against the context of concerns and abilities of an agent. The identification of concern-relevant changes is followed by the creation of a *relational action tendency* (RAT). This term denotes a state of readiness to achieve or maintain a given kind of relationship with the environment [Frijda 1986]. A situation-driven (and not a goal-oriented) meta-level process, it tries to (re)configure the current activities and behaviours. The role of these emotional processes can be understood as embedded human mechanisms for committing an agent to interactions in line with these action tendencies in order to *cope* with the appraised influences by employing higher and lower levels. Furthermore, RATs are manifested in unconditional affective expression and thereby directly observable by other agents, a prerequisite for social awareness. If the current activities do not fit the needs of an action tendency, it may remain inconsequential unless problem-directed coping activities adopt the originating appraisal because of a RAT's intensity and instigate new goal-directed behaviour.

Another important part of an appraisal-based agent are *regulation* strategies [Gross John 2003; Ochsner Gross 2005] that influence the appraisal and coping processes. Regulation comprises, e.g., cultural and individual display rules (including specification of circumstances when to allow, or even fake emotions) and mood-dependent influences. [Dawkins 2006] gives a short report about a person with an autistic condition who claims to be able to better analyse the difficulties that animals face in highly disturbing settings. This could indicate a deficit in regulatory abilities. Opaque simulators might be used to perform parts of appraisal and planning operations, cf. the use of Theory of Mind models in PsychSim [Marsella et al. 2004].

A fundamental aspect of this conceptualisation of an affective agent architecture is the parallelism of interaction. This poses the problem of control as one of management of asynchronous components and, thus, motivates the use of the term behaviour coordination.

3.2.2 Embodiment

Embodiment has been accepted as a crucial aspect of intelligent and emotional behaviour and it is a central notion in the research programme of cognitive science [Anderson 2003; Clark 1997; Núñez 1999]. Computational applications, however, have interpreted the notion of embodiment in vastly different ways (see e.g. [Martin et al. 2005]). In

this section, I start with a critique of different views of embodiment in cognition [Wilson 2002] and present their implications for the design of affective agent architectures, especially those that need to interact directly with humans.

Considerations of embodiment and situated cognition have spawned a lot of research efforts, especially in robotics. These have originally been inspired by a rejection [Brooks 1991] of classical approaches towards Artificial Intelligence. Integration attempts that seek to combine approaches to embedded and embodied cognition—often called behaviour-based—with 'higher-level' capabilities such as counterfactual reasoning and long-term planning face several problems. Methods from symbolic AI deal with scenarios that can be characterised as problem-solving in which the goal, as well as the problem description and the produced solution, are represented in a human-interpretable code. Agent architectures that directly consider the theoretical notion of embodiment are often targeted at environments in which fast decisions and 'unreflected' perception-action capabilities are needed [Arkin 1998], i.e., tight sensorimotor coordination, a term going back to [Dewey 1896], see also [Pfeifer Scheier 1994]. [Wilson 2002] tries to dissect the claims that have been presented using the notion of embodied cognition. Among these is the claim that human cognition leverages embodiment by off-loading work into the environment. This connects embodiment to the importance of an agent's physical and social lifeworld (see section 3.3.3).

I argue that control is a crucial aspect of embodiment when applied to complex software agent architectures [Jones Wray 2006]. Embodied agents in a sufficiently complex environment are limited by constraints about tractability and real-time behaviour. One such constraint is the amount of resources (time, attention, effort) that should be used for the different issues the agent faces, i.e. the arbitration between different concerns. An architecture for embodied agents has to coordinate influences arising from the environment and those stemming from the agent itself. High-level planning, e.g., cannot be the locus of control in such an agent, it has to be dissolved into the complete agent design A principled implementation of embodiment in software agents entails giving up high-level control. This means that no single component of an agent architecture can be assumed to have complete control of the agent (see also section 3.3.2). This connects embodiment to recent trends in continuous planning [Avradinis et al. 2005; Rank et al. 2006].

Models of emotion in cognition and action, that can be incorporated into an embodied agent, further need to consider the role of bodily information in emotion. [Frijda 2005] details reasons why bodily information is central for all emotion experience, but also that other components such as affective valence are needed to complement this infor-

mation. I argue that these distinctions are important to computational modelling and can fruitfully be applied to affective agent architectures.

3.3 The Action Selection Problem and Why It Is Not

The title of this section is a tongue-in-cheek reference to its conclusion: that 'action selection' is not a good term for the dominant problem of an autonomous agent as it evokes the association of a series of decisions of the same kind; it is rather the problem of coordination and orchestration of an agent's behaviour that is concurrent in time and space. To motivate this argument, this chapter introduces the issues surrounding the implementation of synthetic behaviour.

3.3.1 Synthetic Behaviour

Action selection is the term that is often used to characterise the most basic problem of intelligent systems in general: *what to do next?* An action selection mechanism is then described as determining an agent's actions [Brom Bryson 2006]. While action selection mechanisms are often thought of as a layer on top of a behaviour system or sometimes as the whole architecture, the term mostly evokes the picture of a series of isolated decisions. Regarding the design of affective intelligent agents for realistic scenarios, this is the wrong kind of problem statement.

Decision making is another term for the same or at least a related problem in intelligent agents. Classical definitions of decision-making are anchored in a symbolic conceptualisation of intelligent action:

'Decision making is the process of choosing a preferred option or course of action from among a set of alternatives.' [Wilson Keil 1999]

Affect has varied influences on behaviour coordination in humans [Loewenstein Lerner 2003], ranging from the influence of immediate emotions, i.e., the emotions experienced at the time of decision making, to the role of expected emotions in evaluating options. But human, and in general situated, activity involves different processes of control that do not necessarily need conscious decisions between alternatives. Related problems are action selection [Tyrrell 1993] or behaviour selection [Gadanho 2003]. The latter term stresses the fact that what is selected is not a single and atomic action but a collection of current continuously running situated activities. In [Tyrrell 1993] action selection is defined as the problem of 'choosing at each moment in time the most appropriate action out or a repertoire of possible actions'. In order to stress my conviction that for situated activity there is no selection at

'each moment in time', I prefer to use the term *behaviour coordination* (used e.g. in [Althaus Christensen 2003]) to denote the management of ongoing activities of an entity at different levels (which subsumes decision making and action selection).

For the design of agents such as synthetic characters, the first and foremost property to consider is situatedness—the predicament of being in a world. This implies that a character has to obey and is exposed to the spatial and temporal conditions of the (simulated) world, but it also entails that it can exploit what the environment can do for it. Further, it means that such an agent cannot choose not to act, so there is no need to continuously control the agent neither on a low level of single actions, where its behaviour is already structured by the needs of situated activity and routine functioning [Bargh Chartrand 1999], nor at the level of high-level deliberative acts, as these are at the least infrequent, and at the most only a concept conceived in retrospect. The main problem for implementing the synthetic behaviour of an agent then is the allocation of its bounded resources to the coordination of its behaviour [Schut et al. 2004].

The importance of clarifying terminology for agent architectures (see section 3.2.1) is paramount: of both, the concepts describing the agent and its behaviour, as well as the concepts used when implementing the behaviour. For example, the term 'behaviour' is itself as often used to denote the overall behaviour of the agent as well as low-level behaviour*s* of the agent. Overviews of different agent architectures often use a rough classification of architectures, for example, discerning deliberative, embodied, and hybrid architectures. In practice, however, no architecture used in a real scenario falls neatly into one of these categories. In this section, I will therefore cover different aspects of implementing synthetic behaviour, such as planning and lifeworlds, without restriction to a specific class of architectures.

3.3.2 Challenges for a Situated Agent from a Planning Perspective

Resource-bounded agents that are situated in complex environments (such as the interactive storyworlds I propose in section 2.4.1) face challenges of uncertainty, non-determinism, conflicting goals, and social interaction. Planning processes and emotional processes can be key assets in this context and the different perspectives of these processes on the tasks of situated agents are complementary in a whole architectural design.

Complex environments pose several planning challenges that have been disregarded by early planning research: knowledge about the environment might be incomplete or wrong; actions may fail; results of actions may come undone; goals may be achieved to different (partial)

degrees; and there are multiple conflicting goals to be pursued. As a consequence, goals can be of several different types, such as avoiding, achieving, or maintaining a state of the world, as well as verifying or improving upon it. In order to cope with the dynamics of such an environment—including exogenous effects—, planner and executive have to run concurrently, or at least in some interleaved fashion [Doyle 1996].

In the design of such situated agents, motivation-driven continuous planning systems [Avradinis Aylett 2003] are needed: behaviour-based architectures [Brooks 1991] that employ a planner as a separate supportive resource or layer that is not always needed for the agents' functioning, cf. [Ferguson 1992; Jung Fischer 1998]. A behaviour in this context denotes a rather self-contained activity that can run in parallel with others and in an unsupervised fashion, i.e., without central control. The original idea formulated by Brooks also included that these behaviours do not employ any kind of representation at all. This radical view also gave rise to different conceptions of the term representation, as e.g., action-oriented representation [Agre Chapman 1987; Clark 1997]: only locally effective, personalised, indexical structures that stand in stark contrast to classical objective ungrounded symbols.

The challenges that a situated agent faces are more varied than uncertain knowledge and non-deterministic actions. As mentioned above, a rich and dynamic (social) environment entails many and conflicting goals that an agent will have to pursue; different types of goals and partial goal achievement have to be accounted for. Furthermore, goal achievement might be tied to successful social interaction with other agents, and the interplay of commitments and (social) conventions has to be considered [Glass Grosz 2000; Grosz et al. 2004]. Against such a scenario, plans are entities that are mainly used, rather than constructed: while plans abstract away from individual action, they are still a kind of action-oriented representation. This view of plans is informed by the notions of plans-as-commitments and plans-as-communication, augmenting the idea of plans-as-routines, i.e., basic building blocks of higher-level action. The *plans-as-commitments* view regards established intentions as filters for new options to be considered in a continuous execution environment [Pollack 1992; Pollack Horty 1999], thereby containing the need for reasoning. Commitments also influence the estimation of possible costs of new options [Horty Pollack 2001], e.g., when a new option requires only a minor modification of a current plan. The *plans-as-communication* view [Agre Chapman 1991] introduces abstract plans that do not immediately control execution. Instead, they are structures that need interpretations to bind them to a substantial amount of run-time contextual information, achieving effective refinement and re-use. Such indirect employment of (meta-)control information is compatible with hybrid architectures that use deliberative mod-

ules as support for reactive ones: an interpretation might e.g., cause the (re-)configuration of an architectural layer.

An architecture for situated agents has to coordinate influences arising from the environment and those stemming from the agent itself. This is another reason why a planning module for such an architecture has to run concurrently—either continuously, or in an anytime fashion with support for being suspended. To be safely stopped and (re-)started at arbitrary times, a planning module should support partial plans, be capable of plan adaptation[12], in addition to anytime characteristics, ensuring that the module is not subjected to a functional view of solely transforming input to output, but that it allows timely interaction and interleaved execution, a necessity for resource-bounded agents [Doyle 1999].

An important question when thinking about an agent as including a planner, is what has to be in control of the agent at different times. The most radical conception of an integration of planning capabilities into complete behaviour-based architectures would view a planner as simply another behaviour at the disposal of the agent. The functionality corresponding to an emotional process model, however, can probably not be captured adequately by a single behaviour. In the interface of such a planning module, one could distinguish parameters and management facilities. The latter comprise all commands to steer its activity, e.g., *start*, *stop*, *suspend*, and *inspect*. The input parameters of a planning module include descriptions of goals the agent wants to achieve and its current beliefs about the state of the world. Furthermore, the expected utilities of possible action outcomes have to be available. These utility values can in turn be computed using smaller building blocks of state characterisations and may take into account the costs of activities. Additional inputs consist of general and domain-specific heuristics, as well as criteria of success that further differentiate the description of goals by requiring specific degrees of goal achievement or qualitative criteria that concern the process of achieving the goal. Outputs do not only consist of partial or completely refined plan structures, but also include the options abandoned because determined to be unusable at the current moment. In particular, such abandoned plans (and their unfulfilled preconditions) can be instrumental in monitoring the environment for new options. Further, metering and profiling information includes the types and number of alternatives considered, the current level of detail of the plans returned, and the expended effort.

---

[12] The classic term 'plan repair' reminds of the original static and closed environments, where modification of plan structures was the exception, rather than the rule.

3.3.3 Lifeworlds

To analyse the actually possible and necessary behaviours for a specific scenario, I use the concept of a lifeworld [Agre Horswill 1997], to enrich the notion of an environment by focusing on the patterned ways in which a physical environment is functionally meaningful within some activity. Such a lifeworld analysis can be leveraged not only to restrict the complexity of the agent-environment interactions (as necessary in a simulation), but also to identify relevant regularities during design. In the case of emotional interaction, the analysis of the social lifeworld [Rank Petta 2005a] enacted by agents (see below) uncovers the relevant interactions and stresses the requirement for awareness of social relationships in agents.

By operationalising these theoretical notions, I de-emphasise the role of high-level cognition ('thinking') in routine functioning [Bargh Chartrand 1999] and recognise the opportunity offered by available structures to constrain high-level function within tractable bounds. Concerning this scaffolding, [Agre Horswill 1997] identify abstract locatedness and functionally significant relationships grounded in the physical environment. Analogously, social lifeworld analysis considers the potential for interaction with respect to loci of control at the macro level (e.g., power and status [Kemper 1993]) as well as indirect access to (second and higher level) resources (e.g., [Aubé 1998]).

In order to recognise the current and the future opportunities and threats of a situation, agents employ a constant evaluation of their surroundings according to their individual concerns: appraisal ascribes meaning to sensed changes. Emotional processes—ranging from raw affect under rough and undifferentiated circumstances, over fleets of feelings in (yet) unclear scenarios, to fully articulated 'emotions proper' as results of detailed perception—mediate the translation between the subjective worlds of concerns and preferences, the current state of activity in the abstract enacted shared social lifeworld, and the status and offerings of the physical world [Rank Petta 2005a].

Contrary to its physical surroundings, the intangible social lifeworld an agent is embedded in has to be continually enacted and negotiated. My focus on sociality combines the goal of reducing the cognitive load for individual entities populating the environment with emphasising the relevance of coordinative functions [Clancey 1999] mediating between an individual and the (potential) current and future opportunities and threats to satisfy an individual's concerns.

A situated agent's dependence on regularities thus extends beyond the physical world into socio-cultural constructs, whose maintenance can e.g. be modelled as an interplay of conventions (social norms) and evaluative processes (emotions) [Staller Petta 2001], with emo-

tions sustaining social norms, and culturally defined social norms in turn shaping and regulating emotions (e.g., with feeling rules defining which emotions are suitable in which situation, and display rules providing repertoires of how to express them).

Interpretations of situations and developments in the social lifeworld are not a given: both within an individual and in the society, they are the outcome of negotiations and transactions, captured e.g. in sociological models [Kemper 1993] or characterised in terms of personality traits. Purposeful functions such as threatening, sanctioning, and amending, therefore are intrinsic behavioural requirements, along with their affective grounding in the social lifeworld.

The focus on sociality also relates lifeworlds to the concept of interactive storyworlds. The enactment of the social lifeworld is the sphere of activity that is dominant in the context of dramatic interactions. The mechanisms of emotional processes for interpreting and sustaining this lifeworld and their influence on motivation form a fertile ground for drama.

3.3.4 The Arbitration of Behaviour

A single action (actuating or sensing action) is the basic unit of an agent's executive. Such action is always embedded in a structured behaviour, a compound action that can run in parallel with others and in an unsupervised fashion. This ansatz parallels the ideas put forward in, e.g., [Bryson 2000] and [Loyall et al. 2004], of a compromise between plan-like structuring of action and a completely parallel architecture, often using intermediate levels of action patterns.

In accordance with the idea of lifeworlds, I think of all behaviours as embedded in activities, the broader context of what an agent is currently doing. The activities of an agent are generally not goal-directed but only provide groupings of usual behaviours or behaviour sequences. Examples of such an activity could be to 'lead a conversation' or 'to go shopping' with typical more goal-directed behaviours such as 'acquiring an information' or 'acquiring a specific item'. Other typical behaviours sustain the current activity: 'turn-taking nods', 'small-talk', 'scanning a shelf', or 'checking the shopping list' during shopping.

Figure 3.1 shows a conceptual diagram of an appraisal-based agent architecture. The Executive is intended to capture the structure of situated activity. The process of Appraisal covers the interpretation of perceived changes. Perception is the process translating outside information to inside information. It fills in situational meaning structures that are strictly subjective. Note that the resulting possibility of misunderstandings between agents is a desideratum for dramatic scenarios.

Identification of concern-relevant changes is followed by the cre-

Figure 3.1: Conceptual diagram of an appraisal-based agent architecture. The Perception, Expression, and Coping processes all fall into the realm of the Executive; they are singled out due to their special importance.

ation of a relational action tendency (RAT). A situation-driven (and not a goal-oriented) meta-level process, it tries to (re)configure the current activities and behaviours. Furthermore, RATs are manifested in unconditional affective Expression and thereby directly observable by other agents [Reisenzein 2001], a prerequisite for social awareness. If the current activities do not fit the needs of an action tendency, it may remain inconsequential unless problem-directed coping activities adopt the originating appraisal because of RAT's precedence and instigate new goal-directed behaviour.

Another important part of an appraisal-based agent not represented in the diagram are regulation strategies [Gross 1999; Gross John 2003; Ochsner Gross 2005] that influence the appraisal and coping processes. Regulation comprises, e.g., cultural and individual display rules (including specification of circumstances when to allow, amplify or suppress or even fake emotions) and mood-dependent influences.

3.4 Key Aspects of Interest

For the survey of related work in chapter 4, it is necessary to point out those aspects of the theory covered so far that are pertinent to the evaluation of systems similar to the current one. Section 2.3 argued for the comparison of affective agent architectures based on a

scenario analysis. However, in part due to the youth of the field, necessary information for a complete comparison of this kind is often absent from publications. This section collects suitable criteria for a feasible comparison from the scenario concept, from the requirement and objectives presented in section 1.2.1 and from theories of emotion and about behaviour coordination.

- Regarding the target scenario, the overall *purpose and motivation* of the implemented system provides a first categorisation of different systems.

- The *interaction qualities* of the agents modelled provide an important reference for evaluating the potential of a given system. The requirements adopted for the current thesis can also be used to examine related work: Does the system try to avoid *symbolic sensing and acting*? Is *real-time interaction* with other agents or human users an issue in the design? Does the embodiment, if applicable, allow for *multimodality*? If the agent is modelled in a virtual world, are the physical properties of this world respected (and used) by the agent? Another requirement for many emotional phenomena to even occur, is the existence of multiple and potentially *conflicting tasks* for the agents in the system.

- What is the *status of emotion* in a specific system? Since there are several system relevant to this research area that do not explicitly refer to emotion at all, a first basic distinction is if the system acknowledges emotional phenomena explicitly. Concerning the implementation, an important question is if emotions are *reified* entities or if another approach is used that leads to emotional phenomena as, e.g., system-level properties.

  - What functions of emotion does the system try to replicate? What emotional phenomena are of interest and are they stated explicitly (rather than targeting emotion in general)?

  - For the simulation of affective processes, an agent needs to be motivated to act (or to choose not to act). How is the *intrinsic motivation*[13] of agents modelled in the system?

- What role does *social interaction* play in the scenario of the system? Is it possible, essential, avoided, or ignored?

- What *functional decomposition* does the agent architecture use? Are there different *levels of processing* or can the architecture

---

[13] Note that this motivation, an implementation detail of agents, is different from the motivation of scenarios which relates to the reasons for building the whole system.

be assigned to (mostly) one end of spectrum of processing levels? Does the runtime conceptually use a single *locus of control* or is the control distributed in some way?

3.5 Summary

In this chapter, a treatment of the use of levels in agent architectures and the importance of embodiment was presented. The importance of lifeworld analysis and of boundedness, i.e. constraints in time and other resources, are important to consider in the construction of an agent control architecture. This chapter focused on theories of emotion. The next chapter considers how implementations connect to the the theoretical notions presented and to the key aspects of interest identified.

4 Survey of Related Work

Where the uptake of emotion theories in computational modelling is examined; where different agent architectures and their application scenarios are introduced; where implemented mechanisms that involve emotion are pointed out.

In this chapter, I cover a range of agent architectures, applications and systems that are related to modelling emotion. Not every architecture mentioned here is explicitly referring to emotion as a reference point or even claiming that it corresponds to emotional phenomena, but all of them attack the problem of creating synthetic behaviour. In the interest of scientific validity, it would be preferable to cover only those architectures whose implementation details are available for further scrutiny. Unfortunately, this is the case only rarely: often the details given in papers are too vague and the actual source code is not available.

As mentioned in the previous chapter, a complete scenario analysis of all the related work is not feasible. The architectures and applications mentioned here span a wide range of targeted scenarios. On the theoretical end of the spectrum, research efforts aim at flexible architectures for agents that incorporate the understanding and possibly the imitation of emotions. Some projects try to implement a whole framework and tools that can be used in applications that would benefit from knowledge about emotions or from the incorporation of emotional processes. Other projects are oriented towards applications in the pedagogical domain or in the area of interactive fiction. Therefore, the first section of this chapter will introduce the mentioned systems based on the motivation for building them and their purpose, as far as it is known. Later sections highlight specific systems that are relevant to the key aspects of interest collected in section 3.4.

4.1 Overview

Many of the systems relevant for this thesis are naturally those targeting the understanding of emotions. The purpose of early systems was often modelling the understanding of emotions in written text or the generation of emotional text. Examples of such systems are BORIS, its

extension OpEd and DAYDREAMER [Dyer 1987], as well as THUNDER, a program for ethical reasoning about the motives in textual stories [Reeves 1991].

ACRES [Frijda Swagerman 1987] is another system that was directly motivated by the wish to understand emotional phenomena. However, the approach was to create an artificial purpose for the system: it was designed as a command line computer program that can be told and subsequently queried about emotions. While fulfilling its main task it also tried to satisfy other concerns as, e.g., the one not to be killed and not to be kept waiting.

WILL [Moffat Frijda 1995] was developed as a successor to ACRES. For the implementation a different but still suitably restricted test-bed purpose was chosen: playing the prisoner's dilemma [Moffat 1997].

The Affective Reasoning project is a collection of different undertakings to explore the possibilities of computational emotion. The thesis [Elliott 1992] presents the *Affective Reasoner* (AR) itself, an 'emotion engine' based on OCC [Ortony et al. 1988], and the purpose of the system was the generation of emotional responses: it was used in the simulation of emotional taxi drivers. However, the AR project has seen many different applications and was extended to serve for real-time interaction and applications with pedagogical purposes [Elliott 1997a] as well as to analyse emotional structures in stories [Elliott 1997b].

AR is also a direct inspiration for the *EMA* model [Gratch Marsella 2004a,b; Marsella Gratch 2006, 2009]. EMA (short for emotion and adaptation) is a framework for modelling emotion that has seen many different applications and has evolved over time to be used for different purposes. Most notable among those is the Mission Rehearsal Exercise (MRE) [Hill et al. 2001; Swartout et al. 2006], as it involved the integration of many different technologies including a virtual 3D environment, speech and language technologies [Traum et al. 2004]. MRE was funded by the U.S. army and intended to teach decision-making skills to soldiers. The 'virtual humans' in this project incorporate EMA as a model of emotional appraisal and coping mechanisms implemented in the reasoning system SOAR [Laird et al. 1987]. Similar virtual humans using SOAR had been used previously in the virtual tutor Steve [Johnson Rickel 1998]. Precursors to EMA include work on emotional planning like Émile [Gratch 1999, 2000], and applications of pedagogical drama [Marsella et al. 2000].

All of the systems mentioned so far serve, at least in part, directly the purpose of enhancing the understanding of human emotions. This does not apply to the rest of the systems below but there is one more piece of related work that fits this category. *COPE* [Zuckerman Gagne 2003] does not describe an agent architecture or even any system at all, but 'an inventory of coping strategies', a scale measure for testing

coping strategies (COPE) employed by subjects in experiments. It is interesting in the context of this thesis, as its purpose is to classify coping behaviours.

The *Oz project* at Carnegie-Mellon University aimed at a different area of application: to facilitate the creation of interactive fiction including synthetic characters [Bates 1992]. Intelligent emotional agents that live in virtual worlds should act as believable characters. The underlying system used to generate emotions, called Em [Reilly Bates 1992], was therefore a means to an end. The Oz project led to a range of scenarios and architectural designs, including a realtime animation version and a text-based one similar to classic computer games. The latter gave rise to Lyotard, a simulated house cat. An early manifestation were the Woggles in the installation Edge of Intention [Bates et al. 1993], ball-like creatures that display emotions by their morphing movements and do not act according to a predefined narrative. The original Oz project has been discontinued as of December 2002. It was, however, the birth-place of a range of different follow-ups, including Façade and the creatures of Zoesis.

Façade [Mateas 2002; Mateas Stern 2005] tries to achieve an interactive drama in a realtime 3D world. It was publicly released in July 2005[14]. The player experiences a narrative from the first person perspective, taking the part of a longtime friend of a married couple that invited him or her for dinner, an occasion where problems unfold. The goal of the project is a very rich interaction, not only through dialogue but also through embodied interaction, to create a dramatically interesting world populated by emotional agents. [Loyall et al. 2004] describe *Gertie and WB*, a programming language and motion synthesis system, developed and used at the company Zoesis, that are useful for authoring interactive characters. These systems are, similar to Façade, successor to the behaviour authoring system developed in the Oz project.

Another group of projects targets the interactive storytelling scenario using planning formalisms developed in AI. In such an application, the actors' roles are formalised in a planning system that continually monitors execution and performs dynamic replanning according to user actions. A project at the University of Teesside, called *Interactive Storytelling*, is aimed at creating dynamic narratives capable of reacting to the interaction with a user. [Cavazza et al. 2002] describes a system that uses real-time hierarchical planning techniques to simulate the roles of actors. *Haunt2* [Magerko et al. 2004] is a similar attempt to create a game in which AI characters are central to the game experience. The goal of the game is to escape a house by influencing

---

[14] See the website http://www.interactivestory.net/ for different download options.

other characters indirectly.

Several projects use the term virtual human to describe the goal of their research, i.e., developing sufficiently advance ECAs that can be used in diverse applications. *SimHuman* is one platform for such intelligent virtual agents, [Avradinis et al. 2004] describes a hybrid agent architecture for believable behaviour to be applied in the SimHuman environment. An important part of this architecture are mechanisms to model and handle emotions, mainly by perceiving and generating signs of emotion. Another model of emotion that is targeted at virtual humans is the *PME* (Personality-Mood-Emotion) model [Egges et al. 2003, 2004] developed at MIRALab. The intention of this model is to simulate personality, mood and emotion in virtual humans. Two more systems of the same kind, but developed in different virtual human research projects, are ALMA [André et al. 2000; Gebhard 2005] and WASABI [Becker-Asano Wachsmuth 2008].

Another system that can be described as modelling virtual humans is *Twig* [Horswill 2008]. Twig is relevant to the topic of this thesis not because it models emotion but because it uses a scenario that is very similar to the one proposed here. One or two children and a parent are modelled, the behaviours that the system should believably generate are related to the attachment and caregiving behavioural systems (see section 6.1).

On the boundary between entertainment applications on the one hand and the AI topic of behaviour control architectures on the other, the Synthetic Characters Group at the MIT Media Laboratory has developed an architecture for a synthetic 'brain' called C4. Two interactive installations have been implemented using C4: 'Sheep|Dog: Trial by Eire' and 'Clicker' [Burke et al. 2001; Isla et al. 2001], both starring a virtual dog the user may train using vocal commands. Similar to earlier developments for Hamsterdam [Blumberg 1994], their focus lies on the imitation of animal behaviour, which reduces complexity compared to modelling a mind for virtual humans. However, the same group also applied their work to more human-like creatures: '(void*) : a Cast of Characters' was an installation presented at SIGGRAPH99. It featured creatures the user could dance with using a simple interface. The different characters reacted according to their personality and explicitly modelled emotional states. The presented system also included a separate camera and music creature that dynamically directed the virtual camera and the musical accompaniment respectively [Tomlinson 1999].

The Affective Computing division at the MIT Media Laboratory explores possibilities to enhance computing in general by including emotional information. This includes communicating affective information from and to the computer, sensing, and recognizing affective reactions in humans also by using new interfaces. Synthesizing affect in

machines is also one of their research interests, albeit not the main one. Projects include the attempt to incorporate affective signals into a gameplay mechanics based on the gaming engine Quake2 and the development of an Affective Learning Companion. This latter agent is intended to assist children in their learning efforts by recognizing their affective states and responding accordingly, encouraging them, posing questions, or pointing [Kort Reilly 2002; Kort et al. 2001a,b].

Several applications in the area of *robotics* are mentioned throughout this chapter, as the objectives set for the current work are closely related to those of behaviour control in robotic applications. The *Remote Agent* architecture [Gat Pell 1998] is targeted at use in the control of autonomous spacecrafts and [Gat 1997] describes a general blueprint for *three-layer architectures*. The *Humanoid Robotics* Group at MIT is dedicated to research about robots with human qualities. As far as emotions are concerned, these robots mainly try to either recognize them in humans or to imitate them in speech and gestures. The *Cathexis* model [Velásquez 1998, 1999] and its application in Yuppie, an emotional pet robot (a Yamaha puppy), goes beyond their aim by focusing on basic control circuits for specific emotions, rather than higher level emotion processes. [Avila-Garcia Cañamero 2005] test their behaviour coordination mechanism in the *competitive two-resource problem*: two robots face the problem of managing their two competing survival-related needs. The metaphor of hormonal feedback is used to modulate perception with affect in this architecture. The *TAME* (Traits, Attitudes, Moods, Emotions) framework [Moshkina Arkin 2003] for affective robot behaviour has been applied in scenarios including a robotic dog that protects its human owner. [Althaus Christensen 2003] discuss the problem of behaviour coordination in the context of robotic navigation applications. [Barnes et al. 1997] present the *Behaviour Synthesis Architecture*, an example of a robot control architecture that has to solve issues of behaviour coordination at the lower level. The latter two are examples of work in behavioural control that do not mention emotion or affect at all, but that are still relevant to the current thesis.

Just as work in robotics can provide many insight into issues of behavioural control, more general AI control architectures developed for purposes very different from the modelling of emotion highlight aspects that are critical for a principled approach towards the simulation of affective behaviour. One example is the work on the *Edmund architecture* and Parallel rooted, Ordered, Slip-stack Hierarchical (*POSH*) action selection [Bryson 2000, 2001, 2003], originally applied to a virtual grid-world foraging scenario but then also to robot control. Other approaches explicitly acknowledge the importance of accounting for emotion: *PECS* [Urban 2001], an acronym for Physis, Emotion, Cognition, Social Status, is described as a reference model of a component

oriented architecture for decision-making and behaviour control; *Emotional pathfinding* [Donaldson et al. 2004] modifies a classic pathfinding algorithm to take into account the emotional state of the agent; and [Gadanho 2003] describes the use of emotional processes for learning of behaviour coordination in the architecture *ALEC* (Asynchronous Learning by Emotion and Cognition).

Finally, it is helpful to relate the current work to the existing accounts of human minds as a whole. Aaron Sloman leads the very broad research effort at the University of Birmingham, entitled the Cognition and Affect project [Sloman 1999]. Their aim is not to search for one specific architecture that implements a specific capability comparable to a human one. Rather they want to map the space of possible agent designs and their applications ('design space' and 'niche space' in terms alluding to evolutionary biology), always considering as broad an architecture as possible, even if it may remain shallow, or only be a reference blueprint as the *H-CogAff* architecture. The minder-scenario is one of the test-beds developed. In this synthetic world, one agent is the minder or nursemaid, whose task it is to look after 'baby'-robots in a limited area, keeping them out of trouble until they are mature enough to leave this nursery. Sources of trouble are e.g. ditches the babies could fall into or their innate need to be recharged regularly. Dedicated research efforts focused on the relations of such an architecture to high level emotions like grief, seen as a state of perturbation resulting in a partial loss of control [Wright et al. 1995]. Furthermore, Sloman addresses the question how a well-funded conceptualization of emotion could result out of such architectural constraints [Sloman et al. 2005].

Another high-level architectural account of the human mind very similar to H-CogAff has been proposed by Marvin Minsky, in the book *The Emotion Machine* [Minsky 2006]. He argues for the importance of which sets of resources are used for which problems and that emotional mechanisms are involved in the selective activation and deactivation of these sets. The proposal postulates six levels of mental processes that reuse similar mechanisms for increasingly complex tasks. *EM-ONE* [Singh 2005] is an implementation based on Minsky's proposal that focuses on common-sense thinking in a physical scenario with two agents.

Another general model of human intelligence is SOAR which, as mentioned above, is also the basis of EMA. The analysis in [Marinier Laird 2006] presents affect as a natural component of the *PEACTIDM process* of SOAR. The acronym captures the fundamental processing stages of the SOAR architecture control loop: Perceive - Encode - Attend - Comprehend - Tasking - Intend - Decode - Motor. Based on the assumption of this mostly fixed sequence, appraisals (automatic and

deliberate ones) can be understood as an inherent part of the agent architecture, in terms of the stages of comprehension and coding.

PSI [Bartl Dörner 1998; Dörner 2002, 2003] is a psychological model of the human mind that originated at the University of Bamberg. [Bartl Dörner 1998] describes an evolution of the PSI model that features 'cognitive processes that are modulated by emotional states and processes'. The primary goal of the architecture is to provide a theory of action regulation, and the term 'motives' is used to describe the roots of the system. MicroPSI [Bach 2002, 2003; Bach et al. 2006] is a more recent implementation of the concepts of PSI. The main test scenario for the creatures developed with MicroPSI is an island-like environment similar to classic foraging scenarios. Other applications have turned to PSI as a guide to implement affective agents. For example, [Lim et al. 2005] describes a system that includes a virtual tour guide that uses an emotion model based on PSI to allow for different emotional personalities in its presentations.

4.2 Analysis of Key Aspects

In this section, the different systems introduced in the previous section will be analysed in terms of the key aspects of interest laid out in section 3.4. Not all the systems are relevant to all the aspects so only a selection will be mentioned in the different subsections.

4.2.1 Emotional Focus and Targeted Phenomena

There is a fundamental difference between psychological research and computational applications. Based on the notion of a scenario, as presented in section 2.3, you could say that psychological research has a very different motivation to follow and a different purpose to fulfill. Nevertheless, models of affect are used in both areas and can be very similar, notably in the subset of human emotion that they target.

The *Affective Reasoning* project is based on OCC and as such shares its claim to cover most of the spectrum of human emotions. The virtual Taxi-World that the original implementation was used for restricted the emotional displays possible to the prototypical emotion labels used by the OCC theory for its 24 different emotion types.

The scenario of use for the implementation of *WILL* was playing the prisoner's dilemma. While the theory that was the basis for WILL covers human emotions in their entirety, the expressive actions to choose from in the prisoner's dilemma as implemented are limited to influencing the future strategy and to choosing from a set of textual expressions. Examples of emotions simulated in WILL are happiness, anger and pride.

EMA as a framework for modelling emotion claims to be domain-independent, i.e., it also needs to cover the whole range of emotion possible in the current target domain. The theoretical focus of EMA are coping activities defined as the inverse operation of appraisal, i.e., the identification and influencing of the believed causes for what has been appraised as significant in the current context. Appraisal and coping in EMA are mechanisms that operate on a causal interpretation of ongoing world events in terms of past, present, and future states and actions. The actual coverage of emotional phenomena is therefore directly dependent on the granularity and resolution of the planning representation.

For the *Oz project*, emotion was a means to the end of enabling the creation of interactive fiction. The phenomena targeted therefore depended on the possibilities of interaction that the system should be capable of and, rather than simulating human emotion, synthetic characters 'only' needed to believably emotional according to their role. For the ball-like *Woggles* creatures, emotions are shown as changes in their morphing movements. Although the Woggles focus on engaging interaction with characters and not on emotionality, [Bates 1994] stresses the importance of appropriately timed and clearly expressed emotion as a prerequisite to turn software agents in virtual worlds into *believable agents*, a term coined by Bates, alluding to the notion of a believable character used in the art of writing. Such believable characters give the 'illusion of life' which allows the 'suspension of disbelief' for the audience, the ultimate goal of Hollywood-style cinema. Similarly, Façade's goal is such a rich interaction with believable characters. Even though Façade is a descendent of the Oz architecture it does away with the emotion module and only uses carefully crafted displays of emotional interaction. Instead, there is a separate component in the system that monitors the conformance to a theory of drama, i.e. a drama-manager that influences all agents to achieve a coherent structure of the narrative according to dramaturgy.

Projects that aim to develop virtual humans, by definition, need to cover a wide set of emotions. However, as the research focus of virtual humans often lies with the production of visually believable behaviour, the simulation of affective behaviour is sometimes neglected or added as an afterthought. Such models are often based on the OCC categorisation or use a representation of emotion as a low-dimensional vector. One such example is the use of the PAD (Pleasure, Arousal, Dominance) space in the ALMA and WASABI systems.

The robot control architecture *Cathexis* is based on 'affect programs', a term that represents evolutionarily successful stimulus-response patterns. These are to correspond to the so-called basic emotions, a disputed concept in the psychology of emotions. For the robot Yuppie, the

affect programs corresponding to the following emotions were implemented as behaviour systems: anger, fear, distress/sorrow, joy/happiness, disgust, and surprise. The scenario of the *competitive two-resource problem* is used to test behaviour coordination mechanisms based on hormonal feedback and only allows for very rudimentary 'robot versions' of emotion, by design. In the introduction to [Arkin 2005] the author argues that, because of a ethological background, the use of the term motivation normally replaces emotion in many robotic architectures. He characterises emotions as a subset of motivations that provide support for an agent's survival. Consequently, the effect of what is called emotions (and traits) in the *TAME* framework is limited to affecting gain values in motor control in a task relevant way.

PSI refers to emotions as modulators in a motivated agent architecture for 'action regulation'. The PSI theory is notable on the one hand as it uses a different and idiosyncratic terminology compared to other psychological theories of motivation and emotion, on the other hand because it does not use emotions explicitly but constructs an agent architecture that consists of motivators and modulators with parameters that vary over time. The interaction of these components produces complex behaviour that can often be interpreted as emotional As such, the theory does not exclude any human emotion from the range of possible phenomena modelled. The implementation of the model also includes a mapping of the current agent state onto an human-like face that uses eight dimensions named after emotion labels to determine the expression.

The other general models of human intelligence and their accounts of emotion, by design, do not limit themselves to modelling specific emotional phenomena. This holds for the analysis of emotion as part of SOAR's *PEACTIDM* process as well as for Sloman's *H-COGAFF* architecture blueprint and for Minsky's *Emotion Machine*. Similarly, the addition of an emotion model to a virtual human normally tries to cover emotion in its entirety, e.g., by relying on a representation based on a dimensional space of emotions as in the *ALMA* and *WASABI* models. *PECS* claims to be a reference model of a component oriented architecture for decision-making and behaviour control and as such also does not limit itself in any way as to what emotional phenomena it can account for.

COPE is another example that, by design, tries to provide an exhaustive coverage of emotion, in their case of the coping strategies associated with emotional behaviour. The classes for coping behaviours they mention are *problem-focused*, *support-seeking*, and *emotion-focused* coping, later extended by *emotional processing*, and *emotional expression*.

On the contrary, the *Affective Learning Companion*, even though it

uses a fairly general dimensional modelling of emotion, suitably limits itself to emotions relevant in learning situations. Since the agent is intended to assist children in their learning efforts, it only needs to model those affective states and the relation to suitable intervention strategies.

The *emotional pathfinding* algorithm refers to emotion only as three distinct states (fear, hunger, and thirst) that are seen as inputs. The processes that lead to these states in the first place are not considered as relevant to the problem at hand.

4.2.2 Sociality

The social nature of many emotions renders scenarios of emotions in social interactions an obvious target. The virtual taxi-world of the *Affective Reasoner*, for example, is inhabited by many agents that have different personalities and try to fulfill goals like earn money, avoid getting robbed, and avoid getting speeding tickets. A central modelling element of the AR agents are so-called Concerns-Of-Others databases to model other agents they encounter for up to twice removed reasoning. Similarly in *EMA*, the causal interpretation of ongoing world events in terms of past, present, and future states and actions includes the actions and states of others and most of the example scenarios that EMA has been applied to feature such representations of social constellations. The first main scenario of the Oz project, the installation Edge of Intention, used the constellation of three different personalities as a central element. The same holds true for the descendent project, Façade.

However, for some systems that explicitly target emotional phenomena, the initial scenarios do not regard social interaction as a central element. This applies to many architectures for robotic applications, such as *Cathexis* that relies on the so-called basic emotions, or the scenarios of the *hormonal modulation* architecture that features two competing robots that do not interact or perceive each other directly at all. Another example is the *MicroPSI* model whose initial test scenario is a solitary agent on an island.

4.2.3 Multimodality and Real-Time Sensing and Acting

Part of the central objectives presented in section 1.2.1 were a focus on multimodality and the consideration of real-time human-computer interaction. The different systems mentioned in this chapter have very different approaches to these aspects.

The *Affective Reasoner* has been applied to many different scenarios many of which have been exclusively based on textual interaction. Text-

based interaction can be used for real-time interaction but it does not allow for genuine multimodal interaction. Newer incarnations of AR agents can interact with users in realtime using multi-modal perceptive and expressive channels including speech-recognition, text-to-speech, morphed schematic faces and music [Elliott 1997a].

ACRES and *WILL* are two more examples of systems that rely on exclusively textual chat-like interfaces. Many more recent ECA interfaces use a asymmetric interaction mode that presents a multimodal display in the form of the embodied agent but receives information from a user as typed text only. Emotion models such as *ALMA* or *WASABI* are used in such ECA interfaces mostly to inform the visual presentation but they can only appraise changes in the sparse input information.

One of the central claims of the *EMA* framework for modelling emotion is domain independence. The central method for achieving domain independence is reliance on classical planning representations of a causal interpretation of world events. Thus, from the point of view of the emotion model, the modality for interaction with the rest of the system is this representation. Further, the underlying architecture SOAR assumes centralised symbolic representations as the foundation for interacting with the environment. Such an approach leads to strong partitioning of the architecture and a clear separation of central processing and input and output paths. This, by itself, does not preclude real-time interaction applications and multimodality as the different projects that EMA has been customised for show. However, the approach is different from the one taken up in this thesis and relies on the assumption of fast and sufficiently independent coding and decoding processes to connect to the environment. [Marsella Gratch 2006] provides an example of how EMA can be applied to situations that involve rapid dynamics in interaction.

Similarly to the central representation used in EMA, *plan-based interactive storytelling* systems rely on the planning representations to connect to an agent's emotion model. Further, such systems often use a complete game engine as a rapid prototyping platform and thus inherit the specific simulation of agent modalities that the engine supports.

The *Oz project* focused on interaction with humans from the start. The agents created should act believable in the real-time interaction with humans. The modalities of interaction, however, were still limited. *Façade* is the one descendent project that tried to enrich the interaction considerably in the direction of a real-time 3D world, but it also abandoned the emotion model used in Oz and used hand-crafted interactions to achieve believable, i.e. dramatically appropriate, displays of emotion.

Robotic architectures, per definition, deal with real-time interaction. The modalities supported and the possibilities for interacting with

humans depends on the particular sensors and actuators used. The approach towards simulating real-time interaction and multimodality in this thesis can be seen as an attempt to reproduce the richness that robotic platforms potentially need to deal with while at the same time avoiding the complexities that can reasonably be simplified. In the virtual world, simulating another modality is cheap compared to testing and adding another type of sensor and actuator.

The *Affective Computing* division at MIT, however, is doing just that, exploring new sensing and actuating possibilities to enhance computing albeit without the integration into the coordination processes of an autonomous robotic architecture. [Nehaniv et al. 2002] focuses on how the meaning of sensory information for an agent depends on the interaction of the sensing entity with its environment, i.e. its coupling to the environment. Sensing in itself does not provide meaningful information but a channel for meaning is connected with the appropriate processing, as for example the usefulness of the sensing to satisfy homeostatic and other needs. In the terminology of some emotion theories, this corresponds directly to the current concerns of an agent.

Broad accounts of human minds, such as *H-COGAFF*, Minsky's *Emotion Machine*, or *PSI theory*, often need to disregard questions of real-time interaction and information specific to modalities. Similar to the SOAR-based EMA and to interactive storytelling systems based on planning formalisms, they can postulate a central form of representation or they abstractly refer to the importance of modality-specific systems. Only the application to concrete scenarios can choose to value real-time interaction and multiple modalities as central elements.

Finally, the *Twig* system is interesting as its approach to simulating modalities is very similar to the approach of this thesis. In the case of Twig, physical interaction is realised by using a simplified physics simulation. However, the system's purpose is generation of animations of believable behaviour (as opposed to realistic behaviour). On the other hand, this thesis tries to establish a system to model cognitive, perceptive and activity processes in a relatively realistic way, using shortcuts and simplifications where necessary, disregarding any disadvantages that might result from that approach regarding believability of the generated animations. A practical consequence of this difference is the use of such a simplified physics simulation in Twig, while the system described here employs a standard physics engine that makes e.g. believable character movement comparatively harder to achieve.

4.2.4 Functional Decomposition

The most important aspect of the current work is the investigation of coordination and control mechanisms suitable for modelling emotional

phenomena. The functional decomposition of systems can provide insights into the status of emotional states and/or processes in a control architecture.

The *Affective Reasoner*, following the OCC model, uses 24 discrete emotion types. The intensity of emotion instances is determined by several different variables, and emotions that were elicited can be expressed through different channels, ranging from involuntary somatic reactions—like turning pale—to highly conscious plans like taking revenge. The agents also use so-called Concerns-Of-Others databases to model other agents they encounter for up to twice removed reasoning. Possible response actions are selected out of 'action databases' influenced by the personality of the agent. The process structure of AR differs considerably for the different application domains that range from offline text analysis to real-time control.

WILL is an example of an emotion model that focuses on the architectural perspective and is based on the appraisal theory of emotion of Frijda. The model consists of separate modules running in parallel on a shared database of facts, a blackboard. All facts the system knows about are regularly checked for relevance against the concerns of the system. Relevant ones are then subjected to secondary appraisal, in the terms of Frijda, possibly leading to a change in action readiness, i.e., the probability of specific actions changes. Emotion and personality are seen as closely related, emotions being described as temporal inconsistencies, personality as temporal consistency.

As described above, *EMA* is a model of emotion using appraisal and coping mechanisms implemented in SOAR. In EMA, coping is defined as inverse operation of appraisal, i.e., the identification and influencing of the believed causes for what has been appraised as significant in the current context. Appraisal and coping in EMA are mechanisms that operate on a causal interpretation of ongoing world events in terms of past, present, and future states and actions. This interpretation is based on classical planning representations, i.e. symbolic action representations (these actions are high-level actions), plans, and task networks, namely linking actions and states. Appraisal augments this causal interpretation with appraisal variables. Actions are partially ordered, can have duration, and asynchronous effects. The rest of the architecture is derived from SOAR (Newell's Unified Theory of Cognition) with several special purpose operators updating the mental state including evaluation check operators. EMA, therefore, uses the same basic control mechanism as SOAR, i.e. a single evaluation loop that is expected to complete several times per second.

The analysis of affect in the *PEACTIDM process* is also built on SOAR. However, in contrast to EMA, appraisal is not implemented as a separate module but as an inherent part of the agent architecture, in

terms of comprehension and coding, two of the fundamental processing stages. The acronym captures the stages of the SOAR architecture control loop: Perceive - Encode - Attend - Comprehend - Tasking - Intend - Decode - Motor. Based on the assumption of this mostly fixed sequence, appraisals (automatic and deliberate ones) can be understood as part of the encoding and comprehension stages.

In the control architecture used by the *Oz project*, the emotion component *Em* is based on a subset of the OCC model and integrated with the behaviour control component Hap to form the overall architecture called Tok. The Em system includes standards and attitudes and generates discrete emotions when sensing events that conflict or conform with these.

Façade abandons the use of an explicit emotion model in favour of components that coordinate the selection of appropriate behaviours above the level of single agents. There is a separate component in the system that monitors the conformance to a theory of drama, i.e. a drama-manager that influences all agents to achieve a coherent structure of the narrative according to dramaturgy. The drama is managed as a 'bag of beats' that structures the plot events to achieve a dramatic curve that builds up tension until a breaking point is reached, and tries to find a resolution at the end. [Mateas Stern 2004] presents *ABL* (A Behaviour Language), the language used for implementing the believable agent layer of the interactive drama Façade. ABL is the successor of Hap used to implement the character part of Façade. It allows sequential and parallel behaviours (although a selection mechanism orders the basic actions again), and reflective programming. While performing these behaviours, reactive moment-by-moment interaction has to be possible. An activity is represented as a goal, i.e. a desired state, and each goal is supplied with one or more behaviours to accomplish its task. A behaviour itself is a series of sequential or parallel steps, there is a pre-written behaviour library and a tree of currently active behaviours. Single acts are implemented externally to ABL and form the interface to an agent's environment. *Gertie and WB* are the programming language and motion synthesis system used by Zoesis for authoring interactive characters including emotion states, and they are thus also successors to Hap, as well as Em and Tok. The system is described as a behaviour authoring system, behaviour being used as the general term for the activities of agents. There are active goals that can be executed by running behaviours (not used as the general term here) in service of goals, and issuing parameterised acts to the low-level system. Gertie is described as a general purpose programming language with special features for concurrency, reflection, continuous monitoring, backtracking, and action and sensing. Similar to ABL, the behaviours of Gertie need to be grounded in acts and sensing

expressions. Acts are the domain of WB (warping and blending) the motion synthesis system. There are several overlapping acts issued and completed or aborted each second. Behaviours (basically functions in Gertie) and acts can be annotated with priorities and mutual-exclusion information. There are elaborate failure handling possibilities, similar to exception handling, but allowing recurrent actions, or random selections. The reflection capabilities are instrumental for the emotion system to generate emotion instances in response to failure or success of actions that have been marked by the behaviour authors as emotionally important. Continuously monitored (context) expressions provide reactivity where needed.

In a *plan-based interactive storytelling* system, a real-time planning system provides the central control mechanisms that continually monitor execution and perform dynamic replanning according to user actions. As described in section 3.3.2, traditional planning system disregarded runtime constraints in favour of complete problem analysis. A continuous planning system therefore needs an integration with the rest of the system that allows complete replanning on each invocation or strategies for deciding how to interleave with other runtime components.

The *SimHuman* platform for intelligent virtual agents is described as including a hybrid agent architecture organised in different layers [Avradinis et al. 2004]. A list of their generic design entities mentions low-level emotional responses; encountered situations that concern goals, standards, preferences, and expectations; other processes like memory, perception, decision making, and learning. Actions that are processed by their physical layer are either extrovert or introvert. The cognitive layer of the architecture also includes higher-level processes, like appraisal of situations, belief revision, or causal attributions.

Such a layered architecture is prevalent in robotic applications. The *Remote Agent* architecture for the control of autonomous spacecrafts is composed of an executive that translates high-level commands into low-level ones in the context of an unknown dynamic environment. The plan-like commands are abstract descriptions of parallel activities, the term 'timeline' is used to describe separate state informations used during execution. Such timelines are subdivided into 'tokens' which are able to react to time-dependent contingencies. This executive capability is a also step away from open-loop execution of operations as in traditional planning towards reactivity.

The *Cathexis* model aims to base higher level cognitive emotions on 'affect programs', a term that represents evolutionarily successful stimulus-response patterns. It, therefore, focuses on basic control circuits for specific emotions, rather than higher level emotion processes.

Cathexis is described as a system that consists of separate nonlinear computational units, that feature inputs, outputs, and an 'appraisal mechanism'. The latter mechanism filters data and identifies special conditions by exciting or inhibiting its containing system. Behaviours are equated with goal-directed 'activities', the emotional system is responsible for providing the 'motivational context' of behaviours used for arbitrating conflicts with other behaviours. Behaviours issue low-level motor-responses (operations), more than one 'activity' is allowed at one time if the behaviours are non-conflicting in terms of resource usage.

[Althaus Christensen 2003] presents the prevalent approaches to behaviour coordination in the context of robotic navigation applications. The simple task that is posed to the agents in their scenario entails only a single activity. The evaluation of sensory information is only done at the level of single behaviours. The term behaviour is used, in the tradition of behaviour-based AI, for continuous translations from sensory input to motor output. The agent uses a (pre-existing) list of tasks to be accomplished in order to achieve goals. A single task is implemented as a composition of behaviours. Their approach to coordinating behaviours distinguishes between behaviour *arbitration*, i.e. switching between behaviours at discrete points in time, and the *fusion* of the output of several behaviours. Their method integrates these two types by using a dynamical systems approach that combines a value for the 'competitive advantage' of a behaviour (depending on sensory input) and its 'competitive interactions', i.e. the degree to which other active behaviours suppress it.

[Avila-Garcia Cañamero 2005] use the terms behaviour selection and action selection as synonyms. As common strategies for this selection problem they cite winner-take-all and voting-based approaches. Their model is separated into a 'motivational' and a 'behavioural' layer linked through the 'physiology'. Behaviours are described as coarse grained subsystems (embedding simpler actions) that implement different competencies and these are further categorised as either consummatory or appetitive, defined as either achieving or preparing the achievement of a goal.

The *Behaviour Synthesis Architecture* is another example of a robot control architecture for behaviour coordination at the lower level. At this level the tasks of agents are constrained to a single task in order to restrict complexity: specifically, in this case the cooperation of two robots to transport an object. A behaviour is defined by a behaviour pattern that not only includes a translation from sensory input to motor output but also a utility function depending on sensory data. As the robot has only one output channel, i.e. it can only execute one behaviour at a time, the values for different behaviour pat-

terns are linearly superimposed. As related problems on successively higher levels, the architecture refers to behaviour conflict resolution, behaviour adaptation, and behaviour scheduling. For the execution of longer tasks in changing environments the use of behaviour adaptation and scheduling is expected to be needed.

In the *TAME* framework, a 'personality and affect' module sits on top of a behaviour-based robot architecture and alters its output. The effect of emotions and traits on robot behaviour are modelled and implemented as gain values in motor control. The components modelling affect are only used to enhance behaviour selection and behaviour coordination. The control architecture consists of behaviours and a coordination mechanism, primitive behaviours can be parameterised and may be combined into behavioural assemblages. This structures could correspond to the notions of operation and behaviour presented in section 3.2.1. The description of TAME also uses the notion of currently active behaviours. The affective modules are implemented as another set of behaviours that output parameters for the primitive behaviours instead of motor control vectors. They also mention the technical detail of clamping parameter output to reasonable (predefined) values, which might be interpreted as a function of regulation.

The *PSI theory* is notable on the one hand as it uses a different and idiosyncratic terminology compared to other psychological theories of motivation and emotion, on the other hand because it does not use emotions explicitly but constructs an agent architecture that consists of motivators and modulators with parameters that vary over time. The interaction of these components produces complex behaviour that can often be interpreted by an observer as emotional [Bartl Dörner 1998] describes an evolution of the PSI model that features cognitive processes that are modulated by emotional states and processes. The primary goal of the architecture is to provide a theory of action regulation, and the term 'motives' is used to describe the roots of the system. The 'motive selector' chooses one motive and turns it into an intention, which can be 'run' to achieve 'the respective goal'. *Modulators* can parameterise all these cognitive processes, e.g., the *resolution level* influences the degree of exactness of data comparisons, and the *selection threshold* determines the strength of defending the current intention against competitors.

The *C4 cognitive architecture* consists of a perceptive subsystem that receives only input the creature could realistically get from its subjective point of view. It includes a history of percepts and the expectations about future ones, thus enabling the system to be surprised about novelty in its surroundings, which is also one of the key appraisal criteria in emotion theories. At the end of the processing stages lies a complex navigation and motor component. The system also features

a reinforcement learning system. High level emotions like pride, hatred, or gratitude are not used, as the architecture is mainly intended to model a dog's mind, and, of course, because the rest of the architecture provides big enough a challenge already.

The functional decomposition is the central element of most accounts of human minds. For the *Cognition and Affect project*, the aim is not to find one specific decomposition that corresponds to the human mind but to map the space of possible designs. Their reference blueprint, the H-CogAff architecture, discerns three different layers of processing and three different stages on all layers that correspond to input and output as well as central processing. The role of affect in this kind of architecture corresponds to interrupts that can cross the boundaries between levels. Marvin Minsky's *emotional machine*, as already mentioned, discerns six different levels of mental processes. Apart from this abstract categories, he mostly avoids concrete architectural descriptions focusing on mechanisms that are applicable on all levels. One example is the duality of critics and selectors, mechanisms that propose a set of possible processes and then weed out less suitable contenders.

A very influential source for layered control architectures in general is [Gat 1997]. The author describes a blueprint for *three-layer architectures*, consisting of a controller (a reactive feedback control mechanism), a sequencer (a reactive plan execution mechanism), and a deliberator (for time-consuming computations). These are conceptualised as parallel processes. In this framework, a primitive behaviour or skill is a transfer function, one or more of which are active in the controller. Gat poses constraints on these behaviours that parallel my characterisation of a lower level: e.g. minimal use of internal state and constant-bounded time and space complexity. The sequencer is performing behaviour coordination. Each level is operating at a different time-scale, e.g. the deliberator's time-scale is much larger than those of the other two components.

The *ALEC* architecture is separated into an emotion component and a cognitive component. The agents abilities rest on small hand-designed low-level behaviours. The term 'task' is used to describe the overall behavioural requirements for the agent in the scenario. The emotion component contains a goal-system that is responsible for monitoring a set of homeostatic variables. Deviations in this goal system are used to drive behaviour selection and a derived value of 'well-being' is used for reinforcement learning. The cognitive layer uses the same inputs, i.e. the state of the goal system and reinforcement values to learn decision rules based on past experience.

[Bryson 2000] juxtaposes completely parallel architectures and those that exploit hierarchy for action selection. The *Edmund architecture*

compromises reactivity and commitment by using a hierarchical control structure. Behaviours in this architecture are a group of related functions for sensing and acting complemented with any state information needed. This notion is similar to my definition of a behaviour, but in Edmund behaviours do not run without supervision. The lowest level of the executive in Edmund is a primitive action or sensing predicate, they composed into action patterns (a set sequence of actions). A competence denotes a more flexible structure: a prioritised collection of action patterns or other competences with triggers. The root of the behaviour hierarchy is a collection of drives that are processed in parallel.

PECS is described as a reference model of a component oriented architecture and thus focuses exclusively on functional decomposition into a set of smaller model components. The model consists of three layers, the first holding Perception and Sensor components; the second consisting of Social Status, Cognition, Emotion, and Physis; and the third sporting Behaviour and Actor. The latter two hold the repertoire of possible actions and the action selection process. PECS agents can possibly include planning processes which are said to determine in an abstract way which activities have to be undertaken in order to reach the given goal. There is mention of a 'currently triggered activity' although it is unclear if this is related to the notion of activity, as presented in section 3.2.1. PECS is a suitable example of the kind of premature specification of architectural detail without reference to application scenarios that the approach presented in this thesis tries to avoid.

4.2.5 Representation of Motivation

Apart from the general functional decomposition, one aspect is especially important for agents in a system that allows for modelling of emotion: the representation of motivation in the agent architecture. Affective agents need to be motivated to perform activities as a prerequisite for any emotional phenomena. Without the motivation to do something, there is no way to evaluate the subjective significance of changes. Thus, how such motivations are implemented, and how explicit and accessible such representations are for the system, gives an indication of the status of emotion in the system.

The *Affective Reasoner* adopts the notions of explicit goals, standard, and preferences as used in the OCC theory. The main motivation of AR agents is, thus, to follow their goals and respect their own standards. The same hold true for the agents developed as part of the *Oz project* and the descendent architectures of *Façade* and *Zoesis*. Façade is notable in that any motivations of a single agent can be completely

replaced at the end of a dramatic beat and motivational consistency over longer time-scales is not the responsibility of the agent anymore.

WILL is based on the appraisal theory of emotion of Frijda and directly implements its notion of concerns. Architecturally, concerns are symbolically represented setpoints that are used to evaluate all facts as soon as the system knows about them for relevance. Relevant ones are then subjected to secondary appraisal, in the terms of Frijda, possibly motivating a change in action readiness, i.e., the probability of specific actions changes.

Both *EMA* and the analysis of of affect in the *PEACTIDM process* inherit the motivational mechanism of SOAR based on explicitly represented goal states. EMA shows how classical AI planning considerations can be extended when emotion theories are taken into account. The basic symbolic representations do not change, but appraisal augments the causal interpretation with further variables and introduces several special purpose operators updating the mental state including the evaluation check operators.

In the context of the *SimHuman* platform, [Avradinis Aylett 2003] proposes a planning system centred around the notion of motivation. Motivations are described as emotional states or long-term, high-level drives that affect a situated agent's plans and actions and are, in turn, affected by these actions or by changes in the environment. A view that also features in [Aylett et al. 2000] and [Avradinis et al. 2005] which summarises the view on motivations and continuous planning in virtual environments. Motivations in this context are very similar to my conceptualisation as the active part of concerns, see section 3.2.1. In this continuous planning framework, actions are the basic execution unit, and plans combine them into higher level structures. Actions have an explicitly authored relation to named motivations that they are beneficial or detrimental for.

In the *Cathexis* model, drives are the motivational system that is activated by deviations of variables from their setpoints and the emotional system is responsible for providing the 'motivational context' of behaviours used for arbitrating conflicts with other behaviours.

The *hormonal modulation* architecture also uses an explicit 'motivational' layer that refers to homeostatic variables defined in the 'physiology' component of the robot that correspond to basic concerns of the agent that influence the other layers. Motivations are defined as abstractions that represent tendencies to behave in particular ways when a variable deviates from its ideal value.

PSI theory uses the term 'motives' for the motivational roots of the system. The system itself is described in metaphors, e.g. the motivator is compared to a water-tank that is sensitive to the level of a homeostatic variable. There are basic needs (for water and energy), and there

are informational needs such as the needs for certainty, competence, and affiliation present in any agent (cf. the notion of basic concerns [Frijda 1986]). Similarly, the emotion component of the *ALEC* architecture relies on a monitoring a set of homeostatic variables.

The root of the behaviour hierarchy in the *Edmund architecture* is a collection of drives that are processed repeatedly and in parallel. This persistent collection of drives can also be compared to a set of concerns that can be either active as motivations or passive by only checking sensing predicates.

4.3 Summary

This chapter gave a short survey of the implementation of emotion theories in computational models. As many systems do not refer to emotion theories explicitly, this overview also presented several systems that are indirectly related to the modelling approach in this thesis. A complete scenario analysis of the presented systems was not possible, but they were presented with consideration of the key aspects of interest identified in section 3.4. The final part of the thesis applies some of the concepts that are said to be components of emotional experience to an agent architecture.

Part III

Affect, Processes, and Resource Management

5 Building Blocks for Computational Modelling

Where the procedural primitives and the details of the chosen implementation are described. Where the modelling of the limitations of an agent is explained.

According to [Barsalou et al. 2007]: coordination between processes is probably the most complex part of understanding human intelligence. Based on this assumption, the goal of the system described here is to arrive at a principled implementation of coordination processes in a cognitive architecture. The foundation of this implementation, as with any cognitive architecture, are specific building blocks. This chapter presents the procedural aspects of the building blocks, parallel processes, and also introduces resources and resource management as the second foundation for the architecture.

The guiding idea for the choice of primitives is to avoid imposing an arbitrary static structure on the architecture as far as possible. The implementation is in a way designed from the outside: I start with the environment and the idea of a 'body' that is always active (or 'running' in software terminology). Interaction with the environment, including humans, i.e. 'interference', can happen at any point in time, and it does happen continuously. This rules out any architecture that assumes a conceptually singular locus of control, as such an architecture might *block*, i.e. stop interacting. This leads to the assumption of a multi-processing architecture, with more or less independent processes that need to be coordinated.

Another important goal is to explicate any shortcuts that were necessary to arrive at a feasible implementation. Among other things, these shortcuts necessarily include:

- Predispositions that a natural organism would develop during its lifetime, i.e., I need to assume agents that are developed to a certain level of maturity;

- Perceptual modalities and motor activities are only simulated in a crude form, with the guiding idea to simulate the complexity of the processing involved;

Several types of scenarios (including related purposes of use) are already 'excluded' by these aims, in the sense that they are hard to achieve. Examples involve any scenarios that call for realistic modelling of human physical behaviour; as a consequence, physical movement of agents is simulated as a caricature of human movement. On the other hand, an important part of the definition of the scenario are the regularities in the environment: in the current case, the most basic source of regularity is the use of a physics engine to drive all movement, which is thereby constrained to be relatively realistic (as compared to cartoon animation for example).
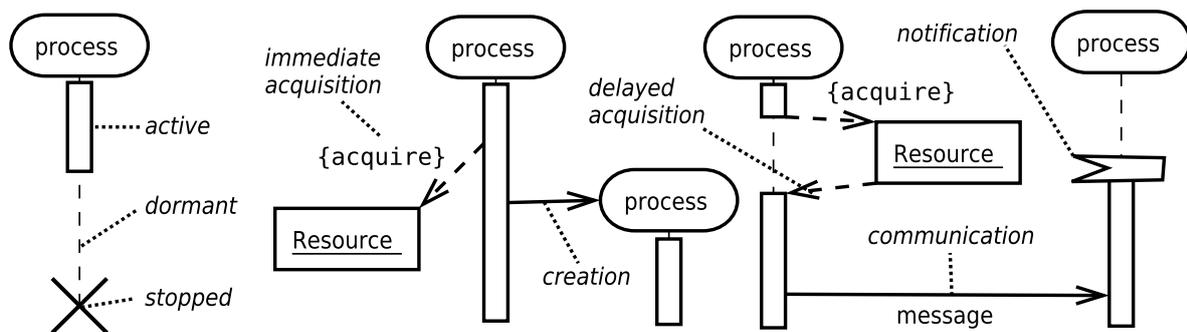
5.1 Modelling Primitives

Figure 5.1:   Examples of the notation for processes and resources.

I propose resources and concurrent processes as building blocks for the implementation of models of emotional phenomena. Figure 5.1 shows an exemplary graphical notation of the basic features of theses processes and resources based on UML activity diagrams.

I use concurrent processes directly to model the activities within an agent, thereby avoiding the use of parameterised states to model effects that arise immediately from interactions of processes over time. An example of such a state is a resolution level that determines the level of detail in comparing perceptions to an agent's knowledge (cf. PSI theory in section 4.2.4). Modelling the comparison process as active and limited in time and resource use allows for the same effects as a resolution level simply by the influence of the activities of the rest of the system that limit the comparison process and force it to come to a result earlier or with less resources. The resource bounds of the modelled agent are represented explicitly by distinct resources that

need to be acquired explicitly by processes for use. The most prominent examples of such resources correspond to relatively independent parts of the physis of the agent such as its eyes, feet, and hands.

5.1.1 Processes

A *process* models one concurrent strand of activity in the agent[15]. It can be *active*, i.e. allocated a 'processing time' resource, or *dormant*. A change to the dormant state is triggered by *waiting*: for a specific resource; for communication from another process; or for the notification of a change in an observed process. A process can request control of resources. A process can *start* or *stop* another process, and transfer its processing time and other resources to it. Similarly, a (meta-level) process can act as a scheduler for other processes. Processes can actively observe and inspect the status of other processes, their communication patterns, their resource use, or they may request to be notified of such changes. The acquisition of resources can either succeed immediately or it might be delayed because the resource is currently occupied. In the latter case the process is suspended until the resource becomes available. A non-blocking request for acquisition is also possible.

Building on these capabilities of processes, I can model what Frijda [Frijda 2007] calls competences: a behaviour system that includes sensitivities for triggering activation, and criteria for well-functioning used in system monitoring. Such competences in turn form the basis for on-line appraisal processes and major concerns of an agent.

5.1.2 Resources

A *resource* is an abstraction for the limited control an agent has over the part of the world that constitutes it. Processing capacity is an abstract resource that limits the number and the speed of concurrently running processes. In the context of a simulation on a serial computer, this translates to how frequently and then for how long a process is allowed to run[16]. Communication channels are a specific form of resource that allows direct communication between processes. Further, communication can be used as a means for synchronisation to model interdependencies between processes. [Scherer 2001] assumes that par-

---

[15] I use the computer science term 'process' rather than 'thread', as usually threads indicate strands of control that share global data, while processes communicate by message passing only. The latter explicit communication strategy better reflects the conceptual model chosen in this implementation.

[16] This corresponds to the distribution of time slices in a scheduling algorithm. On potential future hardware with massively parallel computing capacities, different translations of resource limits to process scheduling are possible. The important point is that such limits are explicitly considered in modelling.

allel appraisal component processes, i.e. Stimulus Evaluation Checks (SECs), usually complete in a specific sequence because of such inter-dependencies.

Resources at the boundary to the agent's environment group sensors and actuators, reflecting the fact that the agent is always *interacting* with the environment rather than receiving or sending [Bickhard 1999]. These boundary resources are subject to influences from both, the environment and the agent's processes. Internal resources and processes are only indirectly affected by the environment through communication links (I thus exclude scenarios involving drug use and brain surgery, while conceptually internal effects of e.g. hormones can be modelled as processes). They can be shielded explicitly, i.e. decoupled, from the environment to form emulators [Grush 2004] or prediction machines [Minsky 2006].

5.1.3 Implementation

My implementation of the building blocks is based on cooperative microthreading in Python. Although cooperative microthreading is a technique that does not allow for true parallelism, even if the processing unit of the executing machine would allow limited parallelism its use can be justified by the requirements of a simulation. This cognitive architecture is intended for a simulation environment (albeit a real-time simulation), i.e., the goal is not to recreate actual parallelism but to recreate the interactions typical of parallel processes that are assumed to occur in a cognizer. Therefore, the relative simplicity of cooperative microthreading offers a definitive advantage over a programming system actually capable of massively parallel execution. In addition, and not as the least reason, this choice allows execution on commodity hardware. The current implementation using cooperative microthreading enables strong control over scheduling and incurs significantly less overhead than the use of operating system threads or processes. The set of resources of an agent is assumed to remain unchanged over the lifetime of an agent.

Appendix A describes the Python package developed to provide the basic functionality. The interface to the use of resources and inter-process communication is described in Appendix B.

5.1.4 Environment Interface and Management

The usefulness of building blocks as structuring principle for an agent also depends on the interface between these building blocks and the actual simulation environment of the agents. As explained in section 1.2.1, some of the objectives of the current work directly concern the

relation between an agent and its environment. Multimodality and sensing and acting over time need to be based on a suitable embodiment in the specific virtual world simulated. This section describes the general approach towards a completely independent environment and an agent physiology based on a physics simulation as well as the rudimentary management facilities for starting, stopping and inspecting the agents and the virtual world. Section 5.2 provides more details on the resources used to model the simulated characteristics of the agents implemented for the test scenario.

The virtual world, the agents and a management application have been implemented as separate processes on the operating system level. The communication between these subsystems is implemented using an network library that allows for asynchronous message delivery. Furthermore, a simple explicit time management is introduced to ensure that interaction between subsystems (especially virtual world and agent models) is scheduled in a fair and inspectable fashion. The management application provides facilities for starting and stopping the world and the agents as well as for examining the log traces of the subsystems. It also functions as a known reference point to discover and connect the separate subsystems. Both agent models and virtual world also include an mode for interactive inspection using a command line interface when the simulation is paused.

The agent models, while running, communicate to the virtual world by issuing function calls using the network interface (e.g. for reading the current vision buffer or for applying a force to a part of their body). The function calls have varying temporal characteristics but the time management is used to prevent or at least detect and log deviations outside the assumed upper and lower limits of execution time [#deviations]_.

The virtual world combines a graphics engine and a physics engine to provide a simulated physical world that includes the bodies of agents as well as the avatar body of the user (a simple oblong). Appendix C documents the technical details of the implemented simulation environment. Both the user and the agents can move their bodies only when the simulation is running and only by applying forces to them (rather than directly setting their location). The user has the additional ability to apply forces to other objects visible from her or his viewpoint using a pointer. Agent bodies include a skeleton of *bones* connected by joints simulated by the physics engine (a so-called ragdoll, see also figure

---

[17] Such deviations are always possible on a non-real-time operating system. The approach of signalling such a conditions seems most suitable for a flexible simulation sytem where any implementation change of agents or the virtual world as well as changing conditions of the host machine could lead to a change in the temporal characteristics of agent-environment communication.

6.2) that is used to influence the body, again only by applying forces and torques to the bones. These skeletons of bones are mainly used to control the extremities of agent bodies but they also include special 'bones' for the control of agents' eyes, eyebrows and lips to allow for gaze shifts and rudimentary facial animation.

5.1.5 Configurations

Given a system configuration of dormant and active concurrent processes, a first issue regards the starting configuration. This question is obviously important for designing and running an artificial system. Translated to the context of adult humans, the question is equivalent to obtaining a snapshot of a mind at a specific moment. A starting point for my incremental modelling approach is to assume that there is at least one active process per resource on the boundary to the environment, that tries to identify changes (i.e., sensors are active by default). There also is one process for each sensitivity of the major concerns [Frijda 2007] of the agent, which in turn creates specific monitoring processes that operationalise the motivation to act on opportunities and threats.

The question of 'configurations' needs to distinguish between the configuration of all processes and the internal configuration of one specific process out of the set of all processes. While both are important, the focus on process coordination in this thesis also entails a focus on the overall configuration of all processes.

Based on these building blocks, the following architectural questions about configurations can be posed for any specific scenario:

- What (types of) processes and resources are there, and how many instances of the different types (e.g. regulation processes, monitors, different schedulers) exist?

- What processes run periodically and most frequently?

- Which processes compete on which resources?

- Which processes monitor what of other processes?

The intention of the incremental methodology of this thesis is to arrive at a set of common processes and process configurations that are helpful to model the chosen emotional phenomena. As a first set, table

---

[18] I use the term interactuator to emphasise the absence of separation of sensors and actuators in the simulated virtual world. Sensing and acting are not separate interface channels, but are combined in the use of resources, as described in the following section.

Table 5.1: Typology of processes (the types are neither necessarily distinct nor organised hierarchically).

Process type	Interesting characteristics
One-off processes	How long do they take (already did and probably will), what 'goal' are they related to, and what process originally created them?
Resource processes	Which resource do they access, what strategy do they use for resource acquisition (fail-fast, persistent, or escalate to another process)?
Single resource processes	Long-running processes that regularly trigger updates of *interactuator*[18] resources.
Multiple resource processes	Processes that combine multiple resources often need to be given only a limited lifetime, as otherwise the probability for livelocks increases.
Eternal processes	Both high-level concerns and processes for maintaining a specific base state of the bodily resources (such as an upright posture) have to be protected against a potential premature end.
Communication processes	These processes mainly wait for and route notifications.

5.1 presents a typology of processes encountered in the implementation of the BehBehBeh scenario described in chapter 6. Note that the types mentioned are not necessarily distinct, e.g., many of the eternal processes are processes tending to one specific resource, and communication processes can be one-off processes or more long running ones.

The different strategies for resource acquisition mentioned in the table will be covered in section 6.2. The next section presents an overview of the resources modelled for the agents that populate the specific scenario adopted as an initial testbed in this thesis.

5.2 Resources of Physically Modelled Agents

In the BehBehBeh scenario, two agents were modelled to populate a world implemented using a freely available graphics engine and a physics engine. In this section, I describe the resources that were implemented as part of these agents as far as they are the same for both and conceptually independent of the concrete scenario. The individual differences between the two agents will be covered in the next chapter. In the context of the research method proposed in this thesis, these resources can be seen as a first step towards the identification of de-

sign patterns for cognitive components of physically modelled affective agents.
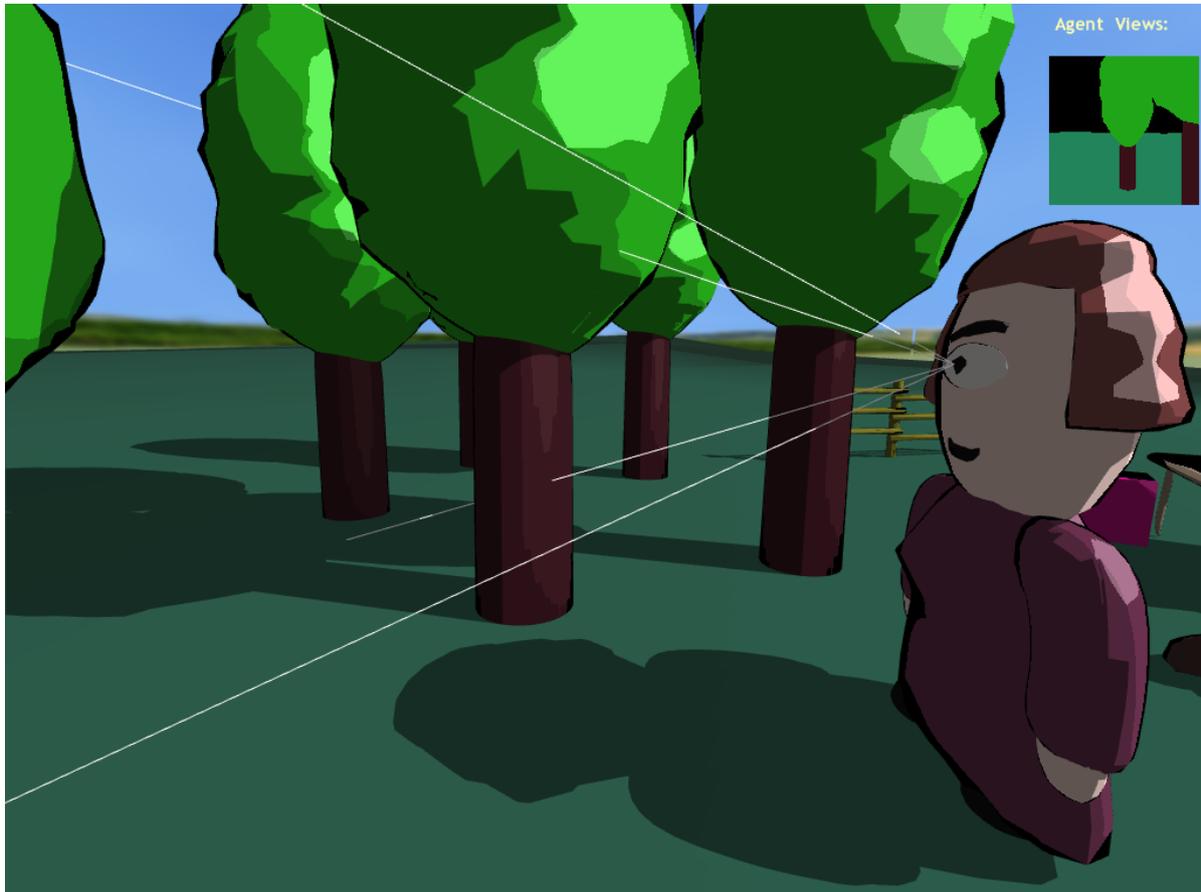

Figure 5.2: An example of the viewpoint of an agent as rendered by the BehBehBeh system.

Resources are abstractions of the physical limits of an agent. Therefore, the main resources correspond to the simulated physiology of the agents (see table 5.2). In addition, more abstract resources are also modelled: processing capacity is restricted by limiting the num-

---

[19] To allow for more direct control of processing capacity, an estimation based on the number of executed instructions of the Python virtual machine was also implemented but not used in the actual implementation. This count of instructions is an estimation of processing effort as valid as any, since for the granularity of many of the processes implemented here, no reasonable performance data (such as the one used in ACT-R) is available.

Table 5.2:   Agent resources corresponding to physiology.

Resource	Characteristics
Head	The head provides access to bones that control the orientation and twist of the head. Continuous control is necessary to keep the head from falling over.
Trunk	The trunk provides the most simple access to movements of the whole body, i.e., turning and pushing forwards or backwards, by accessing the central and upright spine bone. Fully realistic movement should actually use the leg resources, so this is a deliberate simplification, without dropping the requirement of physicality completely.
Spine	The spine combines access to all other bones in the agent trunk, i.e. those that correspond to the spine of a human skeleton. No attention to these bones results in a constantly crouched figure.
Leg (2x)	The legs provide access to the leg bones, as mentioned for the trunk they are not strictly necessary to control full-body movement.
Arm (2x)	The two arms, each consisting of a single bone for simplicity, can be affected using these resources.
Eye	The eye is the most complex resource. It can trigger false colour rendering of the agent's viewpoint and report the sensed colours, their pixel count (based on a fixed resolution), and the centre of the pixels relative to the viewport. The properties and limitations of the implementation are explained in Appendix C. Figure 5.2 shows an example of the false colour rendering for one agent eye.
Face	The face provides access to 'bones' controlling the mouth and eyebrows of the agent. While this resource does not allow sophisticated facial animation, it allows to increase the distance between eye and, respectively, eyebrow or mouth, which can be detected by another agent.
Ear	The ear is a simple resource that can be queried for the presence of any sound source in the vicinity. As a simplification, sound sources are assumed to have a strictly spherical propagation.
Nose	The nose functions analogous to the ear.

ber of processes that can be active at the same time[19]; communication channels and notifications are also limited in number. These more abstract restrictions only have minor effects on the simple initial scenario modelled here, but are expected to be crucial for the extension of the method to further emotional phenomena. Another notable more abstract type of resource are emulators that encapsulate emulation knowledge of the typical behaviours of other elements of the world. The only implemented variant is a predictor for the behaviour of another agent (used as a resource *Child Predictor* in the next chapter). The implementation in the current case is simplified to the detection of movement towards or away from an object. Such emulators are expected to form the basis for more complex reasoning about the world as well as for theory of mind modules in further extensions of the agent architecture and are thus modelled explicitly as a limited resource.

5.3 Summary

In this chapter, parallel processes were introduced as the procedural primitives of the agent architecture presented in this thesis as well as the importance of process configurations in the description of agent instances. Further, the modelling of limitations of the agent as resources was described followed by the description of the implementation of these building blocks as realised in the BehBehBeh system. Finally, the types of processes used in the implementation have been described and the resources that make up the abstract limitations and the physiology of the BehBehBeh agents have been introduced. The next chapter will describe these agents in more detail and how they are used to model an initial emotional phenomenon.

6   Disgust as First Emotional Phenomenon

Where the instantiations of the modelling primitives in an executable environment for reproducing a specific emotional phenomenon are described. Where the extension of the implemented methods to further areas of emotional life is described. Where the relevance of the concepts developed in previous chapters is tested in the analysis of further scenarios.

BehBehBeh is the testbed system developed for the purposes of this thesis. It is used to simulate an episode of the emotion disgust, involving, mainly, a mother, a child, and a piece of dog poop. The name is an abbreviation of 'Behaving affectively Behoving affective Behaviour'[20]. Figure 6.1 shows the two agents in the 3D environment.

As explained in section 1.2, the approach of this thesis is not to create a general framework and then derive instances of models using it, but to bootstrap a framework by incrementally modelling different emotional phenomena while reusing successful structures. The BehBehBeh testbed system described is the first instance of an emotional phenomenon modelled using this approach. Structures and design patterns that can already be argued as scenario-independent are described in the previous chapter. This chapter describes the structures used for a specific emotional phenomenon as a basis for long-term future work that will repeatedly test them for different emotional phenomena.

6.1   The Disgust Episode

This section introduces the emotional episode that is the target of the proof-of-concept application of the method presented in this thesis.

---

[20] The name choice was motivated as a continuation of my master thesis project and to emphasise the shift in focus towards modelling agent behaviour rather than dramatic acting. The name of the previous project is 'Acting Affectively affecting Acting', abbreviated as ActAffAct.

Figure 6.1:   The 3D environment of the BehBehBeh simulation world.

6.1.1 A Disgusting Encounter

The two agents that populate the BehBehBeh world are a mother and her child. The specific emotion instance that the simulation should be able to support can be described in terms of the following episode that mentions only externally observable behaviour:

- Child approaches the dog poop.

- Mother looks at child.

- Mother looks at dog poop.

- Mother initially expresses disgust.

- Mother retreats from the source of disgust.

- Mother reverses direction to approach child.

- Mother's expression intensifies.

The emotion word *disgust* subsumes response dispositions aimed at protecting the body (*core-disgust*, [Lawrence Calder 2004]), and social and personal values (*socio-moral disgust*, [Simpson et al. 2006]). Disgust is assumed to have evolved from a distaste response, but the current form of human disgust is largely social. It seems to require inculturation since it is not present in infants, and is a rather common topic of adolescent humour [Rozin et al. 2000]. According to [Frijda 1986], the expression of disgust reduces sensory contact with distasteful substances in the mouth and tends toward expelling those substances. The associated action tendency is labelled 'Rejecting (closing)', with the function of protection and the end state 'Removal of object'. The appraisal pattern predicted by [Scherer 2001] involves low levels of familiarity, predictability, and need relevance; very low intrinsic pleasantness; medium urgency; and a very high outcome probability.

The episode laid out above incorporates both the aspect of core disgust and a social variant of showing disgust as a means of social signalling that is related to socio-moral disgust[21]. To elaborate, in this episode, the mother tries to determine whether her child is safe by looking at it. She detects a (mildly) disgusting object in its vicinity. The immediate reaction is one of disgust (facially and bodily, by retreating), but this reaction is soon counteracted by her concern for the safety of her child: this leads to her approaching the child. In addition, the expression of disgust is still shown for the benefit of the child that should learn that the focus object is in fact a disgusting object for its mother.

The behaviour of the mother is motivated by what has been described in attachment theory [Bowlby 1969]. More specifically, what is described as the mother's *caregiving behaviour system* is taken to be responsible for her concern for her child's safety and well-being. The *attachment behaviour* of the infant is part of the complementary behaviour system also described by attachment theory. Even more interesting for the emotional phenomenon described above is the notion of *social referencing*, i.e. the ability to use the emotional reaction of another agent as a basis for one's own appraisal of a novel situation [Campos Stenberg 1981; Feinman 1982]. This allows the amplified expression of the mother to potentially (and likely) influence the future behaviour of the child with respect to dogpoop (as well as other situations judged to be similar).

---

[21] The example is constructed to illustrate the potential of modelling with concurrent processes. It is not directly based on examples from the psychological literature.

6.1.2 Scenario Analysis

The motivation for the scenario described here is to provide an initial emotional phenomenon that can be modelled using the chosen building blocks so that structures usable for further models can be identified. The purpose of the implemented system is to allow for experimentation by the scientist using the system in the form of changing the starting conditions or interacting with the running system as well as inspecting log traces and internal states. A closely related aspect to keep in mind is the goal of identifying 'emergent' patterns of behaviour for the phenomenon disgust, i.e. behaviour that is to be expected if you model the basic underlying affective mechanisms respecting the fidelity requirement, see section 2.4.2. The possibility of the emergence of such patterns, such as the one described above, is also the basis for evaluating the presented system.

For a complete scenario analysis, as argued for in section 2.3, I also need to consider the details of deployment. The interaction qualities are mostly described in section 5.1.4: the user interface for the whole system as well as the interaction of agents with their environment are both based on simulated physics and the simulations are not time-stepped from the point of view of the agents (even though the physics simulation and the graphics engine do of course use discrete interpolation internally). Interaction modalities include movement, seeing, hearing, and smelling. There are two agents and one user present in the scenario, see the present section for a typical interaction script. Any interactions where physical contact between the agents has a functional role (as could be expected for disgust episodes) is explicitly excluded.

The following sections describe the setup of the agents, i.e. their bodies and their embedding in the environment as well as their initial sets of processes and resources. The rest of the chapter will present the actual sequence of the simulated phenomenon in more detail.

6.2 Disgusting Agents and Processes

For BehBehBeh's disgust scenario, two different agents were modelled. Both have the same kinds of physical resources, i.e. similar anatomies, with some differences in the initial configuration of their bone structures as the child is much smaller and in a bent position suitable for crawling. The following two sections describe, respectively, the structures common to both agents and the specific processes of the individual agents.

6.2.1 General Structures and Criteria

The general structures of both agents consist of the resources already described in section 5.2 as well as a set of processes. Part of setting up the general structure of such an agent model is the decision of what will be modelled as a process. The main characteristic of the processes used as building blocks is that they are independently and concurrently active, i.e. the demise of one process does not influence any other process unless an explicit link has been established in some form. This characteristic is also used as a intentionally wide criterion for deciding on what should be a process in the model: any activity that can exist independently of other processes.

Eternal Processes

The most basic type of process active in the agents are those that will be active regularly over the whole lifetime of the agent. Three subtypes of these eternal processes (see section 5.1.5) are used in modelling the two agents in the scenario: *body-related control processes*, *high-level concerns*, and *management processes*. The defining element of these eternal processes is that they are protected from ending prematurely, i.e. they cannot be killed by other processes.

Both agents use eternal processes that attend to one specific bodily resource. The task of these processes is to keep the corresponding bodily resource in a nominal working base state unless it is acquired by some other process. For each resource this involves proprioception, i.e. the regular sensing of its current state (e.g. the current orientation of the head, or updating the visual sensing data). In addition, for some resources, regular corrective action is aimed at maintaining the desired base state: e.g., in the case of the head, its upright position. These bodily processes correspond to the autonomic base activity of any living entity.

The body-related control processes cannot be abstracted away fully from the concrete agent; they are conceptually the same for both agents but parameterised to fit the particular physiology. In the case of the child, it moves by crawling, using both legs and hands, and even though whole body movement has been simplified compared to the natural example of human toddlers, the interaction with the physics calculation entail a distinct change in movement control behaviours. The additional contact points with the ground, feet *and* hands, influence the results of full body movement even when applied as a force to the trunk resource.

Figure 6.2 shows the bone structure used in the control skeleton of the agents and figure 6.3 depicts the corresponding collision volumes

used in the physics calculations. Both agents use simulated false-colour vision to get information about their surroundings (using a single Eye resource). This provides them with a suitable subjective point of view in the simulated world, as required by the premise of situatedness. The details of the implementation of sensing are covered in Appendix C.
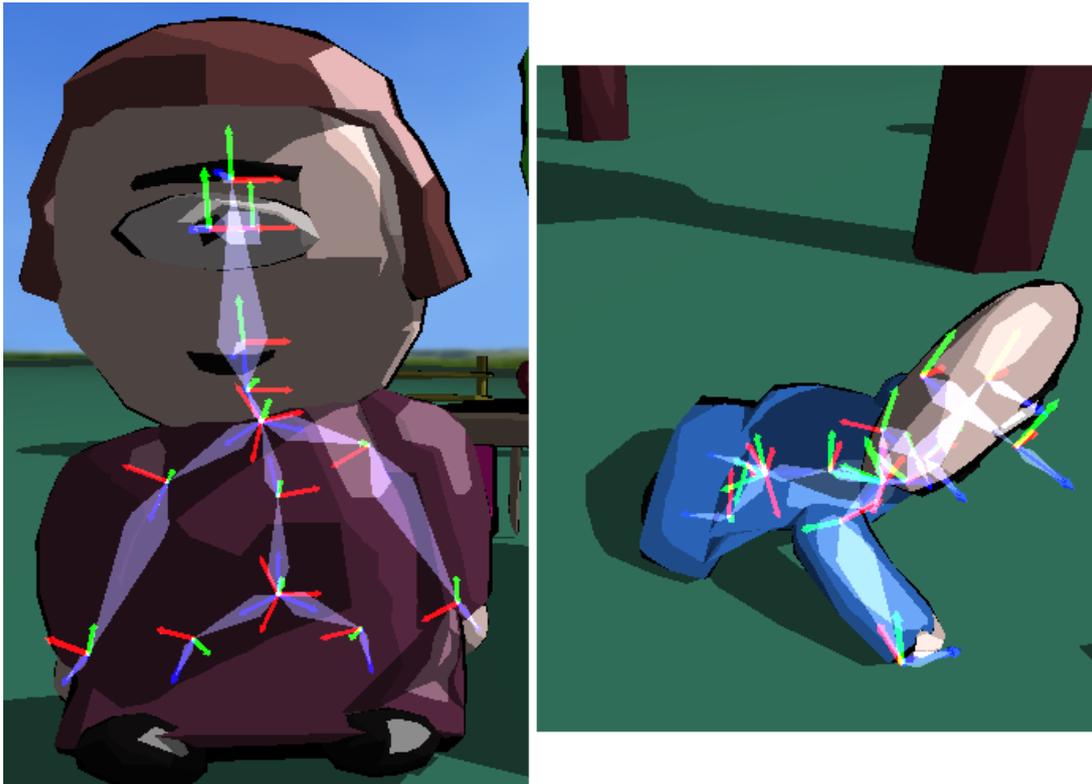

Figure 6.2: The agents' bone structures used to influence their movement.

High-level concerns of an agent implement its motivations. Conceptually, concerns are dormant and only become active if a suitable opportunity for satisfaction is noticed or if a change is appraised as signalling non-satisfaction of the concern. As the building blocks of the model are active processes, 'dormant' here actually means regular checking of specific conditions and/or registering for notifications of specific conditions. The actual conditions that are relevant for a concern and how they are detected can be very different: this is the subject of appraisal in emotion theories and the main idea of my approach towards modelling is to select suitable implementation strategies based on their merits in the chosen scenarios. The processes that are denoted
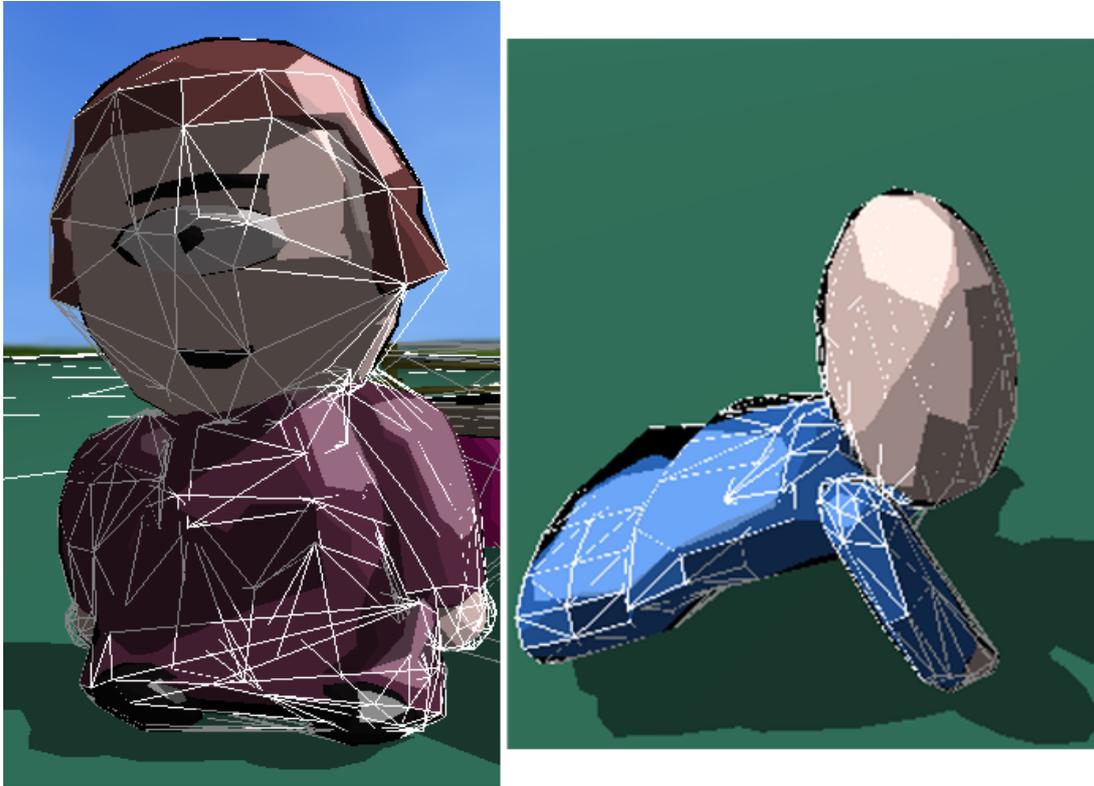
Figure 6.3: The collision volumes used in the calculations of the physics engine.

concerns in this implementation can thus be seen as including a realisation of the monitoring processes of specific concerns. A separation of monitoring and activation into different processes was not necessary. If a concern becomes active it creates other derived processes that are intended to fulfill specific tasks to cope with the detected situation.

Derived processes, described in the following section, often need to acquire and use the resources of the agent. Management processes are used to arbitrate between these derived processes in the case of resource conflicts. They provide the methods for escalation strategies when resource acquisition fails as described below.

Derived Processes

I use the term 'derived process' for any process that was created by one of the eternal processes to complete a specific task. In relation to emotion theory, some of these derived tasks are notable since they corre-

spond to the theoretical notions of *relational action tendency* and *coping*. The former are processes created by active concerns that are relatively flexible in deciding on a way to fulfill their task, i.e. to achieve, maintain, or end a specific relationship with the environment. Coping processes are those created by relational action tendencies to achieve specific goals.

The most important task of many derived processes is to acquire specific resources. Regarding the methods for acquiring a resource three different general strategies for dealing with failures can be distinguished:

- Fail-fast: The acquisition process tries to acquire the resource and aborts immediately if it does not succeed, thus ending its task and potentially allowing other processes (e.g. the one that created it) to detect that failure.

- Persistent: The acquisition process tries to acquire the resource until it succeeds, either by blocking on the acquire call or by repeatedly pausing and trying to acquire (both strategies can make sense, e.g., the latter would correspond to a low-priority task in priority-based scheduling).

- Escalation: If the acquisition process does not succeed it can notify a separate management process to arbitrate the conflict. Many different strategies, in turn, are possible for choosing when to notify this process and for the arbitration decisions of this process. One specific such management process was implemented for the current scenario that decides in favour of the notifying process if it can determine that the current resource holder is in a waiting state and not using the resource actively.

Such a forceful release of a resource can lead to a premature end of the process previously holding the object. Most derived processes are designed so that they can be restarted gracefully if this happens.

6.2.2 Individual Process Structures

Regarding the structure of processes, the child is the simpler agent. Its main 'high-level' concerns involve seeking interesting objects and if possible touching them, as well as regularly looking at its mother. This provides for the possibility of runtime conflicts between competing processes and corresponding conflict resolution processes that are typical of affective processing. However, since the child agent is not the main focus of the simulation, it is only described superficially here.

In the following, the rest of the process structure is described for the mother agent. As a starting point, the mother agent has several concern processes relating to the child's well-being:

- look_after_child

- protect_child

- teach_child

These processes are responsible to regularly check for the current activity of the child, to protect it from harm, if there is any indication of possible harm, and to use opportunities that present themselves to teach the child about good and bad. How the processes fulfill their responsibilites is not fixed. These processes have that status of base concerns, i.e. they are eternal and search for opportunities to use the current activity of the agent to their advantage. For example, while the first concern is named look_after_child it might use either current visual data or aural data to satisfy its motivation to know about the current state of the child. If there is no other way, concerns can use strategies to escalate their influence, as described below.

As mentioned in section 2.3, an interesting property of scenarios for affective agents is if they involve conflicting tasks as the resolution of such conflicts is an important part of emotion. To ensure that task and resource conflicts will happen in this scenario, another base concern was used that motivates the mother agent to mind the table in the scene by looking and approaching it regularly:

- mind_table

Another eternal process was used as part of one of the escalation strategies:

- regulate_conflicts

This process can be notified by a concern process, when attempts to acquire a resource have failed, creating a situation of urgency that needs arbitration. As mentioned above, the strategy implemented in the regulate_conflicts process arbitrates a conflict by looking at the current activity of the current holder of the resource. If it is waiting for a response from the resource, the request for conflict resolution is denied; if the resource is not directly used, e.g. because the process is waiting for another process, or simply waiting for a timeout, the request is granted and the resource is temporarily transferred to the other process. Note that this arbitration strategy does not prevent occurence of deadlocks or livelocks, both are possible and the present

architecture is not intended to prevent them. On the contrary, the possibility of livelocks for example is even a worthwhile property of a simulation of affective behaviour. In practice, there are areas where locks need to be prevented. For this purpose, the implementation employs a specific order of resource acquisition for processes that acuire more than one resource. Competition between different processes notifying the management process cannot occur in the current implementation due to the nature of cooperative microthreading (i.e. only one process can be notifying it at a time). For future models using other implementation strategies such meta conflict resolution could be an interesting topic to investigate.

All other processes are started indirectly by bodily processes or one of the concern processes above. Typical processes used in the implementation include finding an object, checking whether a specific object is currently visible, approaching an object, checking the status of a resource, and evaluating the relation of two objects in view (e.g., the distance of their object centres or whether they are moving towards or away from each other).

6.2.3 Disgust in Detail

This section presents a detailed walk-through of the unfolding of the modelled phenomenon of disgust in the BehBehBeh system. Figure 6.4 shows a sequence diagram of the processes active in the mother agent during a possible interaction sequence corresponding to the emotion instance described in section 6.1.

As shown here, the relevant sequence of processes can be started by the concern look-after-child This is, however, only one possible starting point, as noted in the following list that describes this specific sequence in more detail.

1) Initially, there are several processes active in the mother-agent. One process, look-after-child, is responsible for satisfying the mother's concern for the well-being of the child. This process is active periodically to look for matches with its sensitivities, as are other concern-relevant processes: protect-child, teach-child, and the regulation process regulate-conflicts.

2) The process look-after-child acquires the resource Eye, actively searches for the child visually, and detects that the child is visible[22]: a novel event. Thereupon, it spawns a process to interpret the current physical activity of the child as movement.

---

[22] Another possibility would be that some other process that uses the Eye detects

Figure 6.4:   Processes in a mother in the child and disgust scenario.

3) Interpret-move uses the resource Eye to find target objects near the child. If a potential target is identified, it spawns a process (identify-object) to identify the object and another one (predict(child)) to predict what the child will be doing.

4) As soon as identify-object has a first identification (*dog poop*) it starts an evaluation process. Similarly, the prediction process, after using a Child Predictor resource and the current result of the identify-object process, spawns a process to evaluate the predicted action.

5) The evaluation of the *dog poop* appraises it as unfamiliar and intrinsically unpleasant. The resulting relational action tendency (RAT) triggers the process for expressing disgust and receding from the object as coping action. Each of these processes starts by acquiring the relevant resources, Face and Legs.

6) The evaluation of the predicted action (the child touching the *dog poop*) has identified a threat: possible harm for the child. The process protect-child is notified of the potential threat and reacts by creating a new process to remove the threat (an action readiness) which, in turn, instantiates a process for approaching the child.

7) The approach process tries to acquire the Legs but fails, because the recede process is controlling this resource. The process regulate-conflicts is notified by the action monitoring provisions of approach and arbitrates by ending the process recede since it is currently in a waiting state and not sending movement actions.

8) Concurrently, the process teach-child detects that the processes remove-threat (a coping process targeting a bad object) and express-disgust (an expressive process targeting a bad object) are related via the object involved. The responsibility of the process teach-child is to find connections between processes related to the concern process protect-child and in the current implementation this is this is very much a specialised rather than a generalised process. It reinforces the process for expressing disgust by transferring control to it temporarily.

---

movement of the child. Look-after-child could then be notified if it was monitoring any changes involving the 'child object'.

Not implemented, but part of the original idea based on the scenario analysis was the addition of a possible further event:

- Finally, the object identification process finds new evidence (the *dog poop* is actually made of *chocolate*) and reinitialises the evaluation process. Consequences this might normally have (overt signs of alleviation and humour, related internal process reconfiguration) are suppressed by regulation as incompatible with the ongoing, reinforced, expression of disgust.

The possibility of this sequence of agent interaction based solely on the interaction and coordination of processes without the use of a central evaluation structure can be regarded as a first evaluation of the suitability of the present approach towards modelling affective processes as described by process-based emotion theories.

6.2.4 Shortcuts

Several shortcuts have been taken in this process-based modelling of a disgust episode. Most notably, as described in section 5.2, the intricacies of sensing and acting in the environment have been simplified to a process that employs a resource to retrieve current data. Predicting actions of others (the child) and forming expectations by internal simulation has also been abstracted into a ready-made resource. Further, any knowledge required and used by the processes is not accounted for in the model. As an example, the identification of objects has to access some body of prior knowledge, that is here encoded in a one-to-one relationship between objects and colours. In an evolution of the agent architecture, such prior knowledge might be modelled as exclusive and specific to one type of process or as a shared (set of) resource(s). The same holds for the detection of the relation between threat, disgust, and *dog poop* based on the examination of (relations between) active processes. I expect that during the evolution of the framework and building of simulations of different emotional phenomena of increasing detail, shared knowledge will be introduced as resources. The next sections describe the evaluation of the presented implementation and the expectations and plans for the extension of the modelling approach to different emotional phenomena.

6.3 Evaluation of the Modelled Phenomenon

As already explained in section 1.2.3, the presented modelling approach poses a significant problem for short-term evaluation. It is intended as a program for long-term research that would provide substantial re-

sults to be evaluated only in later stages. However, of course an evaluation of the presented system is necessary. Of the three different evaluation criteria mentioned in section 1.2.3, only the feasibility tests make sense in the current context. Section 6.1.2 suggests the definition of a pattern of behaviour of disgust that should emerge from the interaction of the processes modelled. For this purpose the scenario script presented at the beginning of this chapter was used as a reference for the behaviour to be achieved by the system. Analysis of the log traces of the system showed that for starting conditions that included the two agents in the proximity of the dogpoop object an interaction sequence similar to the described one occurred in about half the runs[23].

6.4 Incremental Extension to Other Emotional Phenomena

The disgust emotion modelled in the BehBehBeh system is only a first instance of the research work proposed in this thesis. The methodology presented in section 1.2 states explicitly that rigid behavioural structures in the implementation of agents should be avoided. This decision entails that a specific agent instance runs the risk to be specifically tailored to one scenario. As a counter-measure, my approach involves the incremental extension of the modelled phenomena, i.e. the successive modelling of different emotional phenomena and, most importantly, the identification of commonalities motivated not by the wish for abstraction (a common pitfall of computer scientists in modelling applications) but by the needs of the different scenarios.

Potential candidates for the extension of the presented disgust scenario are emotional episodes of happiness or pride. These phenomena would constitute a natural extension of modelling the behaviour of the two agents by implementing the explicit detection of an agent's own competence. For example, the successful handling of a situation such as the one modelled in the disgust scenario can trigger an evaluation of one's own competence in dealing with the situation. This evaluation could then result in the diffuse feeling of happiness or a distinct feeling of pride, depending on how crisp, or 'conscious', the evaluation was. In the case of the child expressing happiness about its current situation, this in turn can trigger reciprocal confirmation behaviour by the mother.

Another potential avenue of development can be taken from the different behaviour systems described by attachment theory. The child

---

[23] Of course, this is only a qualitative evaluation. A more quantitative analysis would of course be possible, but also absurd for the present system due the lack of a reference point to compare it to.

losing contact with its mother due to the loss of direct sight behind trees can result in episodes of distress.

While no extension of the BehBehBeh agent architecture to further emotional phenomena has yet been completed, the following section describes the application of the basic methodology and the scenario-based analysis method to different related scenarios. These were performed in parallel to the work described above and served as a reference point to keep the methodology general enough to be useful in different applications.

6.5 Application in Different Scenarios

This section describes two specific scenarios that were not targets of implementation of the agent architecture described above, but objects of analysis using the scenario-based method described in section 2.3. The systems are the interactive exhibit 'The Invisible Person' [Petta 1999, 2003] and the music companions developed as part of the EU project RASCALLI [Krenn Schollum 2008].

As described in section 2.5, the Invisible Person includes the simulation of a single child-like agent presented on a stage in a museum so that visitors can interact with it by stepping onto the stage. The agent itself is presented on a projection wall in front of the stage that also displays a digitally reversed camera picture of the stage itself, resulting in a 'magic mirror' [Maes et al. 1994]. The interaction qualities as seen from the agent's viewpoint in this interactive exhibit are very similar to the BehBehbeh system, thus the implementation choices of the resulting sytem served as a very useful reference for the design of the process architecture of BehBehBeh. As one example of a difference that is directly reflected in design decisions, consider the purpose of the two systems. The purpose is decidedly different in that the Invisible Person has to engage visitors continuously and provide entertaining displays of behaviour rather than simulate behaviours with some degree of physical realism. This is a direct consequence of the setting of and the interaction possibilities for interacting with the system. A visitor can, after all, also walk through a virtual character that is only projected onto a screen (even though the system detects these situations and actively tries to avoid them). The design of the different animation components in these two systems are based on these different scenario characteristics: while the Invisible Person has a single queue of animations that needs to be filled continuously and without interruption while being kept short enough to preserve a sufficient degree of reactivity, BehBehBeh agents are animated based on the output of a physics simulation, use a set of basic processes to stabilise their simulated bodies

and can and will be influenced by other objects in their environment.

As part of the project RASCALLI (Responsive Artificial Situated Cognitive Agents that Live and Learn on the Internet) a framework was developed for Internet aware service agents that serve as companions, providing access to knowledge in a specific domain. The example application central to RASCALLI is that of a music companion, i.e. a virtual agent that can be queried about information related to artists, songs, or genres. As part of the RASCALLI project, I implemented an extension of the music companion agent that was based on the evaluation of the affective potential this scenario presented. The original agents developed in RASCALLI use turn-based question and answering as main pattern of interaction with a single user, i.e. from the viewpoint of the agent, this interaction is restricted to receiving a natural language input that is predominantly interpreted as a question. The user can also provide negative or positive feedback (scolding or praise) triggered by pressing dedicated buttons on the graphical user interface. The potential for emotion instances in this scenario is restricted to the feeling of competence when a specific question can be understood and answered correctly (as far as the agent is concerned) and the potential for pride in case such a 'good answer' is rewarded with praise, as well as corresponding opposites. The extension of the music companion added a separate layer of agent concerns that are active over a longer period of time. Apart from the concern to answer questions, the agent also considers possibilities to learn more about the preferences of the user by asking questions proactively, and it considers the need to receive regular input from the user and can choose to shun the user if this need is not met. This concern structure allows for the implementation of a companion with an attitude, i.e., a personality agent. A direct consequence for the agent architecture resulting from this extended scenario was the need for an equivalent of persistent processes in the predominantly reactive architecture.

6.6 Summary

This chapter presented the instantiations of agents in the BehBehBeh systems based on the building blocks described in the previous chapter. The emotional phenomenon of disgust, chosen as a starting point for the modelling effort, was described and the process structures involved in recreating an episode of disgust were presented. Further, the plans for extension of the modelled phenomena as required by the modelling approach were laid out and the application of the scenario-based analysis method to related applications was introduced.

Part IV

Closure?

7   Conclusions

*Where the main contributions of the thesis are summarised
and a plan for potential future work is laid out.*

This chapter summarises the main contributions described in this
thesis regarding the research on intelligent autonomous situated agents
and its relation to and interaction with the modelling of emotion. It also
presents an outlook on potential and planned future work.

7.1   Main Contributions of the Thesis

Regarding the objectives laid out in sections 1.2.1 and 2.4, the virtual
world developed for the purposes of the thesis provides for the com-
plexity of embodiment, sensing and acting over time, multimodality,
and real-time interaction. It respects the limited control a situated
agent has over itself and its environment. A suitable basis to tackle
the longer-term objectives was thus established.

A theoretical contribution of this thesis is the scenario-based anal-
ysis method for characterising the scenario of use of a specific agent
architecture. In section 2.3, I introduce the concept as a new and es-
sential means to capture the *emotional potential* of a scenario. The
main parts of a scenario description are the motivation for building the
system, its purpose and the details of a possible deployment. In chapter
4, I use elements of scenario description for the evaluation of related
work. In section 6.1.2, the initial scenario chosen for the implementa-
tion described in this thesis is presented using the method.

The sections 3.1 and 3.1.3 analysis recent theories of emotion and
their relation to implemented models as a theoretical foundation for
the modelling approach presented. As part of this approach, the con-
tributions of this thesis include the proposal to use explicitly bounded
resources and concurrent processes as building blocks for computa-
tional modelling of emotion in a virtual world. Motivated by the re-
quirements and the potential of the specific scenario chosen and by
concepts used in emotion theories, I proposed these building blocks to
focus on the identification and management of parallel processes and
the resources they access. This approach aims at an incremental mod-
elling of a growing collection of emotional episodes, with a clear doc-

umentation and delineation of technically necessary simplifications of the natural phenomena. Chapter 5 and the appendix provide details on the implementation of these building blocks based on cooperative microthreading. Chapter 6 presents the first scenario implemented. The structures identified as useful in this implementation based on technical considerations and reference to emotion theories provide the basis for extending the framework by modelling further emotional phenomena.

7.2 Future Work

As already mentioned in section 6.4, the continuation of the research endeavour described in this thesis can be carried out in a number of directions. The main part of future work is to continue the modelling approach and continually research and refine the commonalities of process architectures found by extending the covered emotional phenomena. Apart from this avenue of future research, however, several other opportunities can be identified.

One such direction for future work is the use of formalisation methods for concurrent processes, i.e. process algebras such as Communicating Sequential Processes (CSP) [Hoare 1985]. Such methods would offer a more rigorous approach to define the interactions of processes that form the core of this thesis. However, at the current state of development such a formalisation would likely not provide any immediate added value or even hinder the requirement to keep the process structures as flexible as possible and was therefore not yet undertaken.

A promising area for improvement is also the emergence of new technologies and software packages suitable for the implementation of such systems. One example is the project PyPy[24], an open source project, the development of which was also funded as an EU project. Some of the goals of PyPy are support for cooperative microthreading on the language level as well as provisions for real concurrency primitives, constraint programming, and logic variables. All of these can be harnessed to simplify the structure of the modelling framework BehBehBeh. Currently however, the PyPy project is not ready for use in applications with real-time interaction requirements and it does not support the necessary interfaces needed for the integration of the subsystems of BehBehBeh. Another candidate technology is the Erlang programming language[25] and its support for concurrency, but the same caveat about the difficulty of interfacing with other system components ap-

plies.

Another important topic is the relation of any affective agent architecture to neuroscientific knowledge about the architectures of the brains of humans and other animals [Granger 2006]. The imaging techniques used in neuroscience are still very limited with respect to questions of process models of cognitive capabilities. However, recent advances have provided much more detailed insights into the structural organisation of the cortex and brain regions involved in affective behaviour. A correspondence between the components of an agent architecture and the structures identified in neuroscientific studies cannot validate the architecture on its own, but it can provide supporting evidence. Further, considerations of the ontogenetic development of such an architecture could inform the design of future iterations.

A promising candidate aspect for future work is the extension of the coverage of 'higher level' capabilities such as complex memory formation, planning, and deliberation. These capabilities were explicitly excluded as formative components by the proposed architectural design principles, but similar facilities should arise for more complex scenarios using the chosen building blocks.

7.3 Summary

This chapter summarised the main contributions of the thesis. I also laid out some of the many potential avenues of future work.

---

[24] More information can be found at http://codespeak.net/pypy/ regarding the software and at http://pypy.org/ regarding to the EU project.

[25] http://www.erlang.org/

A   Microprocesses Package Based on Python and Twisted

This appendix describes the interface of the `microprocesses` package, a Python package that implements cooperative microthreading based on Python generators and the network library Twisted. I implemented the package mainly for the purposes of this thesis. The features of Twisted most relevant for the `microprocesses` package are its centralised event-loop (the `reactor`) and the concept of a deferred result that can be used as an event notification mechanism. Further, Twisted provides for a solid integration of network communication with other system components.

The following paragraph gives a concise description of the `microprocesses` package that presupposes the reader is familiar with Python and Twisted terminology. The rest of the section documents the basic interfaces of the microprocesses package and provides simple examples of use.

Microprocesses are based on Python generators transformed with `twisted.defer.inlineCallbacks`: this is a decorator that turns a generator function into a kind of tasklet that can be run with twisted. Such a microprocess can yield any `Deferred` and will be resumed when the result is available. This mechanism provides an easy way to build arbitrary waiting behaviours and both one-off and persistent communication channels. The latter type of channels uses channel objects whose methods return `Deferred`s for new data received (both blocking and non-blocking variants are available).

The main interfaces of the microprocesses package are implemented by the following classes:

- `Microprocess`

  A `Microprocess` wraps a coroutine (called generator in Python) adding methods so that it can be run as a process and adds methods callable from within the generator for the interaction with schedulers and other processes.

- `Task`

A `Task` wraps a `Microprocess` while actually running and adds an external interface for activating, pausing, cancelling, and checking the result of the process. Any scheduling decisions are delegated to a `Scheduler`.

- `Scheduler`

  A `Scheduler` is used to create `Tasks` from `Microprocess`es. It is responsible for the scheduling decisions, and subclasses can provide variants, e.g. versions that can be influenced from within proceses.

In summary, a `Microprocess` is the function-like definition of what can be run. To run it, you need to create a `Task`, either by directly adding a `Microprocess` to a `Scheduler` or from within a `Microprocess`. Either way, each `Task` is associated with a `Scheduler` that is responsible for (re)scheduling time-slices. The `Scheduler` can also be used to establish communication channels to other running `Microprocess`es.

Similar to the use in Erlang, the term 'process' is often used when threads of execution share no data with each other, while 'thread' is used when they share data in some way. So while the microprocesses package is an implementation of what is commonly called cooperative microthreading, the user is encouraged to use explicit communication between active elements (although data sharing is not prohibited by the language). It is for this reason that the notion of 'processes' is used in the name.

Figure A.1 shows the most basic interface of a `Microprocess`. Yielding `Deferred`s inside the `__call__` method of an actual `Microprocess` (such as the one returned by the `pause` method) possibly causes the responsible scheduler to halt execution until the occurrence of some event. Examples of other yieldable objects used in the agent implementations are requests to acquire a resource, requests to resources to perform a status update or some other interaction, and requests to communicate with another process. Furthermore, newly created processes expose at least one `Deferred` that allows other processes to wait for their result.

```
class IMicroprocess(Interface):
    '''A microprocess can be run pseudo-concurrently with other
    microprocesses (and anything run by twisted).
    The __call__ method provided by subclasses needs to be a
    generator function (i.e. it must contain the yield keyword).
    '''

    task = Attribute('''The task this microprocess is running as.
                        None until actually scheduled.''')

    def __call__(*args, **kwargs):
        '''This is run when this microprocess' task is activated.
        It needs to be a generator and is automatically wrapped
        (by a metaclass) as a twisted-compatible function
        wrapping the generator.
        '''

    def pause(delay=None):
        '''Pause for delay seconds. If delay is None, pause for
        an interval determined by the scheduler.
        The actual delay (and possibly the meaning of 'seconds')
        depends on the scheduler.
        '''
```

Figure A.1:   The IMicroprocess interface.

B Resources and Interprocess Communication

This section describes the implementation of resources and interprocess communication in BehBehBeh. Resources are implemented as simple objects protected with mutual exclusion locks. However, there are some differences to the usual mechanisms of concurrency primitives in programming languages that support threads. Resources in BehBehBeh can be acquired synchronously or asynchronously, i.e. in a blocking or non-blocking manner, where the latter variant uses the `Deferred` mechanism of Twisted while the former signals non-availability with an exception that also references the current holder of the resource. Blocking calls can be instrumented with a timeout. This allows for checking the availability of a resource.

To allow for intervention into resource acquisition by (for example, affective) processes, the queue of processes waiting for a specific resource can be influenced by promoting a specific process to the front. Figure B.1 shows the interface of a resource provider object that is responsible for providing access to a group of resources. Details about the implementation of specific resources that were modelled in the BehBehBeh scenario are described in Appendix C.

Interprocess communication, including the notification of processes, is implemented using the `Deferred` mechanism. Any process that wants to accept messages can create a communication channel and publishes it under a unique name. Note that this choice of using unique names to denote communication channels was an implementation shortcut that later implementations should replace with an explicit setup of communication links. Any other process that has access to the channel can send messages to the process. The channels themselves can accept either a single message (once or repeatedly) or they can use a queue to accept messages indefinitely and without blocking.

```python
class ResourceProvider(object):
    '''Provide access to a group of resources
    by acquiring and releasing them.'''

    Resource = Enum('HEAD', 'TRUNK', 'EYE')

    def acquireNow(self, proc, resourcename):
        '''Try to acquire a resource immediately,
        raise an exception if already taken.
        `proc` is the microprocess trying to acquire.
        '''

    def acquire(self, proc, resourcename, timeout=None):
        '''Acquire a resource, block if already taken.
        `timeout` can be a positive amount of seconds to wait,
        after the period elapses an exception is raised.
        `proc` is the microprocess trying to acquire.
        '''

    def promoteAcquisition(self, acqdeferred, resourcename):
        '''Move the DeferredAcquisition `acqdeferred` to the
        front of the queue for `resourcename`.
        '''

    def release(self, proc, resource):
        '''Release a resource, fire any queued acquisition attempts.
        `proc` is the microprocess trying to release.
        '''
```

Figure B.1:   Example interface of a resource provider object.

C   Physics and 3D Engine Integration in BehBehBeh

The implementation of the environment simulation for the scenario of BehBehBeh is based on the open source graphics engine Ogre3D and the physics engines Newton Dynamics and Bullet, exposed to Python by the language binding Python-Ogre. 3D modelling as necessary to create the environment and the agent models and animation skeletons was performed using the open source modelling application Blender.

In order to provide a solid foundation for the modelling and simulation approach presented in this thesis, it was necessary to ensure a consistent simulation runtime (e.g. using standard techniques such as frame-rate independent graphics update based on physics engine calculations) and to decouple agent and environment simulation while retaining an interaction that is as immediate as possible. For this I used the network library Twisted and Foolscap, a communication protocol based on it. As mentioned in Appendix A, Twisted was already instrumental in the implementation of a simple and controllable cooperative microthreading package.

Another requirement for the simulation approach presented is based on the assumption of agents that are physically situated. This assumption also holds for the user that is interacting with the system using a character controller. Thus it was necessary to implement a character controller that never violates physics constraints. Although the interaction allowed in the system currently is rather simple (i.e. movement through the world and the application of a force to objects), this is still a deviation from standard character controllers in most interactive 3D programs. One obvious consequence is that when the user pauses the simulated environment, he or she cannot move anymore, i.e. (nearly) the same rules apply for the observing user as for the simulated agents themselves. The user is present in the environment as a physical object (currently only as an oblong trunk). He or she can see parts of his or her body and its shadow, he or she can grab and move objects as well as bump into them, *and* he or she will be sensed as part of the environment by the agents. The second part of controlling characters involves

the control of the bodies of the agents. For that purpose, I implemented a simple ragdoll model (dependent on the capabilities of the physics engine used) that is only controlled by applying forces and torques (and thus does not violate physics constraints). The only concession to ease control is the use of a so-called `UpVector` constraint that keeps the central part of an agent's bone structure upright at all times. Other than that, all control of agent characters is done through the application of forces and torques to individual bones. Note that while this results in awkward movements at best in simple implementations, this is an advantage from the viewpoint of the modelling approach of this thesis. The resulting necessity of continuous control corresponds to the same requirement of real agents in the wild.

The artificial vision of agents also had to comply with the situatedness assumption. To simplify the object identification processes of the agents, false-colour rendering was used, i.e. there is a defined one-to-one correspondence between sensed pixel values and object identities. The *render-to-texture* capabilities of current commodity graphics cards do not allow for a high sampling rate (i.e. reading the rendered texture back into memory to make it accessible to simulated agents) if several agents are to be simulated on the same machine. However, the simplification of false-colour rendering, a reduction of resolution, and adequate caching mechanisms enable access to a rendering of the subjective viewpoint of about 3 agents 10 times per second[26]. While this provides in no way an adequate model of human vision, it fulfills the requirements put down for this work, in that *perception* is not a fully automated process, but rather an active one that consumes resources and is limited to the current subjective viewpoint. The salient features of objects perceived are then their colour, their pixel count, and the pixel centre (i.e., the average of the pixel positions relative to the current viewport). The latter centre is then used to approximate, e.g., the distance between objects and their direction of motion. Artificial hearing and smell were implemented using spheres of influence with a sharp cut-off limit. Future work can include the extension of these models using 3D sound libraries and corresponding hardware support.

Such a simulation environment that depends on the integration of many software components is much easier to manage if there are continuous tests that give an indication about possible regressions or other problems. For this purpose, buildbot was used to setup automatic testing (using unit tests, Python doctests, and pylint) and for building the

---

[26] This is of course only anecdotal evidence and highly dependent on the actual hardware system used (in this case, an ATI Mobility Radeon 9700 graphics card on a Pentium M 735 notebook).

documentation (using epydoc and docutils).

I would like to thank all the contributors of the great software I could freely use to implement the main test scenario of this thesis. Table C.1 lists these packages and their current websites.

Table C.1: Software used to design and implement BehBehBeh.

Python	http://www.python.org/
Ogre 3D	http://www.ogre3d.org/
Bullet	http://www.bulletphysics.com/
Newton	http://www.newtondynamics.com/
Python-Ogre	http://www.pythonogre.com/
Blender	http://www.blender.org/
Twisted	http://twistedmatrix.com/
zope.interface	http://pypi.python.org/pypi/zope.interface
Foolscap	http://foolscap.lothar.com/trac
Buildbot	http://buildbot.net/
docutils	http://docutils.sourceforge.net/
epydoc	http://epydoc.sourceforge.net/
pylint	http://www.logilab.org/project/pylint

Bibliography

Unless otherwise noted, the URLs specified throughout the thesis were last checked on 2008-12-16. At the end of each bibliography entry is a list of the pages that reference that entry.

[Agre Chapman 1987] Philip E Agre and David Chapman: Pengi: An Implementation of a Theory of Activity. In *Proceedings of the 6th National Conference on Artificial Intelligence, AAAI-87 Seattle Washington*, pp. 268–272. Morgan Kaufmann Los Altos/Palo Alto/San Francisco, 1987. 57, 124

[Agre Chapman 1991] Philip E Agre and David Chapman: What are plans for? In Maes P. (ed.): *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, pp. 17–34. MIT Press Cambridge MA, 1991. 57, 124

[Agre Horswill 1997] Philip E Agre and Ian Horswill: Lifeworld Analysis. *Journal Of Artificial Intelligence Research*, 6:111–145, 1997. 15, 53, 59, 124

[Althaus Christensen 2003] Philipp Althaus and Henrik I Christensen: Behaviour Coordination in Structured Environments. *Advanced Robotics Journal*, 17(7):657–674, 2003. 56, 68, 79, 124

[Anderson 2003] Michael L Anderson: Embodied Cognition: A field guide. *Artificial Intelligence*, 149:91–130, 2003. 15, 53, 124

[André et al. 2000] Elisabeth André, Martin Klesen, Patrick Gebhard, Steve Allen, and Thomas Rist: Integrating Models of Personality and Emotions Into Lifelike Characters. In *Affective Interactions - Towards a New Generation of Computer Interfaces, LNCS 1814*, pp. 150–165. Springer Berlin Heidelberg, 2000. 44, 67, 124

[Arbib Fellous 2004] Michael A Arbib and Jean-Marc Fellous: Emotions: from brain to robot. *Trends in Cognitive Sciences*, 8(12):554–561, 2004. 13, 124

[Arkin 1998] Ron C Arkin: *Behavior-Based Robotics*. MIT Press, Cambridge Massachusetts London England, 1998. 54, 124

[Arkin 2005] Ronald C Arkin: Moving Up the Food Chain: Motivation and Emotion in Behavior-Based Robots. In Jean-Marc Fellous and Michael A Arbib (eds.): *Who Needs Emotions? The Brain Meets the Robot*, pp. 245–269. Oxford University Press, 2005. 72, 125

[Arnold 1960] Magda B Arnold: *Emotion and Personality*. Columbia University Press New York, (Vols. I and II), 1960. 42, 125

[Arzi-Gonczarowski 2002] Zippora Arzi-Gonczarowski: AI Emotions: Will One Know Them When One Sees Them? In Robert Trappl (ed.): *Cybernetics and Systems 2002 - Proceedings of the Sixteenth European Meeting on Cybernetics and Systems Research*, pp. 739–744. Austrian Society for Cybernetic Studies, Volume 1+2, 2002. 29, 125

[Arzi-Gonczarowski 2004] Zippora Arzi-Gonczarowski: From Embodiments Back to their Models: An Affective Abstraction. In Alan Schultz (ed.): *The Intersection of Cognitive Science and Robotics: From Interfaces to Intelligence, Papers from the 2004 AAAI Fall Symposium FS-04-05*, pp. 76–81. AAAI Press, 2004. 34, 125

[Aubé 1998] Michel Aubé: A Commitment Theory of Emotions. In Lola Cañamero (ed.): *Emotional and Intelligent: The Tangled Knot of Cognition, AAAI FS-98-03*, pp. 13–18. AAAI Press, Menlo Park CA USA, 1998. 59, 125

[Avila-Garcia Cañamero 2005] Orlando Avila-Garcia and Lola Cañamero: Hormonal Modulation of Perception in Motivation-Based Action Selection Architectures. In Lola Cañamero (ed.): *Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action*, pp. 9–16. AISB The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, University of Sussex, Falmer Brighton UK EU, 2005. 68, 79, 125

[Avradinis Aylett 2003] Nikos Avradinis and Ruth Aylett: Agents with no aims: Motivation-driven continuous planning. In Thomas Rist, Ruth Aylett, Daniel Ballin, and Jeff Rickel (eds.): *Intelligent Virtual Agents 4th International Workshop, IVA 2003, Kloster Irsee Germany, September 15-17 2003, Proceedings*, pp. 269–273. Springer Berlin/Heidelberg/New York, LNCS 2792, 2003. 32, 57, 83, 125

[Avradinis et al. 2004] Nikos Avradinis, Themis Panayiotopoulos, and Spyros Vosinakis: Synthetic Characters with Emotional States. In Vouros G.A. and Panayiotopoulos T. (eds.): *Methods and Applications of Artificial Intelligence, Third Helenic Conference on AI, SETN 2004, Samos, Greece, May 5-8, 2004, LNAI 3025*, pp. 505–514. Springer-Verlag Berlin Heidelberg, 2004. 67, 78, 125

Bibliography

[Avradinis et al. 2005] Nikos Avradinis, Themis Panayiotopoulos, and Ruth Aylett: Continuous Planning for Virtual Environments. In Ioannis Vlahavas and Dimitris Vrakas (eds.): *Intelligent Techniques for Planning*, chapter 5. Idea Group, 2005. 54, 83, 126

[Aylett 1999] Ruth Aylett: Behavioural Virtual Agents. In Michael J Wooldridge and Manuela Veloso (eds.): *Artificial Intelligence Today: Recent Trends and Developments*, pp. 1–12. Springer Berlin Heidelberg New York Tokyo, LNAI 1600, 1999. 34, 126

[Aylett Jones 1996] Ruth Aylett and Steve Jones: Planner and Domain: Domain Configuration for a Task Planner. *International Journal of Expert Systems*, 9(2):279–318, 1996. 52, 126

[Aylett et al. 2000] Ruth Aylett, Alexandra M Coddington, and GJ Petley: Agent-based Continuous Planning. In *Proceedings of the 19th Workshop of the UK Planning and Scheduling Special Interest Group (PLANSIG 2000)*. Open University Milton Keynes UK, 2000. 83, 126

[Aylett 2000] Ruth S Aylett: Emergent Narrative, Social Immersion and "Storification". In Paul Brna (ed.): *Proceedings of Narrative Interaction for Learning Environments NILE, Edinburgh*. Aug.30-Sep.1 2000, 2000. 32, 126

[Bach 2002] Joscha Bach: Enhancing Perception and Planning of Software Agents with Emotion and Acquired Hierarchical Categories. In Gabriela Lindemann, Catholijn Jonker, and Ingo J Timm (eds.): *Proceedings of Modelling Artificial Societies and Hybrid Organizations, MASHO'02, at the 25th German Conference on Artificial Intelligence KI-2002*, pp. 3–12. Karlsruhe Germany, 2002. 70, 126

[Bach 2003] Joscha Bach: The MicroPsi Agent Architecture. In Frank Detje, Dietrich Dörner, and Harald Schaub (eds.): *Proceedings of the 5th International Conference on Cognitive Modeling ICCM-5, Bamberg Germany, April 10-12 2003*, pp. 15–20. Universitäts-Verlag, Bamberg Germany, 2003. 70, 126

[Bach et al. 2006] Joscha Bach, Dietrich Dörner, and Ronnie Vuine: Psi and Micro Psi a Novel Approach to Modeling Emotion and Cognition in a Cognitive Architecture. In *Tutorial Notes at the 7th International Conference on Cognitive Modeling (ICCM 2006), April 5 2006*. University of Trieste Italy, 2006. 70, 126

[Bard Rioch 1937] Archibald P Bard and David M Rioch: A Study of Four Cats Deprived of Neocortex and Additional Portions of the Forebrain. *John Hopkins Hospital Bulletin*, 60:73–147, 1937. 41, 126

[Bargh Chartrand 1999] John A Bargh and Tanya L Chartrand: The Unbearable Automaticity of Being. *American Psychologist*, 54(7):462–479, 1999. 52, 56, 59, 127

[Barnes et al. 1997] David P Barnes, Robert A Ghanea-Hercock, Ruth S Aylett, and Alexandra M Coddington: Many hands make light work? An investigation into behaviourally controlled co-operant autonomous mobile robots. In W Lewis Johnson (ed.): *Proceedings of the 1st International Conference on Autonomous Agents, Marina del Rey CA USA, February 05-08 1997*, pp. 413–420. ACM Press New York NY USA, 1997. 68, 127

[Barsalou et al. 2007] Lawrence W Barsalou, Cynthia Breazeal, and Linda B Smith: Cognition as coordinated non-cognition. *Cognitive Processing*, 8(2):79–91, 2007. 13, 86, 127

[Bartl Dörner 1998] Christina Bartl and Dietrich Dörner: Comparing the Behaviour of PSI with Human Behaviour in the Bio Lab Game. Lehrstuhl Psychologie II (Allgemeine Psychologie und Methodenlehre), Univ. Bamberg, Memorandum 32, 1998. URL `http://www.uni-bamberg.de/~ba2dp1/online.doc/memo32.rtf.zip`. 70, 80, 127

[Bartneck 2002] Christoph Bartneck: Integrating the OCC Model of Emotions in Embodied Characters. In Marriott A., Beard S., Lavagetto F., and Barakonyi I. (eds.): *Workshop Notes, Virtual Conversational Characters: Applications, Methods, and Research Challenges, November 29 2002*, chapter 10. Melbourne Australia, 2002. 44, 127

[Bates 1992] Joseph Bates: The Nature of Characters in Interactive Worlds and the Oz Project. Technical report, School of Computer Science, Carnegie Mellon University, CMU-CS-92-200, 1992. URL `http://www.cs.cmu.edu:8001/afs/cs/project/oz/web/papers/loeffler.ps`. 66, 127

[Bates 1994] Joseph Bates: The Role of Emotion in Believable Agents. *Special Issue: Intelligent Agents, Communications of the ACM*, 37(7): 122–125, 1994. 71, 127

[Bates et al. 1993] Joseph Bates, J Altucher, A Hauptman, M Kantrowitz, A Bryan Loyall, K Murakami, P Olbrich, Z Popovic, W Scott Reilly, Phoebe Sengers, W Welch, Peter Weyrauch, and A Witkin: Edge of Intention. In Linehan T.E. (ed.): *Visual Proceedings Computer Graphics Annual Conference Series SIGGRAPH 1993*, pp. 113–114. ACM New York, 1993. 44, 66, 127

Bibliography

[Becker-Asano Wachsmuth 2008] Christian Becker-Asano and Ipke Wachsmuth: Affect Simulation with Primary and Secondary Emotions. In Helmut Prendinger, James Lester, and Mitsuru Ishizuka (eds.): *Intelligent Virtual Agents 8th International Conference, IVA 2008, Tokyo Japan, September 1-3 2008, Proceedings*, pp. 15–28. Springer LNCS 5208, 2008. 67, 128

[Bickhard 1999] Mark H Bickhard: Motivation and Emotion: An Interactive Process Model. Technical report, Department of Philosophy, Lehigh University, Bethlehem PA, 1999. URL http://www.lehigh.edu/~mhb0/motemotion.html. 89, 128

[Blumberg 1994] Bruce Blumberg: Action Selection in Hamsterdam Lessons From Ethology. In David Cliff, Philip Husbands, Jean Arcady Meyer, and Stewart Wilson (eds.): *Proceedings of the 3rd International Conference on the Simulation of Adaptive Behavior, Brighton UK, August 8-12 1994*, pp. 108–117. MIT Press, 1994. 67, 128

[Bowlby 1969] John Bowlby: *Attachment, 2nd edition, Attachment and Loss (vol. 1, reprinted 1999)*. New York Basic Books, 1969. 98, 128

[Brom Bryson 2006] Cyril Brom and Joanna Bryson: Action Selection for Intelligent Systems. euCognition: European Network for the Advancement of Artificial Cognitive Systems, White Paper, 2006. URL http://www.eucognition.org/asm-whitepaper-final-060804.pdf. 55, 128

[Brooks 1991] Rodney A Brooks: Intelligence Without Representation. *Special Volume: Foundations of Artificial Intelligence, Artificial Intelligence*, 47(1-3):139–159, 1991. 51, 54, 57, 128

[Bryson 2000] Joanna J Bryson: Hierarchy and Sequence vs. Full Parallelism in Action Selection. In Meyer J.-A. et al. (ed.): *From Animals to Animats 6 (SAB 2000)*, pp. 147–156. MIT Press Cambridge MA USA, 2000. 60, 68, 81, 128

[Bryson 2001] Joanna J Bryson: *Intelligence by Design - Principles of Modularity and Coordination for Engineering Complex Adaptive Agents*. PhD thesis, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, AI Technical Report 2001-003, 2001. 68, 128

[Bryson 2003] Joanna J Bryson: The Behavior-Oriented Design of Modular Agent Intelligence. In Kowalczyk R., Mueller J.P., Tianfield H., and Unland R. (eds.): *Agent Technologies, Infrastructures, Tools, and Applications for E-Services: NODE 2002 Agent-Related Workshops, Erfurt Germany, October 7-10 2002, Revised Papers*, pp. 61–76. Springer Verlag, 2003. 68, 128

[Burke et al. 2001] Robert Burke, Damian Isla, Marc Downie, Yuri Ivanov, and Bruce Blumberg: Creature Smarts: The Art and Architecture of a Virtual Brain. In *Proceedings of the Game Developers Conference*, pp. 147–166. San Jose CA, 2001. 67, 129

[Campos Stenberg 1981] Joseph J Campos and Craig R Stenberg: Perception, Appraisal and Emotion: The Onset of Social Referencing. In Michael E Lamb and Lonnie R Sherrod (eds.): *Infant Social Cognition - Empirical and Theoretical Considerations*, pp. 273–314. Erlbaum Hillsdale NJ, 1981. 98, 129

[Cannon 1927] Walter B Cannon: The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *American Journal of Psychology*, 39:106–124, 1927. 41, 129

[Cassell 2001] Justine Cassell: Embodied Conversational Agents: Representation and Intelligence in User Interface. *Special Issue on Intelligent User Interfaces, AI Magazine*, 22(4):67–83, 2001. 25, 129

[Cassell et al. 2000] Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth F Churchill (eds.): *Embodied Conversational Agents*. MIT Press Cambridge Boston London, 2000. 25, 129

[Cavazza et al. 2002] Marc Cavazza, Fred Charles, and Steven J Mead: Character-Based Interactive Storytelling. *IEEE Intelligent Systems*, 17(4):17–24, 2002. 31, 66, 129

[Cavazza et al. 2003] Marc Cavazza, Olivier Martin, Fred Charles, Steven J Mead, and Xavier Marichal: Interacting with Virtual Agents in Mixed Reality Interactive Storytelling. In Thomas Rist, Ruth Aylett, Daniel Ballin, and Jeff Rickel (eds.): *Intelligent Virtual Agents 4th International Workshop, IVA 2003, Kloster Irsee Germany, September 15-17 2003, Proceedings*, pp. 231–235. Springer Berlin/Heidelberg/New York, LNCS 2792, 2003. 33, 129

[Chartrand Bargh 2002] Tanya L Chartrand and John A Bargh: Nonconscious motivations: Their activation, operation, and consequences. In Tesser A., Stapel D., and Wood J. (eds.): *Self and motivation: Emerging psychological perspectives*, pp. 13–41. American Psychological Association Press Washington D.C. USA, 2002. 52, 129

[Chrisley Ziemke 2003] Ron Chrisley and Tom Ziemke: Embodiment. In Nadel L. (ed.): *Encyclopedia of Cognitive Science*, pp. 1102–1108. Nature Publishing Group/Macmillan London, 2003. 14, 15, 129

[Clancey 1999] William J Clancey: *Conceptual Coordination - How the Mind Orders Experience in Time*. Lawrence Erlbaum Associates Publishers, Mahwah NJ London UK, 1999. 59, 129

Bibliography

[Clancey 2002] William J Clancey: Simulating Activities: Relating Motives, Deliberation, and Attentive Coordination. *Special Issue on Situated and Embodied Cognition, Cognitive Systems Research*, 3(3):471–499, 2002. 51, 130

[Clark 1997] Andy Clark: *Being There: Putting Brain, Body and World Together Again*. MIT Press/Bradford Books Cambridge London, 1997. 51, 53, 57, 130

[Cleeremans French 1996] Axel Cleeremans and Robert M French: From Chicken Squawking to Cognition. *Psychologica Belgica*, 36(1-2):5–29, 1996. 14, 25, 130

[Cooper 1999] Alan Cooper: *The Inmates Are Running the Asylum*. SAMS publishing, 1999. 26, 130

[Crowley 2006] James L Crowley: Social Perception - Modeling human interaction for the next generation of communication services. *ACM Queue*, 4(6):34–43, 2006. 26, 130

[Dalgleish 2004] Tim Dalgleish: The Emotional Brain. *Nature Reviews Neuroscience*, 5(7):583–589, 2004. 42, 130

[Dalgleish Power 1999] Tim Dalgleish and Mick J Power: *Handbook of Cognition and Emotion*. Wiley Chichester/London/New York, 1999. 42, 130

[Darwin 1872] Charles Darwin: *The Expression of Emotions in Man and Animals*. John Murray London, 1872. 41, 130

[Davidson et al. 2003] Richard J Davidson, Klaus R Scherer, and H Hill Goldsmith (eds.): *Handbook of Affective Sciences*. Oxford University Press, Oxford New York, 2003. 42, 130

[Davies Stone 1998] Martin Davies and Tony Stone: Folk Psychology and Mental Simulation. In Anthony O'Hear (ed.): *Contemporary Issues in the Philosophy of Mind, Royal Institute of Philosophy Supplements (No. 43)*, chapter 3. Cambridge University Press Cambridge UK/New York, 1998. 41, 130

[Dawkins 2006] Marian Stamp Dawkins: Through animal eyes: What behaviour tells us. *Applied Animal Behaviour Science*, 100(1-2):4–10, 2006. 53, 130

[Dewey 1896] John Dewey: The Reflex Arc Concept in Psychology. *Psychological Review*, 3:357–370, 1896. 48, 52, 54, 130

[Dolan 2002] Ray J Dolan: Emotion, Cognition, and Behavior. *Science*, 298: 1191–1194, 2002. 13, 131

[Donaldson et al. 2004] Toby Donaldson, Andrew Park, and I Ling Lin: Emotional Pathfinding. In Tawfik A.Y. and Goodwin S.D. (eds.): *Canadian AI 2004, LNAI 3060*, pp. 31–43. Springer-Verlag Berlin Heidelberg, 2004. 69, 131

[Doyle 1996] Jon Doyle: Toward Rational Planning and Replanning - Rational Reason Maintenance, Reasoning Economies, and Qualitative Preferences. In Austin T. (ed.): *Advanced Planning Technology: Technological Achievements of the ARPA/Rome Laboratory Planning Initiative*, pp. 130–135. AAAI Press Menlo Park California, 1996. 57, 131

[Doyle 1999] Jon Doyle: Rational Decision Making. In Robert Andrew Wilson and Frank C Keil (eds.): *The MIT Encyclopedia of the Cognitive Sciences*, pp. 701–703. MIT Press, Cambridge/Boston/London, 1999. 58, 131

[Dörner 2002] Dietrich Dörner: *Die Mechanik Des Seelenwagens - Eine neuronale Theorie der Handlungsregulation*. Verlag Hans Huber, Bern Goettingen Toronto Seattle, 2002. 70, 131

[Dörner 2003] Dietrich Dörner: The Mathematics of Emotion. In Frank Detje, Dietrich Dörner, and Harald Schaub (eds.): *Proceedings of the 5th International Conference on Cognitive Modeling ICCM-5, Bamberg Germany, April 10-12 2003*, pp. 75–80. Universitäts-Verlag, Bamberg Germany, 2003. 70, 131

[Dyer 1987] Michael G Dyer: Emotions and their computations: Three computer models. *Cognition & Emotion*, 1(3):323–347, 1987. 65, 131

[Egges et al. 2003] Arjan Egges, Sumedha Kshirsagar, and Nadia Magnenat-Thalmann: A Model for Personality and Emotion Simulation. In *Knowledge-Based Intelligent Information and Engineering Systems: 7th International Conference, KES 2003 Oxford UK, September 3-5 2003 Proceedings, LNCS 2773*, pp. 453–461. Springer-Verlag Berlin Heidelberg New York, 2003. 67, 131

[Egges et al. 2004] Arjan Egges, Sumedha Kshirsagar, and Nadia Magnenat-Thalmann: Generic Personality and Emotion Simulation for Conversational Agents. *Computer Animation and Virtual Worlds*, 15(1):1–13, 2004. 67, 131

[Egri 1946] Lajos Egri: *The Art of Dramatic Writing*. Touchstone Book, New York NY, 1946. 31, 131

Bibliography

[Elliott 1994] Clark Elliott: Research Problems in the Use of a Shallow Artificial Intelligence Model of Personality and Emotion. In Barbara Hayes-Roth and Richard E Korf (eds.): *Proceedings of the 12th National Conference on Artificial Intelligence AAAI-94, Seattle Washington*, pp. 9–15. AAAI Press/MIT Press Cambridge Menlo Park, 1994. 44, 132

[Elliott 1997a] Clark Elliott: Hunting for the Holy Grail with Emotionally Intelligent Virtual Actors. ACM Intelligence, 1(1), 1997a. URL http://www.depaul.edu/~elliott/papers/intelligence/intelligence.ps. 65, 74, 132

[Elliott 1997b] Clark Elliott: Affective Reasoner Personality Models for Automated Tutoring Systems. In *Proceedings of the Workshop on Pedagogical Agents at the 8th World Conference on Artificial Intelligence in Education (AI&ED-97), Kobe Japan*. IOS Press Amsterdam, 1997b. 65, 132

[Elliott et al. 1998] Clark Elliott, Jacek Brzezinski, Sanjay Sheth, and Robert Salvatoriello: Story Morphing in the Affective Reasoning Paradigm - Generating Stories semi-automatically for use with 'emotionally intelligent' multimedia agents. In Katia P Sycara and Michael Wooldridge (eds.): *Proceedings of the Second International Conference on Autonomous Agents, St.Paul Minneapolis USA, May 9-13 1998*, pp. 181–188. ACM Press New York, 1998. 31, 132

[Elliott et al. 1999] Clark Elliott, Jeff Rickel, and James Lester: Lifelike Pedagogical Agents and Affective Computing - An Exploratory Synthesis. In Michael J Wooldridge and Manuela Veloso (eds.): *Artificial Intelligence Today: Recent Trends and Developments*, pp. 195–212. Springer Berlin Heidelberg New York Tokyo, LNAI 1600, 1999. 44, 132

[Elliott 1992] Clark D Elliott: *The Affective Reasoner a Process Model of Emotions in a Multi Agent System*. PhD thesis, Northwestern University Illinois, 1992. URL http://condor.depaul.edu/~elliott/ar/papers/dis/elliott-phd.html. 65, 132

[Ellsworth Scherer 2003] Phoebe C Ellsworth and Klaus R Scherer: Appraisal Processes in Emotion. In Richard J Davidson, Klaus R Scherer, and H Hill Goldsmith (eds.): *Handbook Of Affective Sciences*, chapter 29, pp. 572–595. Oxford University Press, Oxford New York, 2003. 34, 42, 43, 132

[Feinman 1982] Saul Feinman: Social Referencing in Infancy. *Merrill-Palmer Quarterly*, 28:445–470, 1982. 98, 132

[Ferguson 1992] Innes A Ferguson: Touring Machines: Autonomous Agents with Attitudes. Technical report, Computer Laboratory, University of Cambridge UK, Technical Report TR 250, expanded version of 'TouringMachines: Autonomous Agents with Attitudes', IEEE Computer 25(5):51-55 (May), 1992. URL http://www.cl.cam.ac.uk/Research/Papers/reports/TR250-iaf-touringmachines.ps.Z. 57, 133

[Forgas 2000] Joseph P Forgas: *Feeling and Thinking - The Role of Affect in Social Cognition*. Cambridge University Press, Editions de la Maison des Sciences de l'Homme, Paris, 2000. 42, 133

[Frijda 1986] Nico H Frijda: The Emotions. Cambridge University Press, Editions de la Maison des Sciences de l'Homme, Paris, 1986. 29, 34, 42, 43, 46, 49, 53, 84, 98, 133, 152

[Frijda 1993] Nico H Frijda (ed.): *Appraisal and Beyond the Issue of Cognitive Determinants of Emotion*. Cognition & Emotion 7(3&4) May/July, 1993. 42, 133

[Frijda 2005] Nico H Frijda: Emotion Experience. *Cognition and Emotion*, 19(4):473–497, 2005. 54, 133

[Frijda 2007] Nico H Frijda: *The Laws of Emotion*. Lawrence Erlbaum Associates Publishers, Mahwah NJ USA London UK EU, 2007. 12, 15, 16, 34, 88, 91, 133, 153

[Frijda Swagerman 1987] Nico H Frijda and Jaap Swagerman: Can computers feel? Theory and design of an emotional system. *Cognition & Emotion*, 1(3):235–257, 1987. 65, 133

[Fum et al. 2007] Danilo Fum, Fabio Del Missier, and Andrea Stocco: The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, 8(3): 135–142, 2007. 11, 25, 133

[Gadanho 2003] Sandra Clara Gadanho: Learning Behavior-Selection by Emotions and Cognition in a Multi-Goal Robot Task. *Journal Of Machine Learning Research*, 4:385–412, 2003. 55, 69, 133

[Gat 1997] Erann Gat: On Three-Layer Architectures. In Kortenkamp D., Bonasso R.P., and Murphy R. (eds.): *Artificial Intelligence and Mobile Robots*, pp. 195–210. MIT/AAAI Press, 1997. 50, 68, 81, 133

[Gat Pell 1998] Erann Gat and B Pell: Smart Executives for Autonomous Spacecraft. *IEEE Intelligent Systems*, 13(5):56–61, 1998. 68, 133

[Gebhard 2005] Patrick Gebhard: ALMA: A Layered Model of Affect. AA-MAS2005Proceedings, pp.29-36, 2005. 67, 134

[Glass Grosz 2000] Alyssa Glass and Barbara Grosz: Socially Conscious Decision Making. In Charles Sierra, Maria Gini, and Jeffrey S Rosenschein (eds.): *Proceedings of the 4th International Conference on Autonomous Agents, June 3-7 2000, Barcelona Catalonia Spain*, pp. 217–224. ACM, 2000. 57, 134

[Glover 2004] Scott Glover: Separate Visual Representations in the Planning and Control of Action. *Behavioral and Brain Sciences*, 27:3–78, 2004. 50, 134

[Goldman 1993] Alvin I Goldman: The Psychology of Folk Psychology. *Behavioral and Brain Sciences*, 16(1):15–28, 1993. 41, 134

[Gollwitzer Bargh 2005] Peter M Gollwitzer and John A Bargh: Automaticity in Goal Pursuit. In *Handbook of Competence and Motivation*, chapter 34, pp. 624–646. Guilford Press New York/London, 2005. 29, 134

[Granger 2006] Richard Granger: Engines of the Brain: The Computational Instruction Set of Human Cognition. *Achieving Human-Level Intelligence through Integrated Systems and Research, AI Magazine*, 27 (2):15–32, 2006. 18, 115, 134

[Gratch 1999] Jonathan Gratch: Why You Should Buy an Emotional Planner. In Juan D Velásquez (ed.): *Workshop: Emotion-Based Agent Architectures (EBAA'99) at the 3rd International Conference on Autonomous Agents, Seattle WA USA, May 1 1999*, pp. 53–60, 1999. 65, 134

[Gratch 2000] Jonathan Gratch: Émile: Marshalling passions in training and education. In Charles Sierra, Maria Gini, and Jeffrey S Rosenschein (eds.): *Proceedings of the 4th International Conference on Autonomous Agents, June 3-7 2000, Barcelona Catalonia Spain*, pp. 325–332. ACM, 2000. 65, 134

[Gratch Marsella 2004a] Jonathan Gratch and Stacy Marsella: A Domain Independent Framework for Modeling Emotion. *Cognitive Systems Research*, 5(4):269–306, 2004a. 33, 65, 134

[Gratch Marsella 2004b] Jonathan Gratch and Stacy Marsella: Technical Details of a Domain Independent Framework for Modelling Emotion. Technical report, Institute for Creative Technologies, Marina del Rey CA USA, ICT-TR-04.2004, 2004b. URL http://www.ict.usc.edu/publications/ICT-TR-04.2004.pdf. 65, 134

[Gratch Marsella 2004c] Jonathan Gratch and Stacy Marsella: Evaluating a General Model of Emotional Appraisal and Coping. In Eva Hudlicka and Lola Cañamero (eds.): *Architectures for Modeling Emotion: Cross-Disciplinary Foundations, Papers from the 2004 AAAI Spring Symposium SS-04-02, March 22-24 2004*, pp. 52–59. AAAI Press, 2004c. 27, 135

[Gratch Marsella 2007] Jonathan Gratch and Stacy Marsella: The Architectural Role of Emotion in Cognitive Systems. In Gray W.D. (ed.): *Integrated Models of Cognitive Systems*, chapter 16. Oxford University Press New York, 2007. 14, 32, 135

[Gratch et al. 2006] Jonathan Gratch, Stacy Marsella, and Wenji Mao: Towards a Validated Model of Emotional Intelligence. In Yolanda Gil and Raymond J Mooney (eds.): *Proceedings of the Twenty-First National Conference on Artificial Intelligence, Boston MA July 16–20 2006*, pp. 1613–1615. AAAI Press Menlo Park CA, 2006. 14, 135

[Gratch et al. 2009] Jonathan Gratch, Stacy Marsella, and Paolo Petta: *Modeling the Cognitive Antecedents and Consequences of Emotion*. Cognitive Systems Research 10(1), 2009. 14, 135

[Griffiths Scarantino 2005] Paul E Griffiths and Andrea Scarantino: Emotions in the wild: The situated perspective on emotion. In Philip Robbins and Murat Aydede (eds.): *Cambridge Handbook of Situated Cognition*, chapter 23. 2005. 30, 135

[Gross 1999] James J Gross: Emotion Regulation: Past, Present, Future. *Cognition & Emotion*, 13(5):551–573, 1999. 61, 135

[Gross John 2003] James J Gross and Oliver P John: Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85(2):348–362, 2003. 53, 61, 135

[Grosz et al. 2004] Barbara J Grosz, Sarit Kraus, Shavit Talman, Boaz Stossel, and Moti Havlin: The influence of social dependencies on decision-making: initial investigations with a new game. In Nick Jennings, Milind Tambe, Carles Sierra, and Liz Sonenberg (eds.): *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pp. 782–789. Vol.2, 2004. 57, 135

[Grush 2004] Rich Grush: The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27 (3):377–396, 2004. 89, 135

[Heise 2002] David R Heise: Understanding Social Interaction with Affect Control Theory. In Berger J. and Zelditch M. (eds.): *New Directions in Contemporary Sociological Theory*, chapter 2. Rowman and Littlefield Boulder CO, 2002. 30, 136

[Heise 2004] David R Heise: Enculturating Agents with Expressive Role Behavior. In Sabine Payr and Robert Trappl (eds.): *Agent Culture*, pp. 127–142. Lawrence Erlbaum Associates Publishers, Mahwah New Jersey London, 2004. 30, 136

[Hill et al. 2001] Randall Hill, William R Swartout, Jonathan Gratch, W Lewis Johnson, Chris Kyriakakis, Catherine La Bore, Richard Lindheim, Stacy Marsella, David Miraglia, Ben Moore, Jackie Morie, Jeff Rickel, Marcus Thiébaux, Larry Tuch, Richard Whitney, and Jay Douglas: Toward the Holodeck: Integrating Graphics, Sound, Character and Story. In Elisabeth André, Sandip Sen, Claude Frasson, and Jörg P Müller (eds.): *Proceedings of the 5th International Conference on Autonomous Agents, May 28-June 1 2001, Montreal Canada*, pp. 409–416. ACM, 2001. 65, 136

[Hoare 1985] Charles Antony Richard Hoare: *Communicating Sequential Processes*. Prentice Hall International, 1985. 114, 136

[Horswill 2008] Ian Horswill: Lightweight Procedural Animation with Believable Physical Interactions. In Mateas M. and Darken C. (eds.): *Proceedings of the Fourth Artificial Intelligence and Interactive Digital Entertainment Conference*, chapter 8. October 22-24 2008 Stanford, 2008. 67, 136

[Horty Pollack 2001] John F Horty and Martha E Pollack: Evaluating New Options in the Context of Existing Plans. *Artificial Intelligence Journal*, 127(2):199–220, 2001. 57, 136

[Huber 1999] Marcus J Huber: JAM: A BDI-theoretic Mobile Agent Architecture. In Oren Etzioni, Jörg P Müller, and Jeffrey M Bradshaw (eds.): *Agents'99 Proceedings of the 3rd Annual Conference on Autonomous Agents, May 1-5 1999, Seattle WA USA*, pp. 236–243. ACM, 1999. 36, 136

[Hutchins 1999] Edwin Hutchins: Cognitive Artifacts. In Robert Andrew Wilson and Frank C Keil (eds.): *The MIT Encyclopedia of the Cognitive Sciences*, pp. 126–128. MIT Press, Cambridge/Boston/London, 1999. 30, 136

[Isla et al. 2001] Damian Isla, Robert Burke, Marc Downie, and Bruce Blumberg: A Layered Brain Architecture for Synthetic Creatures. In Nebel B. (ed.): *Proceedings of the 17th International Joint Conference on*

Artificial Intelligence (IJCAI 2001), Seattle WA USA, August 4-10, pp. 1051–1058. Morgan Kaufmann San Francisco CA USA, 2001. 67, 136

[Iurgel 2003] Ido Iurgel: Virtual Actors in Interactive Storytelling. In Thomas Rist, Ruth Aylett, Daniel Ballin, and Jeff Rickel (eds.): *Intelligent Virtual Agents 4th International Workshop, IVA 2003, Kloster Irsee Germany, September 15-17 2003, Proceedings*, pp. 254–258. Springer Berlin/Heidelberg/New York, LNCS 2792, 2003. 32, 137

[James 1884] William James: What Is an Emotion. *Mind*, 9:188–205, 1884. 41, 137

[Johnson Rickel 1998] W Lewis Johnson and Jeff Rickel: STEVE: An Animated Pedagogical Agent for Procedural Training in Virtual Environments. *SIGART Bulletin*, 8:16–21, 1998. 65, 137

[Jones Wray 2006] Randolph M Jones and Robert E Wray: Comparative Analysis of Frameworks for Knowledge Intensive Intelligent Agents. *Achieving Human-Level Intelligence through Integrated Systems and Research, AI Magazine*, 27(2):45–56, 2006. 54, 137

[Jung Fischer 1998] Christoph G Jung and Klaus Fischer: Methodological Comparison of Agent Models. Technical report, Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarbrücken Germany, DFKI Research Report-98-01, 1998. URL `ftp://ftp.dfki.uni-kl.de/pub/Publications/ResearchReports/1998/RR-98-01.ps.gz`. 57, 137

[Kemper 1993] Theodore D Kemper: Sociological Models in the Explanation of Emotions. In Michael Lewis and Jeannette M Haviland (eds.): *Handbook Of Emotions*, pp. 41–52. Guilford Press New York London, 1993. 59, 60, 137

[Ketelaar Clore 1997] Timothy Ketelaar and Gerald L Clore: Emotion and reason: The proximate effects and ultimate functions of emotions. In Matthews G. (ed.): *Personality, Emotion, and Cognitive Science, Advances in Psychology Series*, pp. 355–396. Elsevier Science Publishers (North-Holland) Amsterdam NL, 1997. 29, 137

[Kort Reilly 2002] Barry Kort and Rob Reilly: Analytical Models of Emotions, Learning and Relationships: Towards an Affect-sensitive Cognitive Machine. In *Proc. International Conference on Virtual Worlds and Simulation (VW Sim 2002), San Antonio, TX, 2002*. MIT Media Lab Vision and Modeling Technical Report 548, 2002. 68, 137

[Kort et al. 2001a] Barry Kort, Rob Reilly, and Rosalind W Picard: External representation of learning process and domain knowledge: Affective

state as a determinate of its structure and function. In *Proceedings from the Artificial Intelligence in Education Workshops (AI-ED 2001), San Antonio, TX, May 2001*. MIT Media Lab Vision and Modeling Technical Report 546, 2001a. 68, 137

[Kort et al. 2001b] Barry Kort, Rob Reilly, and Rosalind W Picard: An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy—Building a Learning Companion. In *Proceedings from the International Conference on Advanced Learning Technologies (ICALT 2001), Madison, Wisconsin, August 2001*. MIT Media Lab Vision and Modeling Technical Report 547, 2001b. 68, 138

[Krenn 2003] Brigitte Krenn: The NECA Project: Net Environments for Embodied Emotional Conversational Agents. In Stefanos Kollias and Kostas Karpouzis (eds.): *Proceedings of the Workshop on emotionally rich virtual worlds with emotion synthesis at the 8th International Conference on 3D Web Technology (Web 3 D), 10.March 2003*. St.Malo France, 2003. 44, 138

[Krenn Schollum 2008] Brigitte Krenn and Christian Schollum: The RAS-CALLI Platform for a Flexible and Distributed Development of Virtual Systems Augmented with Cognition. In *Proceedings of the International Conference on Cognitive Systems (Cog Sys 2008), University of Karlsruhe Germany*. April 2-4, 2008. 110, 138

[Laird et al. 1987] John E Laird, Allen Newell, and Paul S Rosenbloom: SOAR: An Architecture for General Intelligence. *Artificial Intelligence*, 33(1):1–64, 1987. 33, 65, 138

[Lange 1885] Carl G Lange: *Om Sindsbevaegelser Et Psyko Fysiologisk Studie*. Lunds Copenhagen, 1885. 41, 138

[Lawrence Calder 2004] Andrew D Lawrence and Andrew J Calder: Homologizing Human Emotions. In Evans D. and Cruse P. (eds.): *Emotions, Evolution and Rationality*, pp. 15–47. Oxford University Press Oxford UK, 2004. 98, 138

[Lazarus 1991] Richard S Lazarus: *Emotion and Adaptation*. Oxford University Press, London/Oxford/New York, 1991. 49, 138

[LeDoux 2000] Joseph E LeDoux: Emotion Circuits in the Brain. *Annual Review of Neuroscience*, 23:155–184, 2000. 42, 138

[Leontev 1978] Aleksei Nikolaevich Leontev: *Activity, Consciousness, and Personality*. Prentice-Hall London/New York/Englewood Cliffs NJ, 1978. 51, 138

[Leventhal Scherer 1987] Howard Leventhal and Klaus R Scherer: The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition & Emotion*, 1(1):3–28, 1987. 41, 139

[Lewis Haviland 1993] Michael Lewis and Jeannette M Haviland (eds.): *Handbook of Emotions*. Guilford Press New York London, 1993. 42, 139

[Lim et al. 2005] Mei Yii Lim, Ruth Aylett, and Christian Martin Jones: Emergent Affective and Personality Model. In Themis Panayiotopoulos, Jonathan Gratch, Ruth Aylett, Daniel Ballin, Patrick Olivier, and Thomas Rist (eds.): *Intelligent Virtual Agents, 5th International Working Conference, IVA 2005, Kos, Greece, September 2005, Proceedings*, pp. 371–380. Springer Berlin Heidelberg, LNAI 3661, 2005. 70, 139

[Loewenstein Lerner 2003] George Loewenstein and Jennifer S Lerner: The Role of Affect in Decision Making. In Richard J Davidson, Klaus R Scherer, and H Hill Goldsmith (eds.): *Handbook Of Affective Sciences*, chapter 31, pp. 619–642. Oxford University Press, Oxford New York, 2003. 55, 139

[Louchart Aylett 2003] Sandy Louchart and Ruth Aylett: Solving the narrative paradox in VEs – lessons from RPGs. In Thomas Rist, Ruth Aylett, Daniel Ballin, and Jeff Rickel (eds.): *Intelligent Virtual Agents 4th International Workshop, IVA 2003, Kloster Irsee Germany, September 15-17 2003, Proceedings*, pp. 244–248. Springer Berlin/Heidelberg/New York, LNCS 2792, 2003. 32, 139

[Louchart Aylett 2004] Sandy Louchart and Ruth Aylett: Narrative Theory and Emergent Interactive Narrative. *Int. Journal of Continuing Engineering Education and Lifelong Learning*, 14(6):506–518, 2004. 31, 139

[Loyall et al. 2004] A Bryan Loyall, W Scott Reilly, Joseph Bates, and Peter Weyhrauch: System for Authoring Highly Interactive, Personality-Rich Interactive Characters. In Boulic R. and Pai D.K. (eds.): *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation, Grenoble France*, pp. 59–68. ACM Press New York NY USA, 2004. 31, 33, 60, 66, 139

[Machado 2004] Isabel Machado: *Children, Stories and Dramatic Games: A Support and Guidance Architecture for Story Creation*. PhD thesis, University of Leeds, Leeds UK EU, 2004. 31, 139

Bibliography

[MacLean 1949] Paul D MacLean: Psychosomatic disease and the 'visceral brain': recent developments bearing on the Papez theory of emotion. *Psychosomatic Medicine*, 11:338–353, 1949. 42, 140

[Maes et al. 1994] Pattie Maes, Trevor Darrell, Bruce Blumberg, and Sandy Pentland: ALIVE: Artificial Life Interactive Video Environment. In Barbara Hayes-Roth and Richard E Korf (eds.): *Proceedings of the 12th National Conference on Artificial Intelligence AAAI-94, Seattle Washington*, p. 1506. AAAI Press/MIT Press Cambridge Menlo Park, 1994. 110, 140

[Magerko et al. 2004] Brian Magerko, John E Laird, Mazin Assanie, Alex Kerfoot, and Devvan Stokes: AI Characters and Directors for Interactive Computer Games. In Randy Hill and Neil Jacobstein (eds.): *Proceedings of the 2004 Innovative Applications of Artificial Intelligence Conference IAAI, San Jose CA, July 2004*, pp. 877–883. AAAI Press, 2004. 31, 66, 140

[Manovich 2001] Lev Manovich: *The Language of New Media*. MIT Press Cambridge MA, 2001. 25, 140

[Mao Gratch 2003] Wenji Mao and Jonathan Gratch: The Social Credit Assignment Problem. In Thomas Rist, Ruth Aylett, Daniel Ballin, and Jeff Rickel (eds.): *Intelligent Virtual Agents 4th International Workshop, IVA 2003, Kloster Irsee Germany, September 15-17 2003, Proceedings*, pp. 39–47. Springer Berlin/Heidelberg/New York, LNCS 2792, 2003. 37, 140

[Marinier Laird 2006] Robert P. Marinier, III and John E Laird: A Cognitive Architecture Theory of Comprehension and Appraisal. In Robert Trappl (ed.): *Cybernetics and Systems 2006 - Proceedings of the Eighteenth Meeting on Cybernetics and Systems Research, April 18-21 2006, University of Vienna*, pp. 589–594. Austrian Society for Cybernetic Studies Vienna, Volumes 1 and 2, 2006. 33, 69, 140

[Marsella Gratch 2006] Stacy Marsella and Jonathan Gratch: EMA: A Computational Model of Appraisal Dynamics. In Robert Trappl (ed.): *Cybernetics and Systems 2006 - Proceedings of the Eighteenth Meeting on Cybernetics and Systems Research, April 18-21 2006, University of Vienna*, pp. 601–606. Austrian Society for Cybernetic Studies Vienna, Volumes 1 and 2, 2006. 65, 74, 140

[Marsella Gratch 2009] Stacy C Marsella and Jonathan Gratch: EMA: A process model of appraisal dynamics. *Jonathan Gratch, Stacy Marsella, Paolo Petta (eds.): Modeling The Cognitive Antecedents And Consequences Of Emotion, Cognitive Systems Research*, 10(1):70–90, 2009. 16, 65, 140

[Marsella et al. 2000] Stacy C Marsella, W Lewis Johnson, and Catherine Labore: Interactive Pedagogical Drama. In Charles Sierra, Maria Gini, and Jeffrey S Rosenschein (eds.): *Proceedings of the 4th International Conference on Autonomous Agents, June 3-7 2000, Barcelona Catalonia Spain*, pp. 301–308. ACM, 2000. 31, 65, 141

[Marsella et al. 2004] Stacy C Marsella, David V Pynadath, and Stephen J Read: PsychSim: Agent-based Modeling of Social Interactions and Influence. In Marsha Lovett, Christian D Schunn, Christian Lebiere, and Paul Munro (eds.): *Proceedings of the 6th International Conference on Cognitive Modelling ICCM 2004, Pittsburgh Pennsylvania USA, July 30 - August 1 2004*, pp. 243–248. Routledge, 2004. 53, 141

[Martin et al. 2005] Alan Martin, Gregory MPO Hare, Brian R Duffy, Bianca Schön, and John F Bradley: Maintaining the Identity of Dynamically Embodied Agents. In Themis Panayiotopoulos, Jonathan Gratch, Ruth Aylett, Daniel Ballin, Patrick Olivier, and Thomas Rist (eds.): *Intelligent Virtual Agents, 5th International Working Conference, IVA 2005, Kos, Greece, September 2005, Proceedings*, pp. 454–465. Springer Berlin Heidelberg, LNAI 3661, 2005. 53, 141

[Martin Clore 2001] Leonard L Martin and Gerald L Clore: *Theories of Mood and Cognition: A User's Guidebook*. Lawrence Erlbaum Associates Publishers, Mahwah NJ London, 2001. 42, 141

[Martinez-Miranda et al. 2008] Juan Martinez-Miranda, Bernhard Jung, Sabine Payr, and Paolo Petta: The Intermediary Agent's Brain: Supporting Learning to Collaborate at the Inter-Personal Level. In Lin Padgham, David C Parkes, Jörg Müller, and Simon Parsons (eds.): *7th International Joint Conference on Autonomous Agents and Multiagent Systems AAMAS2008, Estoril Portugal, May 12-16 2008*, pp. 1277–1280. International Foundation for Autonomous Agents and Multiagent Systems IFAAMAS, 2008. 31, 141

[Martinho Paiva 1999] Carlos Martinho and Ana Paiva: 'Underwater Love': Building Tristão and Isolda`s Personalities. In Michael J Wooldridge and Manuela Veloso (eds.): *Artificial Intelligence Today: Recent Trends and Developments*, pp. 269–296. Springer Berlin Heidelberg New York Tokyo, LNAI 1600, 1999. 44, 141

[Martinho et al. 2003] Carlos Martinho, Mário Gomes, and Ana Paiva: Synthetic Emotension: Building Believability. In Thomas Rist, Ruth Aylett, Daniel Ballin, and Jeff Rickel (eds.): *Intelligent Virtual Agents 4th International Workshop, IVA 2003, Kloster Irsee Germany,*

September 15-17 2003, Proceedings, pp. 57–61. Springer Berlin/Heidelberg/New York, LNCS 2792, 2003. 33, 141

[Mateas 2002] Michael Mateas: *Interactive Drama, Art and Artificial Intelligence*. PhD thesis, Department of Computer Science, Carnegie-Mellon University Pittsburgh PA, 2002. URL http://www.lcc.gatech.edu/~mateas/publications/CMU-CS-02-198.pdf. 66, 142

[Mateas Stern 2004] Michael Mateas and Andrew Stern: A Behavior Language: Joint Action and Behavioral Idioms. In Prendinger H. and Ishizuka M. (eds.): *Life-Like Characters - Tools, Affective Functions, and Applications*, pp. 19–38. Springer Berlin Heidelberg Cognitive Technologies Series, 2004. 77, 142

[Mateas Stern 2005] Michael Mateas and Andrew Stern: The Interactive Drama Façade. In R Michael Young and John Laird (eds.): *Proceedings of the first Artificial Intelligence and Interactive Digital Entertainment Conference*, pp. 153–154. AAAI Press Menlo Park CA USA, 2005. 66, 142

[Mateas Stern 2006] Michael Mateas and Andrew Stern: Façade: Architecture and Authorial Idioms for Believable Agents in Interactive Drama. In Jonathan Gratch, Michael Young, Ruth Aylett, Daniel Ballin, and Patrick Olivier (eds.): *Proceedings of the 6th International Conference on Intelligent Virtual Agents IVA 2006, Marina Del Rey CA USA, August 21-23 2006*, pp. 446–448. Springer LNCS 4133, 2006. 31, 142

[Minsky 2006] Marvin Minsky: *The Emotion Machine - Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster New York, 2006. 33, 69, 89, 142

[Moffat 1997] David Moffat: Personality Parameters and Programs. In Robert Trappl and Paolo Petta (eds.): *Creating Personalities for Synthetic Actors - Towards Autonomous Personality Agents*, pp. 120–165. Springer Berlin/Heidelberg/New York/Tokyo, LNAI 1195, 1997. 65, 142

[Moffat Frijda 1995] David Moffat and Nico H Frijda: Where there's a Will there's an agent. In Michael J Wooldridge and Nicholas R Jennings (eds.): *Intelligent Agents, Proceedings of the ECAI-94 Workshop on Agent Theories, Architectures, and Languages, Amsterdam The Netherlands, August 8-9 1994*, pp. 245–260. Springer LNAI 890, 1995. 65, 142

[Moshkina Arkin 2003] Lilia Moshkina and Ronald C Arkin: On TAMEing Robots. In *Proceedings of the IEEE International Conference on Sys-*

tems, Man and Cybernetics, October 5-8 2003, pp. 3949–3959. Hyatt Regency Washington D.C. USA, 2003. 27, 68, 142

[Murray 1997] Janet H Murray: *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Free Press New York, reprint by MIT Press 2000, 1997. 25, 143

[Nehaniv et al. 2002] Chrystopher L Nehaniv, Daniel Polani, Kerstin Dautenhahn, René Te Boekhorst, and Lola Cañamero: Meaningful Information Sensor Evolution and the Temporal Horizon of Embodied Organisms. In Standish R.K., Bedau M.A., and Abbass H.A. (eds.): *Artificial Life VIII, Proceedings of the Eighth International Conference on Artificial Life*, pp. 345–351. MIT Press Cambridge MA, 2002. 75, 143

[Núñez 1999] Rafael Núñez: Could the Future Taste Purple - Reclaiming Mind, Body and Cognition. *Journal Of Consciousness Studies*, 6(11-12):41–60, 1999. 53, 143

[Nielsen 1993] Jakob Nielsen: *Usability Engineering*. Academic Press, 1993. 26, 143

[Norling 2004] Emma Norling: Folk psychology for human modelling: extending the BDI paradigm. In Nick Jennings, Milind Tambe, Carles Sierra, and Liz Sonenberg (eds.): *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pp. 202–209. Vol.1, 2004. 27, 143

[Ochsner Gross 2005] Kevin N Ochsner and James J Gross: The Cognitive Control of Emotion. *Trends in Cognitive Sciences*, 9(5):242–249, 2005. 53, 61, 143

[Ortony 2003] Andrew Ortony: On Making Believable Emotional Agents Believable. In Robert Trappl, Paolo Petta, and Sabine Payr (eds.): *Emotions In Humans And Artifacts*, pp. 189–212. MIT Press Cambridge MA London UK, 2003. 9, 13, 29, 32, 44, 45, 49, 143, 152

[Ortony et al. 1988] Andrew Ortony, Gerald L Clore, and Allen Collins: *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, UK,, 1988. 13, 42, 44, 65, 143, 152

[Papez 1937] James W Papez: A Proposed Mechanism for Emotion. *Archives of Neurology and Psychiatry*, 38:725–743, 1937. 42, 143

[Parkinson 2001] Brian Parkinson: Putting Appraisal in Context. In Klaus R Scherer, Angela Schorr, and Tom Johnstone (eds.): *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 173–186. Oxford University Press, Oxford New York, 2001. 16, 35, 143

[Pervin 1994] Lawrence A Pervin: A Critical Analysis of Current Trait Theory. *Psychological Inquiry*, 5(2):103–113, 1994. 46, 144

[Petta 1999] Paolo Petta: Principled Generation of Expressive Behavior in an Interactive Exhibit. In Juan D Velásquez (ed.): *Workshop: Emotion-Based Agent Architectures (EBAA'99) at the 3rd International Conference on Autonomous Agents, Seattle WA USA, May 1 1999*, pp. 94–98, 1999. 34, 36, 110, 144

[Petta 2003] Paolo Petta: The Role of Emotions in a Tractable Architecture for Situated Cognisers. In Robert Trappl, Paolo Petta, and Sabine Payr (eds.): *Emotions In Humans And Artifacts*, pp. 251–288. MIT Press Cambridge MA London UK, 2003. 11, 15, 36, 51, 110, 144

[Petta et al. 1999] Paolo Petta, Alexander Staller, Robert Trappl, Stephan Mantler, Zsolt Szalavári, Thomas Psik, and Michael Gervautz: Towards Engaging Full-Body Interaction. In H.-J. Bullinger and P.H. Vossen (eds.): *Adjunct Conference Proceedings, HCI International '99, 8th International Conference on Human-Computer Interaction, jointly with 15th Symposium on Human Interface, Munich Germany, August 22-27 1999*, pp. 280–281. Fraunhofer IRB Verlag, 1999. 36, 144

[Pfeifer Scheier 1994] Rolf Pfeifer and Christian Scheier: From Perception to Action: The Right Direction? In Gaussier P. and Nicoud J.-D. (eds.): *From Perception to Action*, pp. 1–11. IEEE Computer Society Press Los Alamitos CA, 1994. 48, 54, 144

[Pfeifer Scheier 1999] Rolf Pfeifer and Christian Scheier: *Understanding Intelligence*. MIT Press/Bradford Books, Cambridge/London, 1999. 50, 144

[Pollack 1992] Martha E Pollack: The Uses of Plans. *Artificial Intelligence Journal*, 57(1):43–68, 1992. 57, 144

[Pollack Horty 1999] Martha E Pollack and John F Horty: There's More to Life than Making Plans - Plan Management in Dynamic, Multiagent Environments. *AI Magazine*, 20(4):71–84, 1999. 57, 144

[Premack Woodruff 1978] David Premack and Guy Woodruff: Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4):515–526, 1978. 40, 144

[Psik et al. 2003] Thomas Psik, Kresimir Matkovic, Reinhard Sainitzer, Paolo Petta, and Zsolt Szalavári: The Invisible Person Advanced Interaction Using an Embedded Interface. In Deisinger J. and Kunz A.

(eds.): *Proceedings of the 7. International Immersive Projection Technologies Workshop and the 9.Eurographics Workshop on Virtual Environments (IPT/EGVE 2003), May 22-23 2003, Zurich Switzerland*, pp. 29–37. ACM Press New York USA, 2003. 36, 144

[Rank 2004] Stefan Rank: Affective Acting: An Appraisal-Based Architecture for Agents As Actors. Master's thesis, Institute for Medical Cybernetics and Artificial Intelligence, Medical University of Vienna & Vienna University of Technology, Diplomarbeit, 2004. URL http://www.ofai.at/~stefan.rank/StefanRank-AAAThesis.pdf. 13, 36, 145

[Rank 2005] Stefan Rank: Towards Reusable Roleplayers Using an Appraisal-Based Architecture. *Applied Artificial Intelligence, special issue: Educational Agents and (e-)Learning*, 19(3-4):313–340, 2005. 13, 33, 145

[Rank Petta 2005a] Stefan Rank and Paolo Petta: Motivating Dramatic Interactions. In Lola Cañamero (ed.): *Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action*, pp. 102–107. AISB The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, University of Sussex, Falmer Brighton UK EU, 2005a. 30, 59, 145

[Rank Petta 2005b] Stefan Rank and Paolo Petta: Appraisal for a Character-Based Story World. In Themis Panayiotopoulos, Jonathan Gratch, Ruth Aylett, Daniel Ballin, Patrick Olivier, and Thomas Rist (eds.): *Intelligent Virtual Agents, 5th International Working Conference, IVA 2005, Kos, Greece, September 2005, Proceedings*, pp. 495–496. Springer Berlin Heidelberg, LNAI 3661, 2005b. 28, 145

[Rank Petta 2006] Stefan Rank and Paolo Petta: Comparability Is Key to Assess Affective Architectures. In Robert Trappl (ed.): *Cybernetics and Systems 2006 - Proceedings of the Eighteenth Meeting on Cybernetics and Systems Research, April 18-21 2006, University of Vienna*, pp. 643–648. Austrian Society for Cybernetic Studies Vienna, Volumes 1 and 2, 2006. 12, 14, 26, 145

[Rank et al. 2005] Stefan Rank, Pablo Lucas Dos Anjos, Paolo Petta, and Ruth Aylett: What is In an Affective Architecture for Situated Agents? LolaCanamero (ed.): Humaine Deliverable D7a: WP7 Workshop Proceedings, King's College London UK EU, July 4-5 2005, 2005. 17, 145

[Rank et al. 2006] Stefan Rank, Paolo Petta, and Robert Trappl: Features of Emotional Planning in Software Agents. In Della-Riccia G., Dubois D.,

Kruse R., and Lenz H.-J. (eds.): *Decision Theory and Multi-Agent Planning, ISSEK Workshop October 2004 on 'Intelligent Agents: Decision-support and planning"*, pp. 185–196. Springer Wien New York, 2006. 54, 145

[Reeves 1991] John F Reeves: *Computational morality: A process model of belief conflict and resolution for story understanding*. PhD thesis, University of California, Los Angeles, 1991. 65, 146

[Reilly Bates 1992] W Scot Reilly and Joseph Bates: Building Emotional Agents. Technical report, School of Computer Science, Carnegie Mellon University, CMU-CS-92-143, 1992. 66, 146

[Reisenzein 2001] Rainer Reisenzein: Appraisal Processes Conceptualized from a Schema-Theoretic Perspective - Contributions to a Process Analysis of Emotions. In Klaus R Scherer, Angela Schorr, and Tom Johnstone (eds.): *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 187–201. Oxford University Press, Oxford New York, 2001. 16, 61, 146

[Roseman Smith 2001] Ira J Roseman and Craig A Smith: Appraisal Theory: Overview, Assumptions, Varieties, Controversies. In Klaus R Scherer, Angela Schorr, and Tom Johnstone (eds.): *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 3–19. Oxford University Press, Oxford New York, 2001. 42, 146

[Roseman et al. 1996] Ira J Roseman, Ann A Antoniou, and Paul E Jose: Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. *Cognition & Emotion*, 10(3):241–277, 1996. 42, 146

[Rozin et al. 2000] Paul Rozin, Jonathan Haidt, and Clark R McCauley: Disgust. In Michael Lewis and Jeannette M Haviland (eds.): *Handbook Of Emotions*, chapter 40, pp. 637–653. Guilford Press New York London, 2000. 35, 98, 146

[Russell Norvig 1995] Stuart Russell and Peter Norvig: *Artificial Intelligence: A Modern Approach*. Prentice-Hall, London/New York/Englewood Cliffs, 1995. 8, 22, 23, 146

[Russell Norvig 2003] Stuart Russell and Peter Norvig: *Artificial Intelligence: A Modern Approach - Second Edition*. Pearson Education Inc., Upper Saddle River New Jersey, 2003. 22, 28, 29, 146

[Sander et al. 2003] David Sander, Jordan Grafman, and Tiziana Zalla: The Human Amygdala - An Evolved System for Relevance Detection. *Reviews In The Neurosciences*, 14(4):303–316, 2003. 29, 48, 146

[Schachter 1970] Stanley Schachter: The Assumption of Identity and Peripheralist-Centralist Controversies in Motivation and Emotion. In Magda B Arnold (ed.): *Feelings and Emotions: The Loyola Symposium*, pp. 111–121. Academic Press New York, 1970. 42, 147

[Schachter Singer 1962] Stanley Schachter and Jerome Singer: Cognitive, Social, and Physiological Determinants of Emotional States. *Psychological Review*, 69:379–399, 1962. 42, 147

[Scherer 2000] Klaus R Scherer: Psychological Models of Emotion. In Borod J. (ed.): *The neuropsychology of emotion*, chapter 6, pp. 137–166. Oxford University Press Oxford New York, 2000. 12, 147

[Scherer 2001] Klaus R Scherer: Appraisal Considered As a Process of Multilevel Sequential Checking. In Klaus R Scherer, Angela Schorr, and Tom Johnstone (eds.): *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 92–120. Oxford University Press, Oxford New York, 2001. 34, 43, 49, 88, 98, 147

[Scherer 2005] Klaus R Scherer: What Are Emotions and How Can They Be Measured? *Social Science Information*, 44(4):695–729, 2005. 16, 17, 147

[Scherer et al. 2001] Klaus R Scherer, Angela Schorr, and Tom Johnstone (eds.): *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford University Press, Oxford New York, 2001. 12, 42, 147

[Scherer et al. 2004] Klaus R Scherer, Etienne B Roesch, and Tanja Bänziger: Preliminary Plans for Exemplars Theory - Humaine Deliverable D3c, 2004. URL http://emotion-research.net/projects/humaine/deliverables/D3c.pdf. 50, 147

[Scheutz 2004] Matthias Scheutz: Useful Roles of Emotions in Artificial Agents: A Case Study from Artificial Life. In George Ferguson and Deborah McGuinness (eds.): *Proceedings of the 19th National Conference on Artificial Intelligence AAAI-04, July 25-29 2004, San Jose CA USA*, pp. 42–48. AAAI Press Menlo Park CA, 2004. 29, 147

[Schut et al. 2004] Martijn Schut, Michael Wooldridge, and Simon Parsons: The Theory and Practice of Intention Reconsideration. *Journal of Experimental and Theoretical Artificial Intelligence,*, 16(4):251–293, 2004. 56, 147

[Simpson et al. 2006] Jane Simpson, Sarah Carter, Susan H Anthony, and Paul G Overton: Is Disgust a Homogeneous Emotion? *Motivation and Emotion*, 30(1):31–41, 2006. 98, 147

Bibliography

[Singh 2005] Push Singh: *EM-ONE: An Architecture for Reflective Commonsense Thinking.* PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology Cambridge MA, 2005. URL http://web.media.mit.edu/~push/push-thesis.pdf. 69, 148

[Sloman 1999] Aaron Sloman: What Sort of Architecture Is Required for a Human-Like Agent? In Wooldridge M. and Rao A. (eds.): *Foundations of Rational Agency*, pp. 35–52. Kluwer Academic Publishers, 1999. 69, 148

[Sloman 2000] Aaron Sloman: Interacting Trajectories in Design Space and Niche Space: A Philosopher Speculates About Evolution. In Goos G., Hartmanis J., and Leeuwen J.van (eds.): *Parallel Problem Solving from Nature PPSN VI, 6th International Conference, Paris France, September 18–20 2000*, pp. 3–16. Springer LNCS 1917, 2000. 13, 148

[Sloman 2004] Aaron Sloman: What Are Emotion Theories About? In Eva Hudlicka and Lola Cañamero (eds.): *Architectures for Modeling Emotion: Cross-Disciplinary Foundations, Papers from the 2004 AAAI Spring Symposium SS-04-02, March 22-24 2004*, pp. 128–134. AAAI Press Menlo Park CA USA, 2004. 13, 148

[Sloman 2005] Aaron Sloman: AI in a New Millenium Obstacles and Opportunities. Technical report, School of Computer Science, University of Birmingham UK, 2005. URL http://www.cs.bham.ac.uk/research/cogaff/sloman-ijcai05-manifesto.pdf. 26, 148

[Sloman Scheutz 2002] Aaron Sloman and Matthias Scheutz: A Framework for Comparing Agent Architectures. In *Proceedings UKCI 02 UK Workshop on Computational Intelligence, Sept 2002*, pp. 169–176. Birmingham UK, 2002. 13, 148

[Sloman Wyatt 2006] Aaron Sloman and Jeremy Wyatt: Cosy Scenario Template. Technical report, 2006. URL http://www.cs.bham.ac.uk/research/projects/cosy/scenarios/scenario-template.txt. 26, 148

[Sloman et al. 2005] Aaron Sloman, Ron Chrisley, and Matthias Scheutz: The Architectural Basis of Affective States and Processes. In Jean-Marc Fellous and Michael A Arbib (eds.): *Who Needs Emotions? The Brain Meets the Robot*, pp. 203–244. Oxford University Press, 2005. 69, 148

[Smith Lazarus 1993] Craig A Smith and Richard S Lazarus: Appraisal components, core relational themes, and the emotions. *Appraisal and Beyond: The Issue of Cognitive Determinants of Emotion, Cognition & Emotion*, 7(3&4):233–270, 1993. 32, 49, 149

[Staller Petta 1998] Alexander Staller and Paolo Petta: Towards a Tractable Appraisal Based Architecture for Situated Cognizers. In Lola Cañamero, Chisato Numaoka, and Paolo Petta (eds.): *Workshop Notes, 5th International Conference of the Society for Adaptive Behaviour (SAB98), Zürich Switzerland, August 21*, pp. 56-61, 1998. 36, 149

[Staller Petta 2001] Alexander Staller and Paolo Petta: Introducing Emotions Into the Computational Study of Social Norms: A First Evaluation. *Journal of Artificial Societies and Social Simulation*, 4(1):2, 2001. 59, 149

[Steels 2004] Luc Steels: The Autotelic Principle. In Fumiya Iida, Rolf Pfeifer, and Luc Steels (eds.): *Embodied Artificial Intelligence: International Seminar, Dagstuhl Castle Germany, July 7-11 2003, Revised Selected Papers*, pp. 231–242. Springer Berlin Heidelberg New York, LNAI 3139, 2004. 52, 149

[Swartout et al. 2006] William Swartout, Jonathan Gratch, Randall W Hill, Eduard Hovy, Stacy Marsella, Jeff Rickel, and David Traum: Toward Virtual Humans. *Achieving Human-Level Intelligence through Integrated Systems and Research, AI Magazine*, 27(2):96–108, 2006. 65, 149

[Thalmann et al. 1997] Daniel Thalmann, Hansrudi Noser, and Zhiyong Huang: Autonomous Virtual Actors Based On Virtual Sensors. In Robert Trappl and Paolo Petta (eds.): *Creating Personalities for Synthetic Actors - Towards Autonomous Personality Agents*, pp. 25–42. Springer Berlin/Heidelberg/New York/Tokyo, LNAI 1195, 1997. 16, 149

[Thórisson 2007] Kristinn R Thórisson: Integrated AI Systems. *Minds and Machines*, 17(1):11–25, 2007. 13, 149

[Tomlinson 1999] William Micheal Tomlinson: Interactivity and Emotion Through Cinematography. Master's thesis, Synthetic Characters Group, MIT Media Laboratory, Cambridge MA, 1999. URL http://badger.www.media.mit.edu/people/badger/Thesis/Thesis.pdf. 67, 149

[Trappl Petta 1997] Robert Trappl and Paolo Petta (eds.): *Creating Personalities for Synthetic Actors - Towards Autonomous Personality Agents*.

Springer Berlin/Heidelberg/New York/Tokyo, LNAI 1195, 1997. 32, 149

[Trappl et al. 2003] Robert Trappl, Paolo Petta, and Sabine Payr (eds.): *Emotions in Humans and Artifacts*. MIT Press Cambridge MA London UK, 2003. 32, 150

[Traum et al. 2004] David Traum, Stacy Marsella, and Jonathan Gratch: Emotion and Dialogue in the MRE Virtual Humans. In Carbonell J.G. and Siekmann J. (eds.): *Affective Dialogue Systems Tutorial and Research Workshop ADS 2004, Kloster Irsee Germany, June 14-16 2004, Proceedings*, pp. 117–127. Springer LNAI 3068, 2004. 65, 150

[Tyrrell 1993] Toby Tyrrell: *Computational Mechanisms for Action Selection*. PhD thesis, University of Edinburgh UK, 1993. 55, 150

[Urban 2001] Christoph Urban: PECS: A Reference Model for Human-Like Agents. In Magnenat-Thalmann N. and Thalmann D. (eds.): *Deformable Avatars*, pp. 206–216. Kluwer Academic Publishers Boston MA, 2001. 68, 150

[Velásquez 1998] Juan D Velásquez: Modeling Emotion-Based Decision Making. In Lola Cañamero (ed.): *Emotional and Intelligent: The Tangled Knot of Cognition, AAAI FS-98-03*, pp. 164–169. AAAI Press, Menlo Park CA USA, 1998. 68, 150

[Velásquez 1999] Juan D Velásquez: An Emotion-Based Approach to Robotics. In *Intelligent Robots and Systems IROS '99, Proceedings*, pp. 235–240. IEEE/RSJ International Conference vol.1, 1999. 68, 150

[Vogler 1996] Christopher Vogler: *The Writers Journey*. Boxtree, London UK, 1996. 31, 150

[Vosinakis Panayiotopoulos 2003] Spyros Vosinakis and Themis Panayiotopoulos: Programmable Agent Perception in Intelligent Virtual Environments. In Thomas Rist, Ruth Aylett, Daniel Ballin, and Jeff Rickel (eds.): *Intelligent Virtual Agents 4th International Workshop, IVA 2003, Kloster Irsee Germany, September 15-17 2003, Proceedings*, pp. 202–206. Springer Berlin/Heidelberg/New York, LNCS 2792, 2003. 37, 150

[Wardrip-Fruin Montfort 2003] Noah Wardrip-Fruin and Nick Montfort (eds.): *The New Media Reader*. MIT Press, Cambridge London, 2003. 25, 150

[Wilson 2002] Margaret Wilson: Six Views of Embodied Cognition. *Psychonomic Bulletin & Review*, 9(4):625–636, 2002. 54, 150

[Wilson Keil 1999] Robert Andrew Wilson and Frank C Keil (eds.): *The MIT Encyclopedia of the Cognitive Sciences*. MIT Press, Cambridge/Boston/London, 1999. 55, 151

[Wooldridge Jennings 1995] Michael Wooldridge and Nicholas R Jennings: Intelligent Agents: Theory and Practice. *Knowledge Engineering Review*, 10(2):115–152, 1995. 23, 151

[Wright et al. 1995] Ian Wright, Aaron Sloman, and Luc Beaudoin: Towards a Design-Based Analysis of Emotional Episodes. *Philosophy, Psychiatry, and Psychology*, 3(2):101–126, 1995. 69, 151

[Wundt 1905] Wilhelm Wundt: *Grundzüge der Physiologischen Psychologie - (5th edition)*. Engelmann, Leipzig, 1905. 41, 151

[Zuckerman Gagne 2003] Miron Zuckerman and Marylene Gagne: The COPE revised: Proposing a 5-factor model of coping strategies. *Journal of Research in Personality*, 37(3):169–204, 2003. 65, 151

Glossary

This glossary also lists relevant abbreviations used in the text. The number at the end of each entry indicates the page on which the term is defined or first used.

ABSM Agent-Based Social Modelling   24
ActAffAct Acting Affectively Affecting Acting   36
Agent In the context of this thesis, agent usually refers to a virtual body and its control architecture. Agents are embedded in their environment, both environment and agents form the virtual world   13
AI Artificial Intelligence   11
AOSE Agent Oriented Software Engineering   23
APA Autonomous Personality Agent   25
Autonomy Ability of an agent to actively limit or allow influences of the environment on itself   29

BDI Belief-Desire-Intention   13

Concern Disposition to desire occurrence or non-occurrence of a given kind of situation [Frijda 1986]   29
CPM Component Process Model   34

ECA Embodied Conversational Agent   25
Environment In the context of this thesis, the environment is the rest of a virtual world as seen from the viewpoint of an agent   13

HCI Human-Computer Interaction   13

JAM Java Agent Model   36

NPC Non-Player Character   25

OCC Ortony-Clore-Collins, an emotion theory published in [Ortony et al. 1988] and referenced by the specialisation for believable agents published in [Ortony 2003].   13

SEC Stimulus Evaluation check   16

Situational meaning structures Situational meaning structures (or appraisal frames) are theoretical constructs [Frijda 2007] that aggregate information of individual appraisal checks   16

SOA Service Oriented Architecture   23

Tabasco Tractable Appraisal-Based Architecture for Situated Cognizers   36

Virtual World In the context of this thesis, a virtual world consists of agents and their environment   13

Index

Colophon

This thesis was typeset using the LaTeX typesetting system,
the memoir LaTeX class, and the docutils Python package.
The body text is set 3.88/4.8mm (11/13.6pt) font-size/line-height
on a 138mm measure (32.6pc) with the DejaVu fonts.
All systems used, including the fonts,
are freely available and open source.