MASTERARBEIT

Methods of Affective Artificial Intelligence and an Application for Traffic Simulation with Emotional Drivers

Ausgeführt am

Institut für Informationssysteme
Arbeitsbereich für Wissensbasierte Systeme

der

Technischen Universität Wien

unter der Anleitung von

Ao.Univ.Prof. Mag.rer.nat. Dr.techn. Hans Tompits

durch

Alexander Feder, Bakk.techn.

Kröllgasse 5/8
1150 Wien

Wien, Juli 2008

Preface

Emotions play a vital role in our lives, they are ubiquitous and matter of course to us, yet they seem so hard to conceptualize. Little value has been attributed to this phenomenon by scientific discourse, until recent findings in neuroscience revealed that emotions potentially play an essential role in cognition, specifically in regards to memory recall and decision making. Insights like these led to the emergence of the interdisciplinary field of *affective computing*, which deals with the recognition and synthesis of emotions, embracing fields such as psychology, affective neuroscience, and computer science. Besides the interest in human cognition, it evolved from a variety of very different motivations. These include the recognition of human emotions, e.g., for better human-machine interfaces; the simulation of affective characters, e.g., in computer games for entertainment or educational purpose; the synthesizing of emotions for creating intelligent agents that display realistic, life-like behaviour; or even the old dream of AI to create human-like intelligence.

This thesis finds its main motivation in the question whether computational models of emotion suffice to test theories of emotional behaviour, i.e., behaviour exhibited by autonomous entities, such as humans, simulated with virtual agents. This question has been asked numerous of times before, and a variety of simulations with affective agents are in existence, which present results that indicate that agents equipped with synthetic emotions, e.g., outperform purely reactive agents in survival tasks. These are set within highly constrained and abstracted domains where an agent has to survive in a environment, with scarce resources while additionally threatened by dangerous creatures or other agents. Unfortunately, these scenarios cannot be tested in real life. Hence, finding a more realistic domain suitable for computer simulation would be desirable, allowing for a direct comparison with real life. In this thesis, we take up this challenge and discuss a traffic simulation as such a domain. This particular application was chosen for two reasons: Firstly, driving is a highly emotional task, and undoubtedly the way a driver is steering his vehicle can be at least influenced by his emotions, if not controlled. Secondly, it is a domain almost everybody is familiar with, thus results can be intuitively understood. The stimulation developed in the course of this thesis is called `SAD`, standing for **S**imulation of **A**ffective **D**rivers. Several questions were relevant for our endeavour:

- How are emotions to be defined?
- Are primitive experiences such as pain considered to be emotions, or are these merely sensual perceptions?
- Are emotions a purely conscious phenomenon?

- Which processes in the human brain result in the elicitation of emotions, and how are these to be modeled?

We provide partial answers to these questions.

The structure of this thesis can be summarised as follows:

Chapter 1 discusses an overview of some of the motivations that resulted in the emergence of affective computing, and outlines areas of research in this field.

Chapter 2 tries to answer some of the questions asked above. A detailed definition of the word *emotion* is provided in Section 2.1, where also historical influences of our understanding of emotion are shown. It continues with an enumeration of emotional aspects as perceived and understood by contemporary science, such as neuroscience and psychology. It does not remain concealed that this is still an ongoing dispute. While Section 2.2 highlights some theoretical concepts, ideas, and hypotheses of the connection between emotions and cognition, Section 2.3 delivers tangible, pragmatic concepts that give answers to some of the question of how to implement artificial emotions. These approaches can be divided into phenomenological models and abstractions of biological systems. The latter approach leads to Section 2.4, where an emotional architecture is presented in greater detail, which is later used in `SAD`.

Chapter 3 deals with our traffic simulation implementation. Section 3.2 contains an introduction to the field of traffic simulations, and outlines methods of calculation used within that domain. Classical car-following models are discussed, and one state-of-the-art model, the intelligent driver model (IDM), is presented, which is later used for the comparison with the emotional agents. Section 3.3 introduces `SAD`, which is based on the hormonal emotion model presented in Section 2.4. In the employed model, the driving agents have the ability to "feel" happy, sad, angry, or fearful—states affecting their driving behaviour. Theoretical and practical concepts are illustrated, and implementation details are revealed and explained. Finally, Section 3.4 presents methods used for testing the emotional agents. Further, different personalities are introduced, which are then analyzed and compared to IDM drivers in regards to driving efficiency, milage, and the formation of traffic jams.

Chapter 4 concludes this thesis, and delivers a variety of insights gained from the experiments with `SAD`.

Acknowledgements

I want to thank all those wonderful people without whom this work, my studies, and my life would not have been possible in this cheerful manner. I will not enumerate the obvious ones, you know how much I love you. I want to thank you, Ludwig van Beethoven, amongst so many things, for the Piano Concerto No. 4—I could not imagine a better companion while doing my literature research. Life would be so much poorer without you, Wolfgang Amadeus Mozart; what would I have done while reading so many boring papers in the search for scientific gems, without your music? Ronan Harris, your futuristic and propulsive, yet warm-hearted music is a joy, and listening to it was the best way to keep me in a good mood for developing my traffic simulation. Further I want to thank the Austrian National Library for being such a great place, my hideaway for the moments when music alone was not enough to keep me focused. Last but not least I would like to thank gravity for being a constant factor in my life. I never failed to believe in you, and I am grateful to know that you feel the same.

Figure 1: A pleased *Cynopithecus niger*, by T. W. Wood from (Darwin 1872).

Contents

Chapter 1

Introduction

1.1 Emotional Life

Emotions play a vital role in our every day life; ranging from immediate desires, long term goals, social interaction—almost all things we deal with possess emotional quality. Surprisingly, with just a few exceptions, until very recently emotions have largely been "ignored" (Charles 2004) and "neglected" (Damasio 2002; Picard 1997) by modern science. They were merely regarded as an unnecessary relict from our evolutionary past, a mechanism located in the oldest brain regions which served survival when times were rougher, redundant compared to pure thought and reason. Stunningly counter-intuitive that the human ability to love is most commonly cited as one of these important and distinctive qualities that supposedly separates humans from animals.

Figure 1.1: 19^{th} century proposal scene.

Obviously, our approach to emotions is ambivalent and not free of contradictions. So, what exactly are emotions, and, which is of our main interest, what are emotions good for? Their influence in regard to social interaction and communication seems intuitively obvious, but what is their influence on cognitive processes? And, how are emotions related to artificial intelligence?

Humans are emotional beings, it is thus self-evident that (real or synthetic) entities that are supposed to fully interact with humans are required to "possess" emotions—or at least act and react appropriately when confronted with human emotions.

However, it is not obvious at all that our rational decision-making process should be guided by emotions—since Aristotelean philosophy, the perfect mind was envisioned to lack any emotional interference; cognition, emotion, and conation[1] were understood as three separate parts.

Modern brain imaging methods like *positron emission tomography* (PET) and *functional magnetic resonance imaging* (fMRI) revealed though, that there is no evidence that the neural circuity underlying components of affect and cognition differ (Davidson 2002; Lane and Nadel 2002).

When we allow ourselves to take a closer look at the decisions we make, we will realize that most of them are not guided by pure rationality, but are heavily influenced by our moods, personality and intuition, which could be explained due to the fact that most decisions we make appear to be quite trivial to us. Choosing what to wear, what to eat, which movie to watch, which work to ignore and which to complete, which friends to meet and what to say during small-talk, how to walk home from university: our every day life is guided by decisions we make without really thinking about them—we just *feel* what is appropriate.

What we tend to ignore in regard to these every-day problems is the extreme complexity that is involved in solving these issues, at least from a computational point of view.

1.2 Motivation

For a long time, research in artificial intelligence (AI) was driven from the hope that once machines are able to solve problems that seem hard to humans, like playing chess or solving mathematical equations, human-like intelligence will just emerge (Brooks 1991). Alas, Deep Blue did not cheer, or talk to Garri Kasparow, once it finally beat him in a game of chess in May 1997 (Hsu 1999). Deep Blue's success was rather the result of processing massive lists of possible moves, utilizing an immense amount of memory capacity and the power of hundreds of CPUs.[2]

It is highly unlikely that human-like intelligence can be achieved while ignoring fundamental properties of humans, like emotions. As stated by Minsky (1988): the question is not whether intelligent machines can have any emotion, but whether machines can be intelligent without any emotions.

Besides this rather illusionary possibility of ever creating human-like intelligence in the years to come, there are several other research fields that could, or already do, benefit from a deeper understanding of emotions and consequential application of emotional models, such as the explanation of the human mind, cognition and consciousness. Sloman (1993) argues that he does (as of 1993) not know of any other discipline that provides tools and techniques

[1]From latin: *conari* (to attempt, to strive), understood as *will*.

[2]Deep Blue was built using 30 *IBM RS/6000* workstations connected over an *RS/6000 SP* switching network. Enhanced by 480 chess chips (*VLSI*), each capable of calculating 2 to 2.5 million chess positions per seconds, the system was capable of evaluating 1 billion chess moves per second. As each chess move requires 40000 general purpose operations, the system was approximately capable of processing $40 \cdot 10^{12}$ operations per second (Hsu 1999).

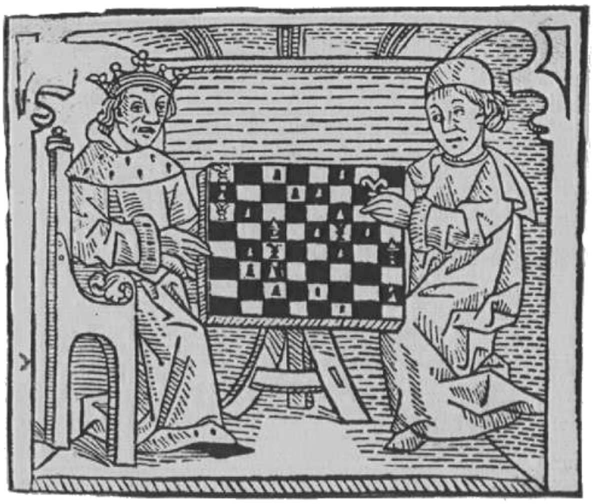

Figure 1.2: Woodcut by William Caxton (1474).

better suited than those of AI to approach the explanation of intricate information-processing capabilities that characterise the human mind.

Rolls (2000) adds that a major reason to investigate the brain mechanisms that underlie behaviour (and therefore emotions) is not only to understand how our own brains work, but also have the basis for understanding and treating medical disorders of the brain.

Insights gained in research for the understanding of the human mind can be utilized to create intelligent systems. Synthetic emotions are used in an attempt to increase the efficiency or realism of agents in dynamic environments. Agents are required to react adequately and in reasonable time to sudden events in dynamic environments like “the real world”. Purely deliberative architectures are not sufficiently equipped to fulfill these requirements for mainly two reasons: Firstly, the extremely large space of possible actions, which requires a long time to evaluate—and in a real life situation, “long” might equal to “too long” in regard to survival or success. Secondly, the outcome of such an evaluation might produce a set of possible actions which have equal priorities, thus creating a conflict that needs to be resolved (Scheutz and Logan 2001; Sloman and Croucher 1981). Emotions might prove to be the ideal way to solve both issues.

Further, in regards to multi-agent systems, emotions may also play a vital role. Additionally to the requirements described above, when multiple agents are to act and interact in an environment, they will need a certain amount of social abilities, which might involve a set of emotions. If humans are involved, the latter is most certainly a requirement.

Within virtual environments, a common technique so far to solve this requirement has been to define protocols which the agents can share. This has several disadvantages, mainly that all agents need to share the same protocol pairwise.

An agent that keeps a representation of its world (goals, achievements, relations, objects, emotions, etc.) can easily use the same architecture to keep approximated representations of other agents, thus enabling it to guess the other agents' subsequent actions (Scheutz and

Sloman 2001).

As computer systems not only communicate with each other but also with humans, another motivation for affective computing is the wish to build better human-machine communication interfaces. As Picard (1997) states, contemporary user interface design rarely take human emotions into account, not to mention recognizing them. *If a piece of technology talks to you but never listens to you, then it is likely to annoy you* (Picard et al. 2001). In this sense, a lot of effort is put into building systems that have an understanding of what a user needs while in a certain emotional state (André et al. 2004; Tao et al. 2005; Paiva et al. 2007). As Picard et al. (2001) notices, machines may never need all of the emotional skills that people need, however, she claims, that there would be evidence that machines would require at least some of these skills to *appear* intelligent in interacting with people.[3] The capability of detecting and distinguishing emotional states is fundamental to most of the issues mentioned above (Picard et al. 2001).

It is a prevailing dispute whether emotions are a hinderance or a benefit to cognition, and their effects on the behaviour of intelligent systems, e.g., humans. There are, however, a variety of arguments that speak for themselves. First of all, regardless of scientific insight, from our experience we can intuitively assert that emotions play a vital part in building a human's character and personality. The personality obviously has an impact on the way a human plans, makes decisions, and interacts with other humans, and it can be observed that the individual success varies magnificently.

This individual success can be translated into Darwinian common sense. Damasio (2002) argues that, since we possess emotions, they brought an evolutionary advantage. According to Kaplan (1991), emotions also provide a way of coding and compacting experience to enhance fast response selection. In evolutionary terms, it is better to respond immediately to the sight of a large animal, perhaps by fleeing, than to take the time to rationally consider the best course of action.

On the other hand, Brooks (1991) argue that evolution brought a lot of unnecessary things too, which is also the conclusion of Darwin (1872); the expression of emotions in adults can occur whether they are of any use or not. Oatley (2004) states that

> programs of emotional expression have been installed into our nervous systems in the course of evolution and they can operate even when, according to reason, there would be no need for them.

Besides these indications, there is also medical evidence. Damasio (1994) proposes that emotions are a necessity for rational decision making, an insight he gained while working with patients who suffer from frontal lobe damage. This, however, is not believed to be a scientific basis by some (Sloman 1999b,a), however, it seems to be the justification for most researchers in this field to further investigate the topic.

Finally, experiments with agents equipped with synthetic emotions indicate that they bring an advantage in survival tasks. E.g., experimental results indicate an increased efficiency (Scheutz and Logan 2001; Scheutz and Sloman 2001) or increased realism (Aylett et al. 2006; Aylett 2004) of agents that utilize a set of emotions.

[3] Interestingly, for a long time the predominant paradigm for research on dialog systems has been the concept of rational agents exchanging rational arguments (Streit et al. 2004).

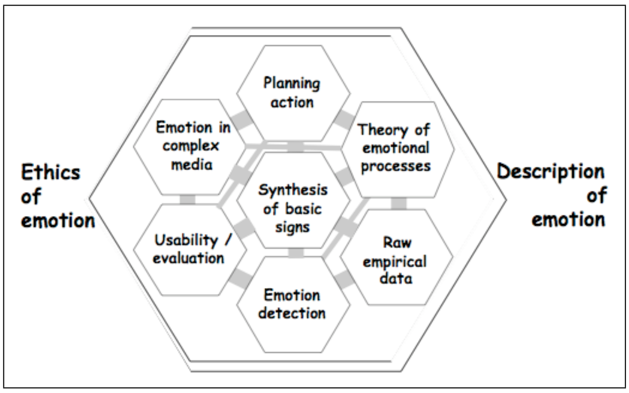

Figure 1.3: Proposed map of the sub-areas involved in emotion-oriented computing (Schröder and Cowie 2006).

1.3 Areas of Research

A wide variety of very different research fields that consider themselves affiliated with affective computing were brought into existence in recent years. They deal with all thinkable aspects of emotions, ranging from recognition of human emotional states to simulation of affective virtual agents.

Schröder and Cowie (2006) describe a dissection into thematic areas which has been agreed upon during the establishment of the HUMAINE[4] European Network of Excellence project (see Figure 1.3). These are:

- *emotion in complex media*, standing for the expression of emotions through higher, cultural channels, like music and language (Subasic and Huettner 2000; André et al. 2004; Becker et al. 2004);

- *usability, evaluation*, describing the augmentation of user interfaces with a concept of emotion, and the required evaluation of its effects (van den Noort et al. 2005);

- *planning action*, the pattern-driven selection or forecast of the possible actions an entity might perform when in a particular emotional state (Aylett et al. 2006; Rank et al. 2006; Gratch 1999; Riccia et al. 2006; Sloman 1999a; Elliott 1992);

- *emotion detection*, for the classification of human emotional states by using a variety of sensors (cameras, microphones, ECGs, EEGs, ...) and methods (pattern recognition, hidden Markov models (HMMs), ...) (Ball 2003; Scherer 2000; Kaliouby and Robinson 2005; Ekman 1993, 1999);

- *theory of emotional processes*, trying to explain human emotion emotions by building models and proposing architectures, as well as creating models—and consequently architectures—based or not based on theories of the human mind, for the use in artifacts/agents (Rolls 2003; Beaudoin and Sloman 1993; Bellman 2003; Charles 2004;

[4]HUman-MAchine Interaction Network on Emotions.

Damasio 2002; Coutinho et al. 2005; Damasio 1994; Davis 2000; Fodor 1983; Minsky 1988; Ortony et al. 1988; Simon 1967; Sloman 2004, 1999a);

- *raw empirical data*, needed to rate the performance of simulations of human emotions, e.g., to see if a model behave like its real counterpart; and
- *synthesis of basic signs*, describing the application of emotional mechanisms for the creation of agents interacting and/or acting emotionally (this is depicted in the center of Figure 1.3 as it requires all the other fields to a certain extent).

In order to bring all these topics onto a mutual basis, a clear definition of emotion seems necessary. The next chapter will discuss this issue.

Chapter 2

Theories of Affective Computing

2.1 Definition of Emotion

We shall not quarrel about the definition of "emotion" since the word is full of ambiguity and vagueness. All we are concerned with is that there are certain features found in some mental states that many people would describe as emotional.

Beaudoin and Sloman (1993)

The alleged vagueness of emotions and feelings is the most frequent excuse offered to justify the difficulty of studying these undesirable phenomena.

Damasio (2002)

2.1.1 Linguistic Definition

The Oxford American Dictionary (2007) defines our words of interest as follows:

e•mo•tion |i'mō SH ən|

noun

a natural instinctive state of mind deriving from one's circumstances, mood, or relationships with others: *she was attempting to control her emotions* | *his voice was low and shaky with emotion.*

- any of the particular feelings that characterize such a state of mind, such as joy, anger, love, hate, horror, etc.: *fear had become his dominant emotion.*
- instinctive or intuitive feeling as distinguished from reasoning or knowledge: *responses have to be based on historical insight, not simply on emotion.*

e•mo•tion•al |i'mō SH ənəl|

adjective

of or relating to a person's emotions: *children with emotional difficulties.*

- arousing or characterized by intense feeling: *an emotional speech.*
- (of a person) having feelings that are easily excited and openly displayed: *he was a strongly emotional young man.*
- based on emotion rather than reason: *sound reason, not an emotional knee-jerk response, is the best recipe for making decisions.*

Especially of interest is the last entry, which clearly shows how emotions are commonly perceived; as something contrary to reason, hindering in making good decisions, thus being something that has no place in the quest for artificial intelligence.

The old approach to the definition of intelligence used by many AI researchers, as Rodney Brooks (1991) taunts, was lead by the thought:

> ...things that are difficult for humans require intelligence, thus [...] intelligence could be defined as something necessary to solve hard problems, like solving math equations and playing chess.

It was believed (or, rather, hoped), that once machines mastered the game of chess, human-like intelligence will just emerge.

Obviously, that did not work out so well. Thus, Brooks (1991) prefers to define intelligence as *being the sort of stuff that humans do, pretty much all the time.* Alas, that definition does not deliver any novel insight, but it contains a very important clue: that *sort of stuff* humans do all the time is also being emotional, thus—from a philosophical point of view, so far without any backup from medical sciences—it might be valid to assume that emotionality is either directly interwoven with human intelligence or a parallel mechanism that is equally relevant.

Whatever it is, it is self-evident that we are affected by our emotions, which is described as such in the Oxford Dictionary by the noun *affect*:

af•fect |a'fekt; ə'fekt|

noun `Psychology`

emotion or desire, esp. as influencing behavior or action.

Thus, it is not surprising that many AI researchers refer to *affective computing* when talking about the recognition and simulation of emotions.

2.1.2 Historical Approaches

Unfortunately, knowing what these words describe in common English does not help in understanding the underlying principles. So, what are emotions?

This question has been asked throughout history, the first useful answer came from the Epicureans in Greece, around 500 BC:

> Emotions are typically caused by evaluations of events in relation to what is important to us: our goals, our concerns, our aspirations (Oatley 2004, pg. 3).

These evaluations are also called *appraisals* by psychologists. A deeper explanation of the appraisal theory will be given in Section 2.3.1.

Also the Stoics came to very useful conclusions, namely that emotions are *valuational judgments* (or "beliefs") and *resulting affective states* (Charles 2004), which clearly attributes a certain *rationality* to emotions, e.g., fear being the judgment that a specific object is potentially harmful and thus better being avoided, also accompanied by physiological reactions. The Stoics rejected the older Aristotelian idea that emotions are based on states more primitive and less rational than belief.

In the 19^{th} century, William James proposed that emotions are merely the perception of physiological changes, like an increase in heart rate, muscular tension, sweat, and sentic modulations like facial expression, vocal intonation, and posture. As such, an emotion had to be cognitively perceived before counting as such. This idea has been rejected by later researchers, for many reasons, particularly as it does not explain the mechanisms leading to these physiological changes, and does not account for unconscious emotions. Modern brain imaging methods further indicate that *the process of generating and executing an emotional response can and often does proceed outside of conscious awareness* (Lane and Nadel 2002).

However it is true that the conscious experience of emotions is a very, if not the most important aspect, especially in relation to homeostatic regulation (see Section 2.4.2) and that emotions are not always conscious before the bodily change has been perceived, e.g., muscular tension showing a yet unconscious anxiety, thus allowing for self reflection. What James described, the perception, is nowadays regarded as *feeling*, the subjective and conscious sensation a specific emotion results in.

This shows the ambiguity the term *emotion* carries, as so many phenomena are described with it. It is thus necessary to discriminate among these phenomena.

2.1.3 Emotional Aspects

According to Scherer and Peper (2001), one should distinguish between emotions, feelings, moods, attitudes, affective style, temperament.

Emotion: referring to a *relatively brief episode of coordinated brain, autonomic, and behavioural changes that facilitate a response to an external or internal event of significance for the organism* (Davidson 2002). As a sudden experience of usually short duration, *emotions give life its urgency, a condition of immediate readiness to act* (Oatley 2004). As Oatley and Jenkins (1996) describe it, *emotion is a state usually caused by an event of importance to the subject*, typically including

1. a conscious mental state with a recognizable quality of feeling and directed towards some object,
2. a bodily perturbation of some kind,
3. a gesture,
4. a readiness for certain kinds of action.

Rolls (2000) defines emotions as states elicited by rewards and punishers, which is a very suitable definition for, e.g., reinforcement learning (see Section 2.2.1). In contrast,

Positive Emotions	Negative Emotions
Amusement	Embarrassment
Satisfaction	Anger
Sensory pleasure	Disgust
Excitement	Contempt
Contentment	Sadness/distress
Pride in achievement	Shame
Fear	Relief
	Guilt

Table 2.1: Basic emotions as defined by Ekman (1999).

Scherer (2000) points out that treating emotions as steady states rather than processes is inconsistent with evidence that is accumulating with respect to the phenomena.

According to Meyer, Reisenzein, and Schützwohl (2003), the following properties suffice for a scientific definition of emotions:

- *Emotions are psychological states, states like joy, sadness, anger, fear, jealousy, pride, surprise, empathy, guilt, disgust, and similar phenomena* (Meyer et al. 2003). Not all researchers would agree with these examples, e.g., surprise is more plausibly defined as a cognitive state accompanied with an emotion like joy or anger, and empathy as a cognitive ability utilizing the emotional mechanisms.
- *Emotions are current, short-time states.* This definition discriminates emotions from moods or emotional dispositions.
- *Emotions are usually conscious states, they have a certain quality and intensity.* This definition too is disputed, for *feeling* is often understood as the conscious aspect of emotionality and according to Damasio (2002), the words *emotion* and *feeling* require a clear discrimination.
- *Emotions are typically directed towards objects.*
- *Emotions possess, besides their subjective aspects physiological and behaviouristic aspects.*

Based on cross-cultural studies, Ekman (1999) was able to outline six basic emotions, namely *joy*, *sadness*, *anger*, *surprise*, *fear*, and *disgust*, but later enhanced that list to 15 basic emotions which all humans can express regardless of heritage. These are depicted in Table 2.1.

Ekman (1999) further provides characteristics which distinguish these basic emotions from one another and from other affective phenomena

1. distinctive universal signals;
2. distinctive physiology;
3. automatic appraisal;
4. distinctive universals in antecedent events;

5. distinctive appearance developmentally;
6. presence in other primates;
7. quick onset;
8. brief duration;
9. unbidden occurrence;
10. distinctive thoughts, memories images;
11. distinctive subjective experience;

However, most researchers who refer to emotions defined by Ekman, refer to the six basic emotions.

Feelings: referring to the subjective perception of emotions, necessarily conscious.

Mood: describing a diffuse state of sometimes uncertain cause, a longer lasting process that can persist for hours or days, e.g., sadness.

Attitudes (or *sentiments* (Oatley 2004)): *enduring, affectively colored beliefs, preferences, and predispositions toward objects or persons* (Davidson 2002).[1]

Affective Style: refers to *relatively stable dispositions that bias an individual toward perceiving and responding to people and objects with a particular emotional quality, emotional dimension, or mood* (Davidson 2002).

Temperament: refers to *particular affective styles that are apparent early in life, and thus may be determined by genetic factors* (Davidson 2002).

Thus it seems obvious that—depending on the field of research—all these emotional aspects are of a certain interest and use, and that it is highly important that especially emotions and feelings are clearly distinguished from each other (Damasio 2002). However, not everyone shares that view. Often a clear definition is avoided, by using the words *emotion*, *drives*, and *feelings* as synonyms (e.g., Delgado-Mata and Aylett (2001)).

The lack of a clear definition of the emotional components leads to a lot of problems and unnecessary discussion, especially when it comes to model emotional behaviour explicitly. Sloman (1999a) asks if an exact definition of emotions is necessary at all, and argues that it depends on what is being tried to achieve. He lists following scenarios:

1. emotional behaviour in robots or software agents;

2. detecting and responding to emotions in human users of computing systems;

3. model and explain human emotions.

[1] Oatley (2004) came up with a beautiful analogy: We can think of reactive emotions and sentiments, respectively, as like the two kinds of neural signal by which our muscles work: phasic signals move a limb; the signals of muscle tone hold the limb steadily in place. Comparably, a reactive emotion causes a change whereas a sentiment maintains an emotional attitude.

The question can not be answered precisely without a clear definition of what is understood with *emotions* and the previously enumerated *emotional aspects*. If we rely on the definition by Scherer and Peper (2001) (see above) and specifically Meyer et al. (2003), we could state that Items 1 and 2 would not necessarily require

- a definition of human emotions, that is, a definition of what exactly, e.g., the emotions anger, fear, joy, and love are,
- in which way a mood differs from an attitude,
- how these things are elicited, and
- what they trigger in the human brain.

These are specific, human states or processes, and it is perfectly valid to assume that other entities, such as animals, newborn children, or an intelligent virtual agent will possess or require (or, come up with, accordingly) completely different types of emotions and fundamentally differing underlying emotional mechanisms. As such, we cannot assume that these underlying emotional mechanisms can be exhaustively described using the words *emotions*, *moods*, *attitudes* and so forth, but we can find mappings of natural language words that describe *things* which are supported by a specific architecture, e.g., which types of emotions and emotional aspects a certain architecture supports, if any at all. A primitive entity relying, e.g., on a subsumption-based architecture might make an observer believe to possess emotions or moods while only following a simple set of rules. Figure 2.1 depicts an example of a light chasing robot called a *Braitenberg vehicle* after Braitenberg (1984), which targets the strongest light source and follows it or, respectively, moves away from it. This might leave the impression that it either dislikes (vehicle a) or enjoys (vehicle b) light, even though its underlying architecture consists of only two photoreceptors, each directly connected with a motor.

This phenomenon can be explained both by emergent properties as well as the human mind, which compulsively tries to attach emotional meaning to almost anything observed. This phenomenon is called *anthropomorphism*. And even further, for a robot capable of mimicking facial expressions like smiles or frowns, it is important to keep in mind that most emotional processing involved does not happen in the robot's software but inside the brain of the human observer.

Sloman (2003) proposes an architecture to formalize all these aspects in order to precisely answer the above question. This architecture is discussed in Section 2.4.1.

The explanation of specific human emotions (Item 3 of the listed scenarios by Sloman (1999a)) will most probably require or at least highly profit from an exact definition, and leads to the issue of categorization and definition of the various emotions.

It is very difficult for researchers from different research fields (or even the same research field) to agree on definitions of a subject that has such a strong an ambiguous connection with the emotion words used in everyday language. Firstly, a fundamentally different understanding of the working of the human brain necessarily leads to a different understanding of its emotional aspects. Secondly, cultural aspects show an influence on the elicitation as well as the understanding of the various emotional aspects. As Scherer (2000) points out, e.g., a word like *anger* can denote a neurophysiological program, a subjective feeling state, an interactive stance, or a value instantiation, depending on the theoretical framework of the respective

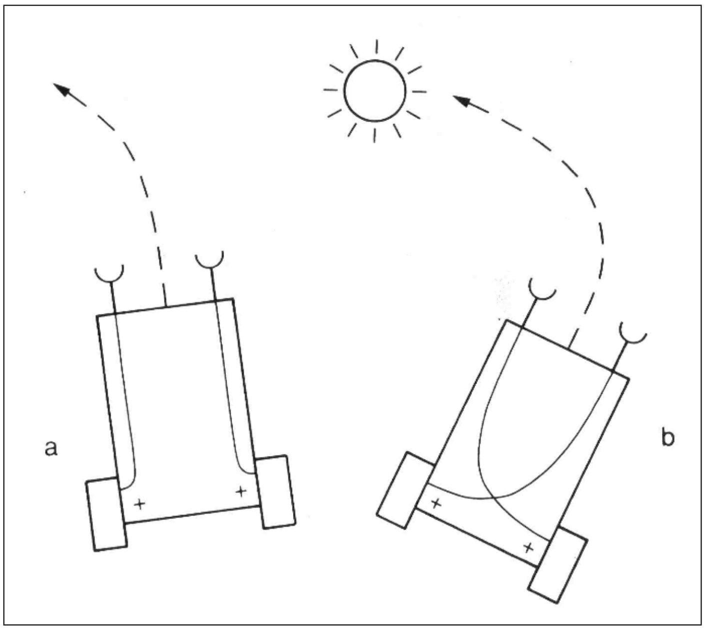

Figure 2.1: Two Braitenberg vehicles: (a) distracted and (b) attracted by light.

scholar. Further, the word *anger* is a label existing in a natural language which will vary among languages and cultures. Thus, similarly to Ortony et al. (1988) and in accordance to the proposed formalization of Sloman (2003), Scherer (2000) claims that a true science needs to keep a distance from the apparently natural categories of affective states as provided by the respective languages.

Nevertheless, e.g., an emotion classifier would be highly unsatisfying if lacking the ability to produce results that are understandable by humans, that is, for instance, mapping recognized states to one or more natural language words or presenting results as vectors in an emotional dimensional representation (see Section 2.3.3 for a discussion on dimensional models of emotion).

If, and to what extent, an *emotion* has to be defined, depends, as Sloman (2004) pointed out, on what is being tried to achieve. Thus, a broad range of very differing definitions are in existence.

Some researchers, like Ventura and Pinto-Ferreira (1999), avoid any definition, e.g., by stating that their model *is oriented towards the emergence of artificial emotional behaviour from a particular architecture,* without an a priori definition of human-like emotions, which is argued by the fact that the purpose of their research is *not explaining human emotions, but rather creating a theoretical framework uncompromised with human emotions.*

Similarly opposed to the assignment of specific symbolic labels are Henninger et al. (2002) who opt to use continuously valued variables ranging between 0 and 1, namely "pleasure/pain", "clarity/confusion", and arousal. Proposals like these fall under the topic of *dimensional models* which are outlined in Section 2.3.3.

According to Ortony et al. (1988, pg. 2), *a theory about emotions has to be a theory about the kinds of things to which emotion words refer, not about the words themselves.* Nevertheless, lots

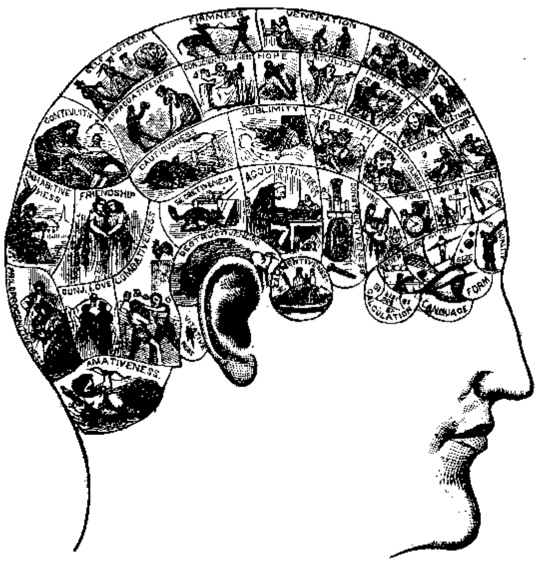

Figure 2.2: 19^{th} century phrenology chart, from Fowlers & Wells (1891).

of especially empirical research is based on emotion words, e.g., in setups where test persons are asked to rate and classify their emotional experiences, as done by Picard et al. (2001). Experiments using dimensional methods, i.e., where a test subject is asked to mark a point (emotional vector) in a two dimensional graph in correlation to his emotional experience prove to be more efficient and reliable.

Sloman (2004) warns that the word "emotion" should not be too broadly construed:

> Sometimes over-generalising the notion of "emotion" is related to a desire to argue that emotions are important in ways not previously acknowledged, e.g., that they are a prerequisite for intelligence. This can be wishful thinking or a trivial truism. If "emotion" is construed so broadly that it covers all goals and preferences, the claim that emotions are needed for intelligence is vacuous. On the other hand if it refers more narrowly to the sorts of processes in which one subsystem interferes with or disrupts the normal functioning of another, as happens in many of the states in which people are described as being 'emotional' then it is false that emotions are required for intelligence: on the contrary, emotions of that sort can get in the way of sensible decisions and actions.

2.2 Emotions and Cognition

Many researchers in the affective computing field were motivated by Antonio Damasio's work with patients who suffer from frontal lobe damage. These patients lack certain abilities, most notably they cannot experience emotions and show impairments in judgment and decision-making in real-life, while maintaining normal intellect. Next to severe difficulties in social situations, like choosing friends, partners and activities, the decisions made are no longer advantageous. The patients often decide against their best interest (Bechara and Damasio 2005).

Damasio (1994) describes a patient, *Elliot*, who “suffers” from this very condition due to the removal of a brain tumor in that specific region. Elliot scores above average on intelligence tests, but fails to solve basic every-day tasks. When asked to meet a person the next day at a specified time and place, he searches an endless space of possible ways to go there, considering absurd conditions that might arise, without the motivation to stop searching, which allows the interpretation that he is unable to feeling emotionally inclined towards a specific solution he already came up with. Elliot himself describes his surprise, as he realizes that he can come up with so many possibly correct answers—e.g., how to act in certain social situations—but yet, being unable to decide which would be most appropriate to pursue.

This lead to the conclusion that emotions are a necessity for the human decision making process. Sloman (2004) argues that while the conclusion might prove correct, that emotions play a vital role in the decision making process, the argumentation is not necessarily scientific or true:

> ...from the premiss: *Damage to frontal lobes impairs both intelligence and emotional capabilities* to the conclusion *Emotions are required for intelligence*. A moment's thought should show that two capabilities could presuppose some common mechanism without either capability being required for the other.

The supposed necessity of emotions in cognition could also be explained in such a way that it reduces the computational complexity of any given task. According to Scheutz (2004), emotions can influence, bias, and direct cognitive processes. He states that negative affect could influence humans to apply local, bottom-up processing, positive affect can bias towards global, top-down approaches. A purely rational decision would need to take all possible factors into account, as any yet so tiny factor can have a minimal influence on the outcome of an observed situation. Thus, a pruning to the space of possible solutions naturally needs to be applied.

Aylett et al. (2006) draws similar conclusions, particularly attributing the tasks of setting the attentional focus, goal selection, and plan evaluation to emotion, which she sees as one aspect of an integrated system.

However, the justifications brought in Chapter 1 should suffice to further persuade the idea. Through introspection we can tell that indeed we are often guided by emotions, ignoring or not considering possible solutions or steps which we feel negatively inclined to, and are strongly biased towards solutions that have a positive emotional value.

2.2.1 Reinforcement

Rolls (2003) defines emotions as *states elicited by reinforcers*. Reinforcers can be positive or negative, which is simply described as *reward* and *punishment* by Rolls, which stands in contrast to the definition of reinforcement in psychology.

Within psychology, a positive reinforcer is understood as a reward given *after* a specific behaviour, e.g., sugar for a diligent horse (representing an environmental reward, a learned reinforcer), or an orgasm after the attempt of reproduction (being an internal bodily reward, an unlearned reinforcer). A positive reinforcer increases the probability that an animal is inclined to repeat that specific behaviour—in that specific situation—again. A negative reinforcer however is understood as the increase in likelihood of a specific behaviour resulting out of the *removal* of an unpleasant stimulus, e.g., a horse moving without its rider hitting it.

In simpler terms, reinforcement can be understood as the increase in likelihood of behaviour through the addition of pleasant stimuli or removal of unpleasant stimuli whereas punishment is understood as the decrease in likelihood of behaviour through the removal of pleasant stimuli or addition of unpleasant stimuli.

Rolls' hypothesis of the brain being designed around reward and punishment evaluation systems (Rolls 2003) is based on the assumption that this is a feasible approach of avoiding complexity; that it is much easier for a system to define goals rather than providing predefined behavioural strategies in order to adapt to a dynamic environment.

This concept of reinforcement can be applied to learning. Sutton and Barto (1998) describe reinforcement learning as learning of what to do, i.e., how to map situations to actions in order to maximize a numerical reward signal. It is a machine-learning technique by which the learner has to chose the necessary actions that result in the highest reward on its own, usually using an trial-and-error approach.

In contrast to *supervised learning*, e.g., neural networks which are trained using samples and an associated desired result, reinforcement learning yields at learning through interaction with an environment; in other words, pain is known to be an exceptionally good teacher.

However, reinforcement learning also introduces the concept of a future reward, and thus making the concept particularly useful for long-term tasks. It can be described as a Markov decision process, which can be described by a tuple $(S, A, P.(\cdot,\cdot), R(\cdot,\cdot))$ where, according to (Feinberg and Shwartz 2002),

- S denotes a set of world states,
- A is an action space, where $A(x)$ denotes a set of possible actions at states $x \in S$,
- $P_a(s, s') = P(s_{t+1} = s' | s_t = s, a_t = a)$ denotes a transition probability, i.e., that an action a at time t, denoted as a_t, leads from state s at time t, denoted as s_t, to state s' at time $t+1$, denoted as s_{t+1}, and
- $R(s, a)$ denotes a one-step reward function, using action a in state s.

Following Feinberg and Shwartz (2002), $v_N(x, \pi, \beta, f)$ is the expected total reward over the first N steps, given by

$$v_N(s, \pi, \beta, f) = \sum_{n=0}^{N-1} \beta^n R(s, a) + \beta^N f(s_N), \tag{2.1}$$

where

- the policy π is a sequence of transition probabilities, $\pi : S \to A$;
- $\beta \in [0, 1]$ is a discount factor,
- $f(\cdot)$ is some terminal reward function (for infinite-horizon problems, $N = \infty$ and the expected total reward does not depend on f).

This reward, v_N, is then to be maximized. In the case of reinforcement learning, an agent achieves this by finding a suitable policy π that maximizes his reward.

Ahn and Picard (2005) show that this concept can be used within the domain of emotions, and propose a model of affective anticipatory reward, where $R(s, a)$ can be interpreted as *feeling good* and *feeling bad*.

2.2.2 The Somatic Markers Hypothesis

The above mentioned observations that Damasio (1994) made while working with patients with lesions of the ventromedial prefrontal cortex, led to the formulation of the *somatic markers hypothesis*.[2] According to Bechara and Damasio (2005), the hypothesis specifies a number of structures and operations required for the normal operation of decision-making, an emotional mechanism that rapidly signals the prospective consequences of an action, and accordingly assists in the selection of an advantageous response option. Without this signal, the brain needs to utilize reasoning to make decisions, which, due to so many possible conflicting options that need to be taken into account, impairs the speed of deliberation—Damasio (1994) describes patients that need a very long time to decide between two brands of cereal, because of endless reasoned analyses of the pros and cons of each brand—as well as the adequacy of the choice, i.e., patients may choose disadvantageously (Bechara and Damasio 2005).

These somatic states can be elicited by so called *primary inducers* and *secondary inducers*. Primary inducers are innate or learned stimuli that cause pleasureable or aversive states (Bechara and Damasio 2005) (similar to the reinforcers described by Rolls (2000, 2003)), such as, e.g., fear of snakes or joy of having won money while gambling on the stock market. Secondary inducers are entities generated by the recall of a personal or hypothetical emotional event, e.g., memories of encounters with a snake.

Somatic states triggered by secondary inducers lead to a "what it feels like" to be in a given situation. Further, acting on a conscious or non-conscious level, they can have an influence on activity in various cortical regions—most noteworthy, regions involved in *working memory*.

> The influence of somatic state signals on the contents displayed in working memory helps endorse or reject "objects" and "response options" brought to mind during the pondering of a decision, i.e., they help *bias* the options and plans for action (Bechara and Damasio 2005).

2.2.3 Emotions and Memory

According to Davidson (2002), memories for emotional events can show more robustness compared to memories for nonemotional events. Cahill and McGaugh (1998) have shown that endogenous stress hormones affect the amygdala to modulate memory storage in other parts of the brain in response to emotionally arousing events. Similarly, Cochran et al. (2006) state that emotional arousal has an impact on memory performance. Their experiments have demonstrated that whereas a too low arousal leads to a lethargic condition, too much arousal may lead to a hyperactive condition that can inhibit concentration. However, high emotional arousal

[2] *Somatic* refers to the Greek word "soma", i.e., body. The term *somatic* was chosen because of the different understanding of the word *emotion* amongst the different fields of (scientific) expertise. *Somatic* refers to the collection of body-related responses that hallmark an emotion (Bechara and Damasio 2005).

Figure 2.3: Drawing by W. H. Drake (1895).

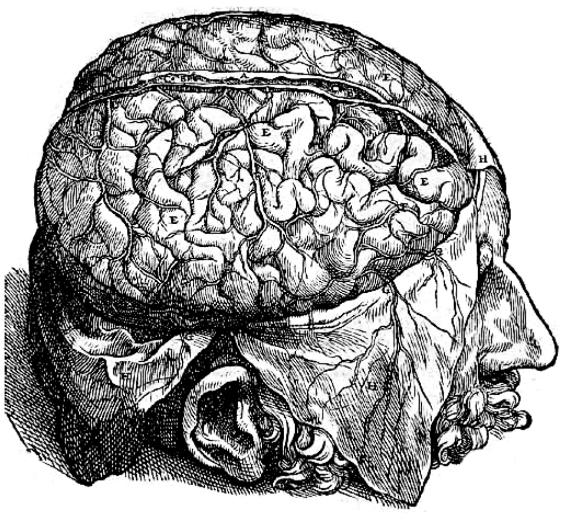

Figure 2.4: A human head with an exposed brain. Picture by Andreas Vesalius (1543).

during encoding can facilitate long term retention, but it is also associated with an inability to retrieve information for a short period of time (< 30min) following the original encoding. Levine and Pizarro (2004) bewail that much of the recent research in emotion and memory has been limited by its treatment of emotion as merely arousal. They present evidence that people process, encode, and retrieve information differently depending upon whether they are feeling happy, fearful, angry, or sad. In an empirical study, Kensinger (2007) found out that negative emotion enhances memory accuracy more than positive emotion does, thus concludes that the valenceof an event seems to be a critical determinant. Within an evolutionary framework this finds its natural explanation in the primary function of emotion to guide action; that potentially threatening situations gain higher attention and specific information is stored more accurately for better future predictive use.

Also Rolls (2003) suggests that emotion facilitates the storage of memories and attributes these mechanisms to relatively nonspecific projecting systems to the cerebral cortex and hippocampus, including the cholinergic pathways in the basal forbrain and medial spetum, and the ascending noradrenergic pathways. He states three ways in which this could occur: Firstly, episodic memory is facilitated by emotional states. This is argued by the assumption that learned behaviour of a specific situation with its emotional context will be useful for similar situations in the future, that is, when a similar emotional state is experienced. Secondly, emotions may affect the storage of memories. Similar to the somatic markers hypothesis which deals with emotions in the decision-making process, it is suggested that the current emotional state is stored with episodic memories. Which memories are recalled is then affected by the emotional state. Thirdly, emotions may guide the cerebral cortex—known as the *gray matter*—which is responsible for the processes of memory, attention, perception, thought, language, and consciousness.

Memorization of stimuli or events associated with appropriate behaviour can be explained by reinforcement learning, or stimulus-reinforcement association learning (Rolls 2000). As explained in Section 2.2.1, stimuli or events act as rewards or punishers (reinforcers), that is, their occurrence, termination, or omission as a consequence of a certain action, alter the probability of the future emission of that action—for example, fear is an emotional state, e.g., elicited by a loud sound, previously associated with a painful stimulus. Rolls (1999) argues that the adaptive value of such emotional states is to allow a simple interface between sensory inputs and action systems and further adds flexibility in contrast to stimulus response or habit—that is, the association of a specific response to a specific stimulus.

Chassy and Gobet (2005) propose a computational model linking emotion and cognition, with an emphasis on learning and memory, which is—according to the authors—based on recent findings in neuroscience. The model is built around a type of neural network called *cell assemblies* (CA) (Figure 2.5), a cognitive central processing unit (CPU) and an emotional CPU (Figure 2.6). The CAs, networks concerned with different type of information such as spatial or auditory stimuli, are distributed in distinct brain areas and are closely interconnected in such a way that an activation of one CA's neuron results in the possible activation of other neurons in that collective. Thus, a subpart of the assembly can activate the entire assembly, which makes the recognition of an entire object possible, even when partial aspects of that object are missing. Chassy and Gobet (2005) state that this activation defines a neural coordinate which serves as a pointer for *short-term memory* (STM) storage. The cognitive CPU's role is to compute a representation of the external environment by connecting all active CAs, which leads to the emergence of a dynamic neural network (DNN) that represents that external environment. These DNNs are the result of the computation's two main consequences: the activation of new pathways between CAs, and the modification of binding strengths within CAs.

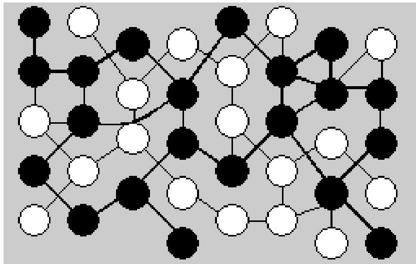

Figure 2.5: A cell assembly. Neurons connected by bold links form a CA. Picture by Chassy and Gobet (2005)).

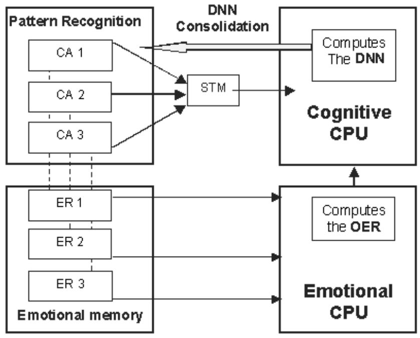

Figure 2.6: The information flow. Picture by Chassy and Gobet (2005)).

Following the computation of a DNN, the CPU supervises its consolidation into durable memory traces, a process inspired by neuroscientific evidence that the medial temporal lobe mediates information storage in the sensory cortices by generating structural changes via mechanisms like long-term potentiation (Kandel et al. 2000), resulting in *long-term memory* (LTM).

According to Chassy and Gobet (2005), this information lacks any utility value, that is, the relation of an object in the environment to a given task. This association they see as one of the roles of emotions in the sense of Rolls (2000, 2003)—emotions being states elicited by rewards and punishers. There are two processes that are performed in parallel, which are depicted in Figure 2.6. Firstly, the previously outlined cognitive process, and secondly the emotional process. The latter is further divided into two steps: the activation of a CA induces the retrieval of an associated *emotional response* (ER1, ER2, ER3), and the calculation of an *overall emotional response* (OER) using all retrieved ERs by the CPU. This OER is the utility value of the environment represented by a specific DNN.

2.3 Theories and Models of Emotion

The computational simulation of any natural phenomenon requires first of all a theory and resulting thereof an abstract model of its functioning. There are countless motivations that lead to the need for simulations; e.g., for entertainment, for predictive or explanatory purposes, or for testing a theory.

If a model is too shallow, the results might be either unsatisfactory or simply wrong. If the model is too fine grained, there might be no results at all because of computational restrictions in the sense of computational complexity, or there might be the need to again build more abstract models of that model in order to understand it. Whereas a deeper model of emotion and cognition is most probably necessary for the understanding of the human brain, a shallow model might prove sufficient for entertainment applications or interface design; however, when empathy is required, that assumption might prove wrong.

Sloman (2003, pg. 46) brings up the question whether it is possible at all to build a theory of the human information processing architecture, considering the possibility that the architecture is a completely unstructured mess. This concern is not new, and was, e.g., articulated by Fodor (1983), but has not stayed unchallenged. It is, however, beyond controversy that the human mind is a perfect example of a natural system where it is hard to find suitable models. As Minsky (2006) states, it has been tried for a long time by psychologists to approach the human mind in the same way as physicists approach physical things, trying to find a compact set of laws to explain what happens inside our brains. Minsky believes that such an approach is impossible to succeed, as he perceives the brain as something consisting of hundreds of interacting parts, specialized in very specific tasks, an idea originally presented in his book *Society of Mind* by Minsky (1988). Accordingly, there could not be a simple set of laws.

This notion is not particularly helpful for the practical goal of simulating emotions, after all, simple rules are the most obvious approach in constructing an algorithmic model.

One method to overcome this problem is by taking a pragmatic stance and only caring about the phenomenal component of the system, that is, applying a so called *black-box* approach by creating a model which—given a predefined set of inputs—behaves similar to the original, but is not deliberatively constructed similarly. These models respond to a purely engineering motivation (Cañamero 2003) and do not suffice as an explanation of the human or animal brain. Minsky (2006, pg. 20) describes such an approach using a *rule-based reaction-machine* where for any thinkable condition recognized by sensors, there is a mapping for appropriate actions, using *"If→Do"* rules. Of course, these machines are highly constrained and inflexible. However, Kaiser and Wehrle (2006) argue that these models are very useful for practical decision making and for providing a sound grounding for theoretical and empirical study.

Another pragmatic approach are *design-based models* (Cañamero 2003) or *process models* (Wehrle and Scherer 1995), which can be inspired by biological systems, but do not necessarily need to be. The goal is to come up with mechanisms that are grounded on a hypothesis of the structure of natural mechanisms, in order to reproduce their specific behaviour. Without relying on a set of inflexible oversimplified rules, these models are supposed to be more suitable for dynamic and unpredictable environments. As Wehrle (2001) points out, any process model will at a certain level reveal black-box properties—e.g., a computational model of neural structures does usually not explicitly model the chemical exchange in the axons. In Table 2.2, the differences between black-box and process models are outlined.

Intermediate between these two definitions lies what Sloman (2003) describes as *design stance*, *physical stance*, and *intentional stance*. The first describes an organism (or artifact) based on its information processing architecture. The physical stance describes the physical components and organization of the system. Finally, the intentional stance ignores internal mechanisms and just assumes that it is useful for a predictive purpose in regard to the behaviour of a system.

In what follows, the most important theories which are used within the field of affective computing will be mentioned and briefly discussed.

2.3.1 Cognitive Appraisal Theories

Emotions *are typically caused by evaluations of events in relation to what is important to us: our goals, our concerns, our aspirations* (Oatley 2004, pg. 3), and only the subjective interpretation of these events by an individual give them their significance. This idea led to a group of theories called *cognitive appraisal theories of emotion*, which are undoubtedly the most influential contemporary theories of emotion (Scherer 2001).

An *appraisal* is a not necessarily conscious evaluation of stimuli, events, or situations and

Approach	Black-box modeling	Process modeling
Purpose	Evaluate theoretical predictions Explore theory Help formalize theory Discover missing postulates and test internal consistency Empirical studies	Same as on the left and: Formalize mechanism Formalize dynamics
Techniques	Rule-based systems Decision tree Neural networks Fuzzy logic Stochastic components Etc.	Neural networks Dynamical system theory Time series, etc.
Constraints on techniques	None	Biological and psychological plausibility
Criterion	Quality of input and output mapping Performance, economy of chosen model	Quality of input and output mapping Dynamics Plausibility of chosen model
Description	Input/output relation	Mechanism, and intermediate states
Explanation	What? Why?	What? Why? How?
Prediction	Final outcome	Final outcome, dynamics, and state transitions

Table 2.2: Black-box and process modeling, by Wehrle (2001).

leads to *the elicitation of an emotion process and its differentiation into different types of emotional qualities such as fear, anger, joy, etc.* (Wehrle 2001).

If the evaluation, or appraisal, of the event was rated as significant, an adaptive action or internal adjustment is considered, a process called *coping*. *Appraisal* and *coping* are the two basic processes that form the appraisal theory, and build on a broad set of cognitive mechanisms like planning, explanation, perception, memory, and linguistic processes (Gratch and Marsella 2005). Ortony (2003) makes use of a finer grained concept, that of *response tendencies* of which coping is one of three cognitive reactions to emotional situations.

The most influential approaches to appraisal theory are the ones by Smith and Lazarus (1990), Roseman, Antoniou, and Jose (1996) (which is briefly discussed below) and Ortony, Clore, and Collins (1988) (the latter will be outlined in greater detail later on).

2.3.1.1 Roseman's Approach

Roseman (1979) defines multiple appraisal dimensions which determine whether the elicitation of a specific emotion occurs. Since its initial introduction, Roseman's theory was refined multiple times; initially, five dimensions were thought as sufficient, in the latest revision (Roseman et al. 1996) nine dimensions are utilized, each can be represented as a discrete numerical scale from 1 to 9.[3] These dimensions can be interpreted, partly inspired by van Dijk and Zeelenberg (2002), as follows:

- *Unexpectedness*: was the event expected (1) or not (9).
- *Motivational state*: is an event relevant to appetitive or aversive motives[4]—the agent wants to keep something pleasurable (1) or get rid of something painful (9).
- *Situational State*: is the state consistent with the agent's appetitive or aversive motives or not—the event improved things (1) or the event made things worse (9).
- *Probability*: are the consequences of the event certain (1) or uncertain (9).
- *Control potential*: can the agent control or influence the event—something could be done (1), or not (9).
- *Legitimacy*: is the agent thinking of himself being morally right (1) or morally wrong (9) in this event.
- *Own power*: does the agent feel powerful (9) or powerless (1).
- *Problem source*: did the event reveal the basic nature of someone or something (9) or not (1).
- *Agency:* who or what caused that event. This dimension can further be divided into *self*, *other* (agents), and *circumstances*, each can be attributed with causing (9) or not causing (1) an event. Appraisal calculations for each agency have to be performed separately (van Dijk and Zeelenberg 2002). According to Picard (1997), this is a shortcoming, as complex

[3] This numerical scale reveals the theory's intended purpose, that of being used for the analysis of surveys.
[4] Similar to the reinforcers described by Rolls (2000, 2003), see Section 2.2.1.

situations where multiple appraisals may have to be made could not be modeled, e.g., when an agent is treated unfair, but himself is guilty to have produced that certain event.

Picard (1997) suggests to combine the dimension of *motivational state* and *situational state*, and the dimensions *legitimacy*, *own power*, and *problem source*.

2.3.1.2 OCC

Ortony et al. (1988) were among the first researchers to identify the necessity of artificial intelligence to deal with emotions, though not to make machines utilize them but rather to be enabled to reason about them. Thus they proposed a model, commonly referred to simply as the *OCC model*, which was explicitly designed for the use in computer simulations, though not for synthesizing emotions. However, it has been demonstrated that the model is suitable for this purpse, e.g., by Elliott (1992) presenting the *Affective Reasoner*.

The model is based on the assumption that the distinct emotion types cannot be arranged informatively into any single space of reasonably low dimensionality, but rather that emotions come in groups, each structured in such a way that it provides a specification of the emotions' *eliciting conditions* (Ortony et al. 1988, pg. 15). Emotions within a group are related insofar as they share the same eliciting conditions, but differ in intensity and behavioural components.

Emotions are understood as valenced reactions; depending on the intensity of the affective reactions, an emotion will, or will not be experienced as such. There are three kinds of things to which an agent can have valenced reactions to. These are the three main aspects in the world recognized by the model: *events*, *agents*, and *objects*—respectively consequences of events (with the valenced reaction of being pleased/displeased), actions of agents (approval/disapproval) and aspects of objects (like/dislike).

The model distinguishes between 22 emotional categories, which are depicted in Figure 2.7. It should be noted that the words used are not to be regarded as exact definitions: *a theory about emotions has to be a theory about the kinds of things to which emotion words refer, not about the words themselves* (Ortony et al. 1988, pg. 2). Thus, they only provide an orientation in regard to the overall meaning of each category, and according to Ortony et al. (1988), the words used are the best representation they could find.

Ortony et al. (1988, pg. 181) propose a (pseudo) formalism that is derived from their emotion characterizations, and demonstrate their ideas by the example of joy.

Let $P_j : \mathbb{R}^3 \rightarrow \mathbb{R}$ be the *joy potential*, that is the potential for generating a state of joy, then

$$P_j(p, e, t) = \begin{cases} f_j(D(p, e, t), I_g(p, e, t)) & \text{if } D(p, e, t) > 0, \\ P_j(p, e, t-1) & \text{otherwise,} \end{cases} \tag{2.2}$$

where $D(p, e, t) \in \mathbb{R}$ denotes the desirability that a person p assigns to some perceived event e, at time t. It defines whether an event is judged as being beneficial to the agent ($D(p, e, t) > 0$) or possibly harmful ($D(p, e, t) < 0$). $I_g(p, e, t) \in \mathbb{R}$ represents the contribution to intensity made by global variables in response to an event. Ortony et al. (1988, pg. 183) do not elaborate on how these global variables (e.g., sense of reality, proximity, and unexpectedness) might be represented. $f_j : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a function, associated with joy, which is supposed to represent the combined effects of $D(p, e, t)$ and $I_g(p, e, t)$ numerically; however, again Ortony et al. (1988) do not elaborate on that.

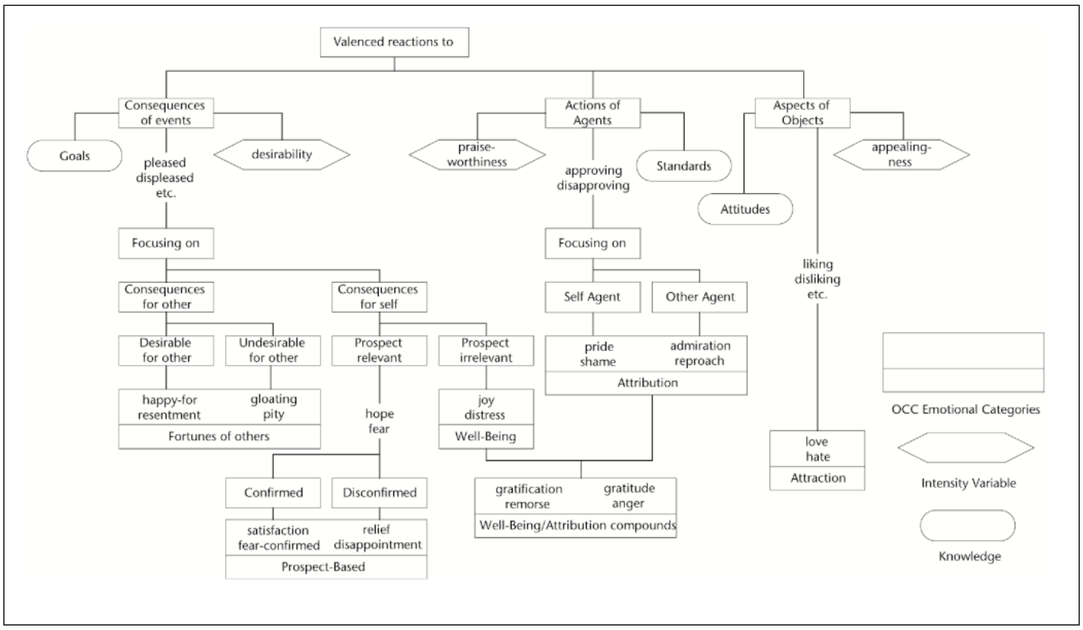

Figure 2.7: The original OCC model, image from Bartneck (2002).

Equation 2.2 shows one of many possible ways of formalizing the elicitation of an emotion-potential. In order to further determine whether any emotion ensues, and if so, how intensively, an *emotion intensity* has to be derived from the emotion potential.

Let $I_j(p, e, t) \in \mathbb{R}$ be the intensity of the emotion *joy* of person p for perceived event e at time t, and let $T_j(p, t) \in \mathbb{R}$ be some threshold value, then

$$I_j(p, e, t) = \begin{cases} P_j(p, e, t) - T_j(p, t) & \text{if } P_j(p, e, t) > T_j(p, t), \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Ortony et al. (1988, pg. 185) suggest that the rules can be augmented in such a way that depending on the calculated intensity, certain appropriate English[5] language tokens can be selected, e.g., *pleased* or *glad* for relatively low values of I_j, or *ecstatic* or *euphoric* for high values, respectively.

Introducing the concept of *prospect*, Ortony et al. (1988) claim that emotions such as fear can be modeled. Let $E(p, e, t) \in \mathbb{R}$ be the prospect (or expectancy) an agent p entertains of an event e at time t, and let $L(p, e, t) \in [0, 1]$ be the likelihood of that event to happen, then the *fear potential* $P_f : \mathbb{R}^3 \to \mathbb{R}$ is defined as

$$P_f(p, e, t) = \begin{cases} f_f(D(p, e, t), L(p, e, t), I_g(p, e, t)) & \text{if } E(p, e, t) < 0 \text{ and } D(p, e, t) < 0, \\ P_f(p, e, t-1) & \text{otherwise,} \end{cases} \tag{2.4}$$

where $f_f : \mathbb{R}^2 \to \mathbb{R}$ is some function, associated with fear, which is supposed to combine the values of desirability, likelihood, and various global intensity variables into one single value.

[5] or Russian, Cantonese, Swahili, etc...

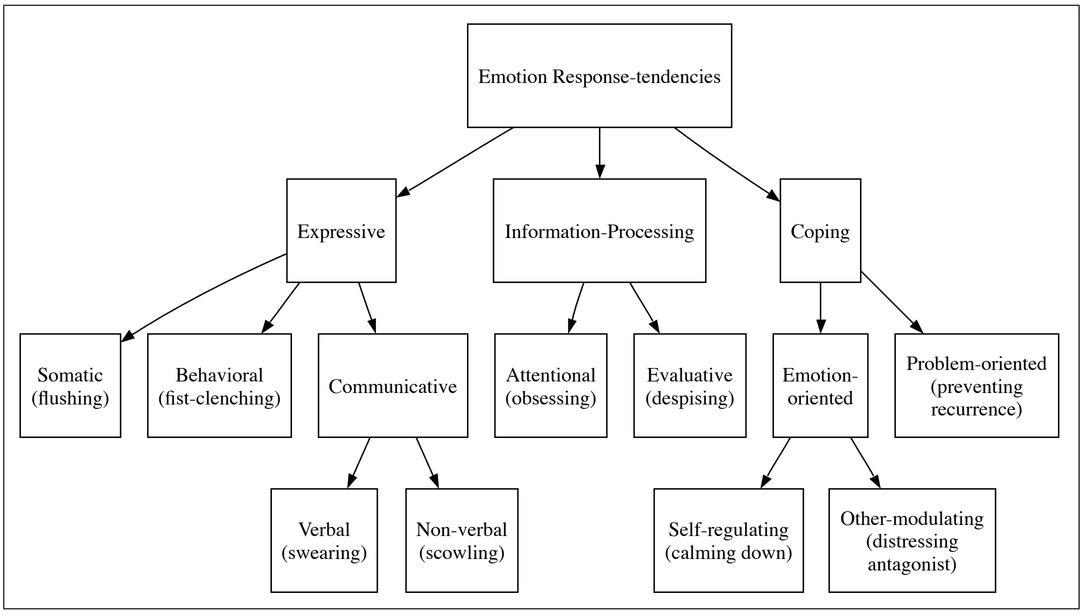

Figure 2.8: Response tendencies of Ortony (2003).

As above, to calculate whether an emotion is ensued depends on its intensity. Let $I_f(p,e,t) \in \mathbb{R}$ be the intensity of fear of person p for perceived event e at time t, $T_f(p,t) \in \mathbb{R}$ again some threshold, then

$$I_f(p,e,t) = \begin{cases} P_f(p,e,t) - T_f(p,t) & \text{if } P_f(p,e,t) > T_f(p,t), \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

It can be assumed that Ortony et al. (1988) did not intent these formalizations to serve as a detailed template for an actual implementation. Low-level details which would be required for a direct implementation, such as the values of the thresholds or the design of the combination functions f_j and f_f are not specified.

Ortony (2003) states that for the purpose of building believable artifacts/agents, a consolidation of some of the described emotion categories would be necessary. Thus Ortony (2003, pg. 196) presented some refinements to the OCC model. Among them is the concept of *emotion response-tendencies* (see Figure 2.8), that is, tendencies to influence certain behaviours which are triggered by the appraisal of an event, i.e., involuntary *expressive* manifestations (e.g., physiological reactions such as flushing), changes regarding *information processing*, or *coping* responses such as goal-oriented, planned actions. Ortony (2003, pg. 204) claims that *all emotion responses have these three kinds of tendencies associated with them*, which does not necessarily mean that an affect has to be observed—after all, *they are* tendencies *to behave in a certain way, not actual* behaviours.

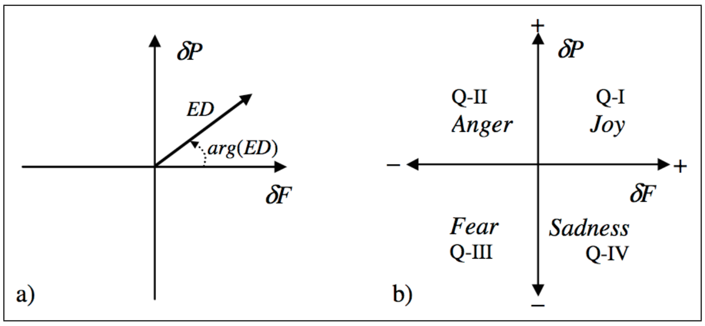

Figure 2.9: Flow model of emotion by Morgado and Gaspar (2005).

2.3.2 Psi Theory

The Psi theory is part of a family of models, dealing with an autonomous agent that grounds its representations in a dynamical environment and derives its goals from a set of drives (Bach et al. 2006). Instead of understanding cognition as something isolated from motivation, emotion and a diverse perceptual environment, the Psi model is based on the idea that a system specifies physiological and/or cognitive demands, from which a motivational system derives goals. These goals then guide the activity of the agent, and reinforcement signals guide the learning process. As such, within the Psi model, cognition is understood as a way of facilitating abstraction in perception and action, establishing high-level goals, finding and following plans and strategies to reach these goals (Bach et al. 2006). A few notable software implementations relying on this theory are in existence, e.g., by Lim et al. (2005).

2.3.3 Dimensional Emotion Models

Besides the various attempts to discretely categorize emotions by their names, e.g., as shown by Kleinginna and Kleinginna (1981) or their eliciting conditions, e.g., as shown by Ortony et al. (1988), there are a variety of approaches that make use of a continuous space of three or more emotional dimensions. Therein, an emotion is described as a vector, further referred to as *emotional vector*.

As claimed by Picard (1997), all approaches at least include *arousal* and *valence* (also understood as *pleasantness*) as dimensions, and further states that merely two dimensions would not suffice for the description of all human-known emotions; for instance, intense fear and anger would be described by the same two dimensional vector, yet these two emotions are clearly different.

However, there are researchers who successfully implemented emotional systems which apply two dimensional models, e.g., Morgado and Gaspar (2005) use a model where an emotional vector is defined by an *achievement potential* P and an *achievement flow* F (see Figure 2.9). Mehrabian (1995) reports that a few researchers in the field of psychology successfully applied two-dimensional scales (with the dimensions *pleasure* and *arousal*) in empirical studies.

Breazeal (2003) presents a model inspired by the somatic markers hypothesis (as discussed in Section 2.2.2). Next to *arousal* and *valence*, it utilizes a *stance* tag which specifies how approachable the percept is to an agent; positive values are associated with advance whereas

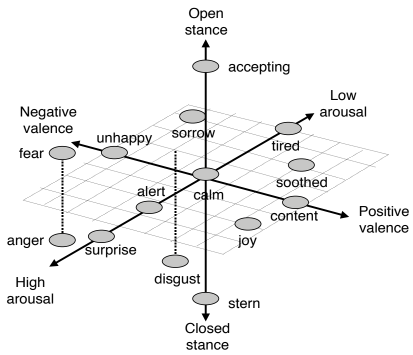

Figure 2.10: Mapping of emotional categories to arousal, valence, and stance dimensions. Picture from Breazeal (2003).

Emotion	P.	A.	D.
angry	-.51	.59	.25
bored	-.65	-.62	-.33
curious	.22	.62	-.01
dignified	.55	.22	.61
elated	.50	.42	.23
hungry	-.44	.14	-.21
inhibited	-.54	-.04	-.41
loved	.87	.54	-.18
puzzled	-.41	.48	-.33
sleepy	.20	-.70	-.44
unconcerned	-.13	-.41	.08
violent	-.50	.62	.38

Figure 2.11: Emotions associated to a PAD emotional vector.

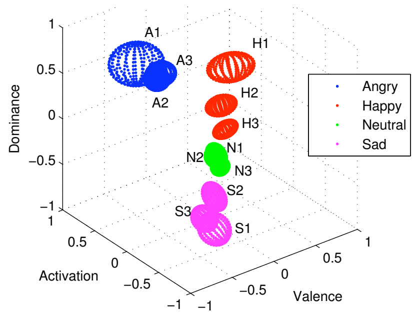

Figure 2.12: Covariance plot of the emotion classes angry (A), happy (H), neutral (N), and sad (S) in a three-dimensional emotion space of three speakers. Picture from Grimm et al. (2006).

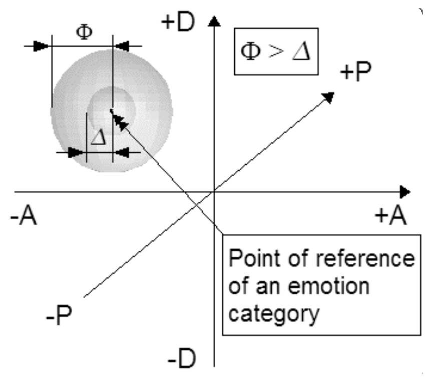

Figure 2.13: Threshold Φ and Δ for each emotion category. Picture from Becker et al. (2004).

negative values are associated with retreat. This allows for a mapping to emotional categories, as can be seen in Figure 2.10.

The most prominent approach of dimensional representation of emotions is the PAD emotion scale. It makes use of three dimensions, namely **p**leasure-displeasure, **a**rousal-nonarousal, and **d**ominance-submissiveness as, among others, developed by Mehrabian (1995) based on ideas tracing back to Wundt (1919). Through extensive empirical research, a fine graded set of emotions were located on the PAD scale.

Mehrabian (1995) asserts a sample of emotion terms with their corresponding emotional vector, which are shown in Figure 2.11.

PAD is used in numerous simulations, e.g., Becker et al. (2004) implemented the PAD model for use in a conversational agent, and Grimm and Kroschel (2007) use it in their system for automated recognition of emotions conveyed in speech. Figure 2.12 shows a plot that

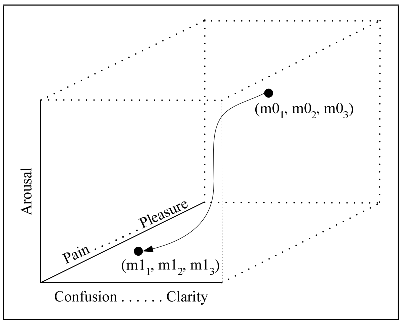

Figure 2.14: Representation of Emotion, from Cochran et al. (2006).

demonstrates the effectiveness of PAD in regard to indulgent classification of imprecise information, in this case the classification of emotions conveyed in sentences spoken by three different speakers. This is a feature also emphasized and exploited by Becker et al. (2004) which can be seen in Figure 2.13.

Similar to PAD, Cochran et al. (2006) propose a model consisting of three core parts: m_1 an arousal system that provides a measure for importance, m_2 a pleasure/pain system that assesses valence and m_3 a clarity/confusion system that provides a measure of competence.

Picard (1997) claims that *the question of whether to try to represent emotions with discrete categories or continuous dimensions can be considered a choice, as each representation has advantages in different applications.* However, as stated by Mehrabian (1995), continuous dimensions can be translated into discrete emotion categories by utilizing a mapping function.

2.3.4 Emotions as Signals

Regardless of the way emotions are elicited, there are basically three ways in which they can be related to the term "signal":

Signal to other entities: The most obvious interpretation is in the sense of the signals emitted by an entity while in a certain mood or while experiencing a certain emotion. These signals allow an observer to guess the emotional state of this entity and might alter decisions and plans. These aspects will be further discussed in Section 2.3.4.1.

Signals to the body: As previously stated, *emotions give life its urgency, a condition of immediate readiness to act* (Oatley 2004)—e.g., a specific emotion might lead to an increase in muscular tension, thus neural signals are transmitted throughout the body.

Emotions modeled as signals: Interesting for this chapter is the aspect of understanding and modeling emotions as signals. As described in Section 2.1.3, emotions refer to *a relatively brief episode [...] a response to an external or internal event of significance for the organism* (Davidson 2002). Triggering events leading to emotional reactions, their effects and their temporal aspects are, according to Picard (1997, pg. 141), best described using the concepts of signals, thus methods from linear system theory and digital sound processing can be applied to model affective signals and systems. Scherer (2000) goes even further by suggesting that researchers should go beyond classic linear process models and evaluate the potential of recently developed nonlinear dynamic modeling notions such as *coupled oscillators*, as used in dynamic system theory, and the concept of *hysteresis*, as used in, e.g., catastrophe theory and traffic simulation.

2.3.4.1 Recognition of Human Emotional States

The recognition of human emotional states is a broad topic covering various aspects, e.g., signal processing, natural language processing, pattern recognition, statistics, and many more. The methods used for recognition vary greatly depending on which feature of human emotional expression is given attention.

Measuring human internal states is not an easy task, which explains why, e.g., the deceivingly easy appearing problem of sleep phase analysis, that is the categorization of merely five sleep stages, requires so much effort and did not yet lead to fully satisfying results due to the ambiguity of the measured signals and the high amount of contained noise (Klösch et al. 2001). Compared to emotions, these stages are clearly defined and almost no ambiguity lies within their physiological expression. Whereas it seems rather straightforward to recognize if, e.g., a human is smiling or not, using a camera and image processing, it proves to be severely difficult when the optical conditions are not optimal, and to distinguish between real and fake smiles, that is, guessing the (internal) emotional state. Measuring the conductivity of the skin is easy, but how are the results to be interpreted, given that some people sweat permanently in high amounts, others never noticeably, some when positively excited, some when in a fearful situation?

These questions indicate the main problem of the recognition of almost any human feature, which the research field of speech recognition first experienced; that of recognizing human signals in a person-independent way. E.g., considering that even among speakers sharing the same language, the same spoken sentence results in different sound signals, due to variances in accent, pitch, speed, etc.

Next to these obvious difficulties, there are a number of factors which are even harder to process; for instance, for a word to gain meaning, an agent requires—next to the understanding of its common meaning—the context in which it was spoken, and knowledge of the speaker's personal history. Without these unspoken yet complex features, a word as simple as "mouse" could not be processed in a satisfying fashion. Did the speaker refer to the animal or the electronic device, a mascot of his favorite soccer team, or his spouse? Does it have negative emotional meaning because of the speakers phobia of furry animals, or a highly positive one because of last night's events?

According to Picard (1997, pg. 33), Nicolas Negroponte came up with the idea that there is no need for person-independent systems, considering the fast pace at which personal computers shrink in size; adapting to its user over a longer period of time, a personal assistant using a person-dependend recognition system will act as a mediator to other agents. However, these are dreams of the future—as of now, systems capable of recognizing and correctly classifying all human features holistically as accurately as humans are capable to do are nonexistent. Yet, some isolated features can be recognized by machines in a satisfying fashion, performing comparable to human classifiers or even better.

2.3.4.2 Measuring Physiological Signals

The bodily reactions that are believed to reveal emotional or affective states are covered by the term *sentic modulation*, according to Picard (1997) a term brought up by Manfred Clynes, 1977.

Not all of these reactions are visible to an untrained human observer, some may not be visible at all; some are beyond a human's direct influence, like heart rate, transpiration or cortical blood flow, others can easily be faked, like smiles, nervous lip or eyelid movement.

Generally, facial expression (Ekman 1993; Messom et al. 2005), vocal intonation and posture are visible (or audible) and can be culturally influenced to some extent and may be faked believably with sufficient training, but only to a certain extent (e.g., a fake smile can be distinguished from a *Duchenne*[6] *Smile* by a trained observer).

Physiological changes like heart rate (Picard et al. 2001) or blood pressure are hard to be estimated without tools that are applied to the subject's skin, although skin coloring can give certain clues in regards to blood pressure, and heart rate can be visible in rare situations.

There are several approaches for detecting these human emotional states, some are totally non-invasive, which makes it possible to perform the analysis without the subject knowing. These include features which can be recorded using cameras and microphones. Other approaches require (at the current state of the art) sensors being applied to the subject's skin, e.g., for measuring the heart rate and blood pressure or to perform an EEG. Thus, however, there is the chance of disrupting emotional responses, due to the subject being aware of being surveilled.

Historically, studies on emotions have usually taken place in fixed laboratory environments using systems developed for medical or psychological purposes (Peter et al. 2005). This however leads to unnatural environments where the subjects of interest are wired and have to sit still over a period of time, performing predefined actions. Thus, devices for mobile acquisition of data have been developed (e.g., Peter et al. (2005); Strauss et al. (2005)). These mobile devices allow for natural acquisition of data without inhibition or affection of the subjects emotional condition, and enable, e.g., software developers to utilize emotional data without having to care about physiological details or technical pitfalls of measurement engineering.

Machines have the advantage that they can utilize a broader spectrum of signals, e.g., in contrast to the human eye they can sense the infrared spectrum using a thermographic camera, and thus measure blood flow.

2.3.4.3 Emotions in Complex Media

As initially mentioned in Section 1.3, emotions can be encoded in or expressed through higher, cultural channels, like music and language.

Huge efforts are being taken in fields like automatic genre identification (Santini 2003) and information retrieval (Amati et al. 2007) for many reasons, e.g., to answer the question of how large amounts of electronically stored text as well as audio and video recordings can be processed and classified for applications like music recommendation systems, automatic

[6] The french neurologist Guillaume Duchenne (1806–1875) was a pioneer in electro-therapeutics. He performed experiments by electrically stimulating the muscles to induce facial expressions. He discovered that true smiles resulting from happiness also utilize muscles of the eye and thus, in his honor, a true smile is called Duchenne Smile

parental advisory systems (for a hypothetical protection of children's innocent eyes), or even automatic filtering of emails containing potential subversive threats.

The field of automatic genre identification is highly related to automatic emotion recognition in complex media, regarding the applied techniques as well as the possible fields of application; e.g., Subasic and Huettner (2000) see a movement to personalized services and decision support. The latter requires the capability to quickly analyze huge amounts of documents or recordings and to deliver an intuitive presentation to the user.

As for written or spoken text, as previously noted, this is not an easy task due to the ambiguity and imprecision which lies not only within the domain itself, natural language, but also the encoded emotions. Traditionally, natural language processing (NLP) has tried to constrain this ambiguity, Subasic and Huettner (2000) however try to explicitly represent and process these both types of ambiguity by using basic techniques from fuzzy logic in combination with NLP. They defined an *affect lexicon* which maps words to certain, possibly multiple, *affect categories* using two numerical weights, namely *centrality* and *intensity*.

An entry in the affect lexicon is defined like this:

```
<lexical_entry> <part_of_speech_tag>
<affect_category> <centrality> <intensity>
```

for instance:

```
"arrogance" noun  superiority 0.7  0.9
"aversion"  noun  repulsion   0.9  0.5
"avenge"    verb  violence    0.8  0.5
```

Explicitly denoting the part of speech (POS) resolves ambiguity that lies in the *how* the word is used in a sentence; e.g., the adjective "alert" describes a kind of intelligence whereas the verb "alert" refers to a warning. As such, the different affective meanings can be modeled by using two variables, namely centrality and intensity. The aspect of centrality is required to model the qualitative affiliation of a word to its affect categories by assigning a numerical value between 0 and 1. For instance, the verb "emasculate" could be described as affiliated with a centrality of 0.7 to weakness but only 0.3 to violence. Similarly, the intensity of a word in regard to an affect category is described in range of 0 and 1. E.g., the verb "abhor" in regard to the affect category *repulsion* shows an intensity of 1, whereas "displeasure" is less intense with 0.3 in regard to *repulsion*. (Related ways of modeling qualitative aspects of an emotion are discussed in Section 2.3.3.)

To calculate the overall emotional intensity and to create a centrality profile, the document is first normalized using NLP techniques and augmented by POS tags. Then, the process of *fuzzy semantic typing* is performed, that is applying the affect lexicon to the text corpus. The resulting intensity and centrality values are then combined and deliver insight into the document's content.

> A document with high overall intensity (> 0.7) in combination with a specific centrality profile (e.g., distaste/0.8 + violence/0.9 + pain/0.8) may indicate offensive and undesirable content (Subasic and Huettner 2000).

In accordance with Ortony et al. (1987), an affective lexicon must take into account that not only words such as "joyful", "hate", or "beloved" may carry emotional meaning, but also words like "plane crash" or "rollercoster", which Subasic and Huettner (2000) considered—their affective lexicon contained 3876 lexical entries.

Alas, approaches like these are not capable of realizing that the sentence

> Bob loved his wife Alice dearly with all his heart, yet she was happy with someone else and made fun about him being so amusingly unsuspecting

is highly sad, however, judging from the work of Strapparava and Mihalcea (2007), using an affective lexicon is still the common approach for affective sentiment analysis, while some approaches concentrate on adjectives exclusively (Kozareva et al. 2007).

2.3.5 Temporal Aspects of Emotions

Most models describe how emotions are elicited and define probabilites of emotion state transitions. However, only few care about what happens when time elapses without any further emotional input.

Picard (1997) describes the decay of emotions using the metaphor of a bell being struck. First, the bell emits a loud sound which then decays in intensity. *It builds up quickly to a peak, becoming very loud, then gradually fades until it is too faint to be heard.* Similarly, an emotion quickly builds up and slowly decays. When a bell is hit again with the same intensity shortly after the first strike, the intensity of the sound increases. With multiple small strikes, the sound can gain an intensity unachievable by a single strike. A strike too hard could break the bell, a strike too soft would not make the bell vibrate at all in order to produce a sound. Accordingly, an event emotionally too intense might produce a shock that could lead to a human's death, whereas an event of insignificance might not be perceived at all emotionally. Events of the same type occurring in short succession can increase the intensity of a given emotion.

2.3.6 Models of Personality

For various purposes, the ability to model different personalities is a vital or at least interesting aspect, e.g., for creating believable virtual characters, or for simulating realistic behaviour of humans in a traffic scenario. Personality is understood as the characteristics that distinguishes an agent from others.

Historically, personality was described by using the four Hippocratean humors: *sanguine* (courageous, hopeful, amorous), *choleric* (easily angered, bad tempered), *melancholic* (despondent, sleepless, irritable) and *phlegmatic* (clam, unemotional), depicted in Figure 2.15. A healthy personality was understood as a balanced mixture of these humors. These concepts are not entirely disregarded nowadays, however recent research in psychology and psychiatry came up with models scientifically more grounded, to describe a human's personality, both for healthy individuals and patients suffering from mental disorders. The *five-factor model* (FFM or *big five*) by Digman (1990) and McCrae and John (1992), amongst others, and the *three-factor model* (*big three* or *giant three*) by Eysenck and Eysenck (1985) have, according to Saggino (2000), emerged as the two most important psychometric theories in the field of personality.

The five factor model is, however, having a greater influence on affective computing. It is a model of personality that persists across time (Oatley 2004, pg. 103), which is based on traits

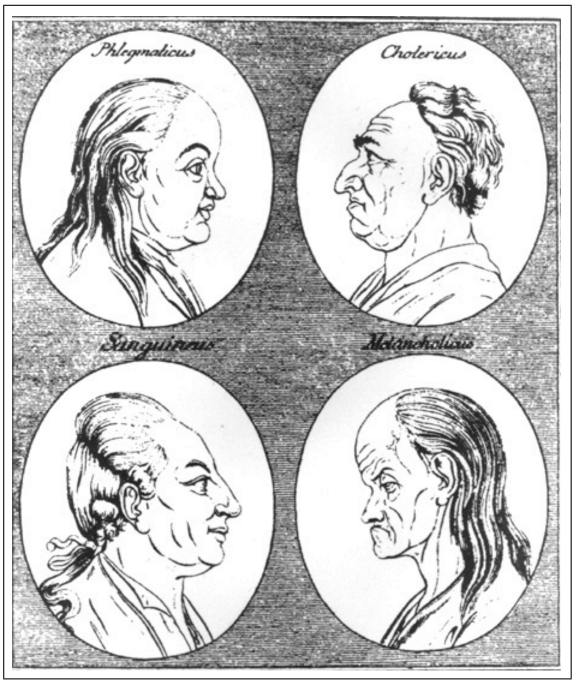

Figure 2.15: The four humors, Image of woodcut from 18th century text by Johann Kaspar Lavater.

influenced by genetics and chemicals. These traits, sometimes referred to with the acronym OCEAN, are

- **O**penness (to fantasies, emotions, aesthetics, ideas),
- **C**onscientiousness (including striving for achievement, dutifulness, self discipline),
- **E**xtroversion (warmth, sociability, cheerfulness),
- **A**greeableness (responsiveness to others including trust and compliance),
- **N**euroticism (proneness to anxiety, hostility, and depression).

See Table 2.3 for a clearly arranged explanation of each factor.

Kshirsagar (2002) implemented the five-factor model using a Bayesian belief network (BBN) for the use in believable virtual humans in a dialogue system. A BBN is a directed acyclic graph where each node represents a set of independent and mutually exclusive states together with condition probability tables for each node. The influence of parent nodes on their child node(s) is represented by a direct link and the states of the child nodes are calculated using the conditional probability tables.

Kshirsagar (2002) claims that the knowledge of emotional personality definitions and their classifications makes it possible to define rules that map personality to emotional states, in other words, modifying action and response tendencies. It is argued that rule-based systems do not account for the uncertainty that is involved in human behaviour. Thus Kshirsagar

Factor	Description	Adjectives
Extroversion	Preference for and behaviour in social situations	Talkative, energetic, social
Agreeableness	Interactions with others	Trusting, friendly, cooperative
Conscientiousness	Organized, persistent in achieving goals	Methodical, well organized, dutiful
Neuroticism	Tendency to experience negative thoughts	Insecure, emotionally distressed
Openness	Open mindedness, interest in culture	Imaginative, creative, explorative

Table 2.3: Five Factor Model, tabular from Kshirsagar (2002).

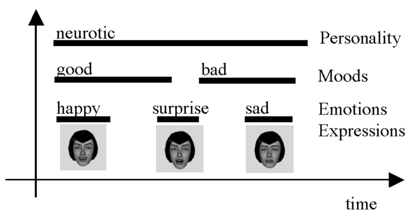

Figure 2.16: A layered approach to personality modeling, image by Kshirsagar (2002).

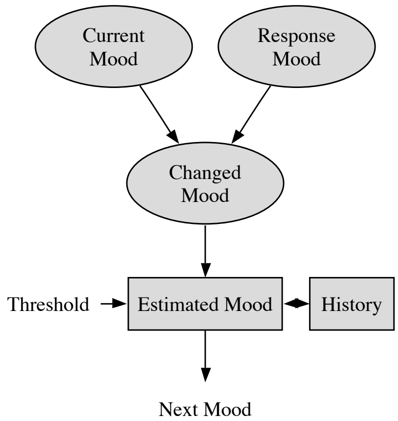

Figure 2.17: Personality model by Kshirsagar (2002).

(2002) states that BBNs were the natural choice to model the *uncertainty while retaining the underlying principles.*

75.5

For each of the five factors, one BBN was created, where each can be used by a user to specify a distinct personality (e.g., 30% neurotic, 70% extrovert). Figure 2.16 shows Kshirsagar's understanding of a model of personality across time; a high layer describing the personality which keeps persistent across time, a middle layer describing moods which are influenced from a higher level (personality) as well as a third, lower level, the emotional states. Kshirsagar (2002) states that he was not able to find sufficient evidence in the literature to differentiate and classify moods, thus proposed to make use of merely three simple states, namely "good", "neutral", and "bad". These moods bias the elicitation of emotions; whereas a good mood more easily leads to positive emotions, a negative mood increases the likelihood of negative emotions to arise.

The personality model π makes use of the above mentioned probability tables to calculate the probability $P(m_n|m_c, m_r)$ for a mood change, where m_c and m_r denote the *current mood* and *response mood*, respectively and m_n is the next mood, $m_n, m_c, m_r \in \mathcal{M} and \mathcal{M} = \{good, neutral, bad\}$. A response mood m_r is the result of an interaction with the environment. This interaction can further elicit a set $\mathcal{E}_r$ of emotions, in accordance to the OCC model mentioned in Section 2.3.1.2, where $\mathcal{E} = \{joy, anger, sadness, surprise, fear, disgust, neutral\}$ and $\mathcal{E}_r \subseteq \mathcal{E}$. $\mathcal{E}$ are the six basic emotions as defined by Ekman (1999) (see Section 2.1.3), plus a neutral state. As the model was applied to a dialogue system, mappings between text patterns and responses are defined. Each response is associated with one or more emotions $e_r \in \mathcal{E}_r$, which are each further associated with a certain probability. E.g., the response “I am busy” could be associated with 30% pride and 70% distress.

Let $P(m_n)$ be the probability of $m_n \in \mathcal{M}$ being the next mood, then

$$P(m_n) = P(m_n|m_c, m_r) \cdot P(e_r). \tag{2.6}$$

This calculation is performed for all pairs $(m_n, e_r) \in \mathcal{M} \times \mathcal{E}_r$, the emotion causing the highest probability $P(m_n)$ is selected.

Let $\mathrm{Ex}(\cdot)$ be a function that maps the 22 OCC emotions[7] to $\mathcal{E}$, and let Γ_{m_n} be the transition probability matrix for mood m_n, that is, a matrix over $\mathcal{E} \times \mathcal{E}$, each cell describing the likelihood that emotion $e_i \in \mathcal{E}$ succeeds emotion $e_j \in \mathcal{E}$ while in mood m_n, then

$$P(e_n) = \Gamma_{m_n}(\mathrm{Ex}(e_p), Ex(e_r)) \cdot P(e_r) \tag{2.7}$$

$P(e_n)$ is computed for every $e_r \in \mathcal{E}_r$, the emotion resulting in the highest probability is chosen as the next emotion.

2.4 Affective Architectures

As argued in the previous sections, it is rather obvious that machines directly interacting with humans would benefit from an at least basic understanding of *emotional semantics*, that is, not “just” possessing the ability to classify the emotional state the user currently is in, but also knowing how to adapt accordingly, and, if useful, react emotionally.

Systems supposed to react and act emotionally require—depending on their purpose—an architecture either augmented by emotional concepts, or based on emotional mechanisms. Sloman (2003) proposed a framework that allows to discriminate between various different approaches.

2.4.1 Sloman's Theoretical Framework

Sloman (2003) argues that one of the main problems in the study of emotions is people talking about different things when using words like “emotion”, “feeling”, etc. Thus, he suggest that there is the need for a framework to explain all the phenomena that are referred to by emotional and affective states and processes.

[7] Following the definition of Ortony et al. (1988), the term “OCC emotion categories” would be more adequate, but Kshirsagar (2002) explicitly speaks of “emotions”

His ideas are based on the assumption that the human mind has a kind of *information-processing architecture*, that generates the phenomena which are commonly described as "emotions". This idea was originally postulated by Simon (1967).

This framework should support a variety of different architectures, where each architecture may support different classes of states and processes, thus being capable of describing a child, a brain damaged adult, animals, robots, software agents, etc. Whereas a newborn child might not utilize the concept of shame of being unable to solve a mathematical equation, an adult human will most probably not be capable of experiencing the fear of insufficient battery power as a basic emotion.

It is important to keep in mind that any observed behaviour can result out of an infinite number of different architectures. Although highly unfeasible, it is thinkable that a human mind could be described with finite state automata.

Thus, Sloman suggests to only consider architectures that are possibly the result of a natural evolution process[8], which *leads to notion of a human architecture composed of many components that evolved under the influence of the other components as well as environmental needs and pressures.* Similar to Marvin Minsky's concepts presented in *Society of Mind*, this explains the mind as a collection of *coevolved suborganisms* (Sloman 2003) or *agents* (Minsky 1988)[9] acquiring and processing different kinds of information in a cooperative or competing way.

Considering such an architecture—which is to be understood as a virtual machine, not defined by physically grounded mechanisms—the question is which sorts of states and processes it allows. Consequently, the question then will no longer be whether robots, software agents, insects or newborn infants can *have* emotions, but rather, *which kinds* of emotions. Sloman opts for using the *design stance* (see Section 2.3), which makes it possible to define different sorts of emotions, different kinds of awareness, different kinds of learning, different kinds of intentionality, etc. in the context of the architectures that produce them.

Such an architecture exhibits components that have evolved at different times, with older portions found in many organisms and newer portions in relatively few animals. These components are described as layers, where the oldest layer (1^{st} layer) is understood as purely reactive. The newer, 2^{nd} layer introduces a "what-if" reasoning which enables the organism to reason about imaginary situations. The newest portion is a metamanagement or reflective layer (3^{rd} layer), which provides a form of self-consciousness by monitoring, categorizing and evaluating and controlling the rest of the system. Some researchers believe that humanlike language is a necessity for such a third layer, a notion brought up by Rolls (1999). Sloman (2003) does not share that assumption, however acknowledges that its functionality will be considerably modified by the existence of such a language, and notes that all information processing mechanisms require a certain type of formalism for encoding information. Similarly, Bach et al. (2006) states that many advanced problem solving strategies cannot be adequately modeled

[8]Sloman (2003, pg. 46) describes a natural evolution process as something that necessarily results in modular architectures: "in producing a really complex, multifunction design, evolution, like a human designer, will be constrained to use a modular organization—so that changes to improve one module will not impact disastrously on the functionality of other modules, and also so that some economy of genetic encoding is possible for a design using several copies of approximately the same module."

[9]Minsky (2006) uses the term *resource* instead of *agent*, which is justified by the observation that many readers assumed that an *agent* is a personlike thing. Nevertheless, we will continue using the original term *agent*.

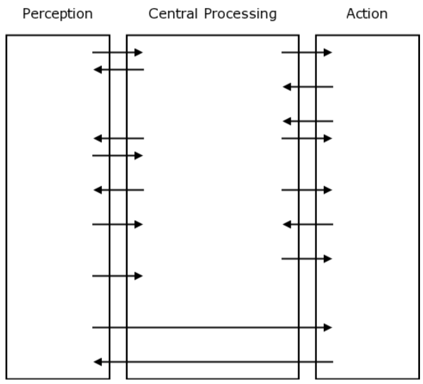

Figure 2.18: Triple tower model.

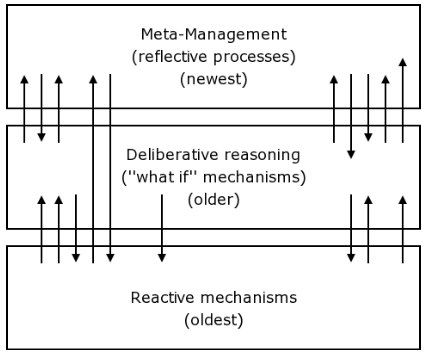

Figure 2.19: Multilayer model.

without assuming linguistic capabilities.

Based on this proposed architecture, between three different types of emotion processes can be distinguished. Namely, *primary emotions*, depending on the reactive layer, *secondary emotions*, depending on the deliberative mechanisms, and *tertiary emotions*, which depend on the reflective, 3^{rd} layer. This classification allows further subclasses, depending on the required precision.

Two very simplistic views of the architecture are presented by Sloman (2003), depicted in Figure 2.18 (*triple tower model*) and Figure 2.19 (*multilevel system*), which result in a third architecture when combined (see Figure 2.20). The triple tower model is a notion of information flowing through an organism, which is schematically constructed using three pillars: a *perceptual mechanism* feeding into *central processing* which further passes information to an *action* pillar. Multilevel systems describe an organism using layers, with the oldest, *reptilian* layer on the lowest level, and two recently evolved layers above it. This is similar to the previously brought definition of 1^{st} to 3^{rd} layer.

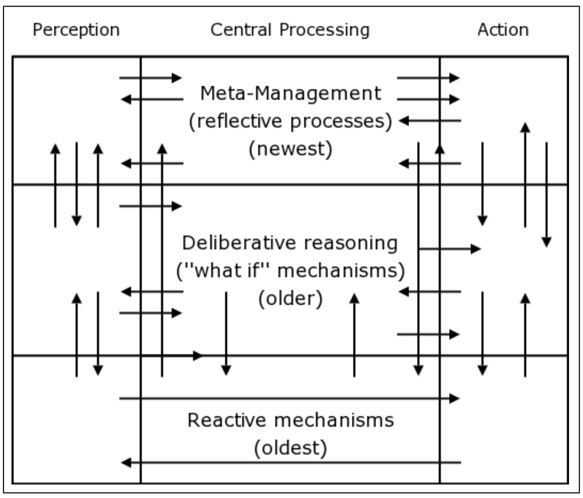

Figure 2.20: Combined model.

I_{e_n}	the intensity of emotion e at step n
I_{f_n}	the intensity of feeling f at step n
H_{f_n}	the hormone system for f at step n
A_{f_n}	the influence of feeling f
C_{e_f}	the weightings for e and f
B_e	a bias for an emotion e
I_{th_a}	emotion activation threshold
I_{th_s}	emotion selection threshold

Table 2.4: The properties used in the model by Gadanho and Hallam (2001).

2.4.2 A Hormonal Emotion Architecture

Gadanho and Hallam (2001) present an emotional architecture that utilizes a hormone system. In terms of Sloman (2003), it could be described as a triple-tower approach, a biologically inspired design-based model. The main features covered by that architecture are:

Valence: emotions can be positive or negative. In the specific model, three emotions with negative valence, namely sadness, fear, and anger, and one emotion with positive valence, namely happiness, are modeled.

Persistence: emotions possess temporal aspects, thus a new emotional state is not just dependent on new sensorial input but also the agent's emotional history (compare this with the bell metaphor by Picard (1997) mentioned in Section 2.3.5).

Bias: what is perceived is dependent on the current emotional state.

Recall that Gadanho's model makes use of four basic emotions,

$$\mathcal{E} = \{happiness, sadness, fear, anger\}. \tag{2.8}$$

Each of these emotions' intensity is calculated in discrete timesteps through linear weighted dependencies based on the agent's current internal feelings:

$$\mathcal{F} = \{hunger, pain, restlessness, temperature, eating, smell, warmth, proximity\}. \tag{2.9}$$

The feelings described are a result of the perception of raw sensory stimuli, but can also be invoked or strenghtended by subjective sensations. The model is mainly based on the properties shown in Table 2.4.

Following insights by Damasio (1994), the agent's emotional states and feelings show mutual reaction; feelings can invoke particular emotions, and emotions can invoke particular feelings when their intensity exceeds the activation threshold, I_{th_a}. This is achieved by implementing a simplified hormone system that associates each feeling with a specific hormone that is discharged in a specific quantity when an emotion is triggered by an external sensorial or internal sensation. Influenced by a coefficient parameter C_h, the sum of a hormone level H_{f_n}, and the sensation's value S_{f_n}, define the resulting feeling's intensity, the hormone level itself is further dependent on the emotion influences A_{f_n} (which is biased by an attack gain, α_{up}, and decay

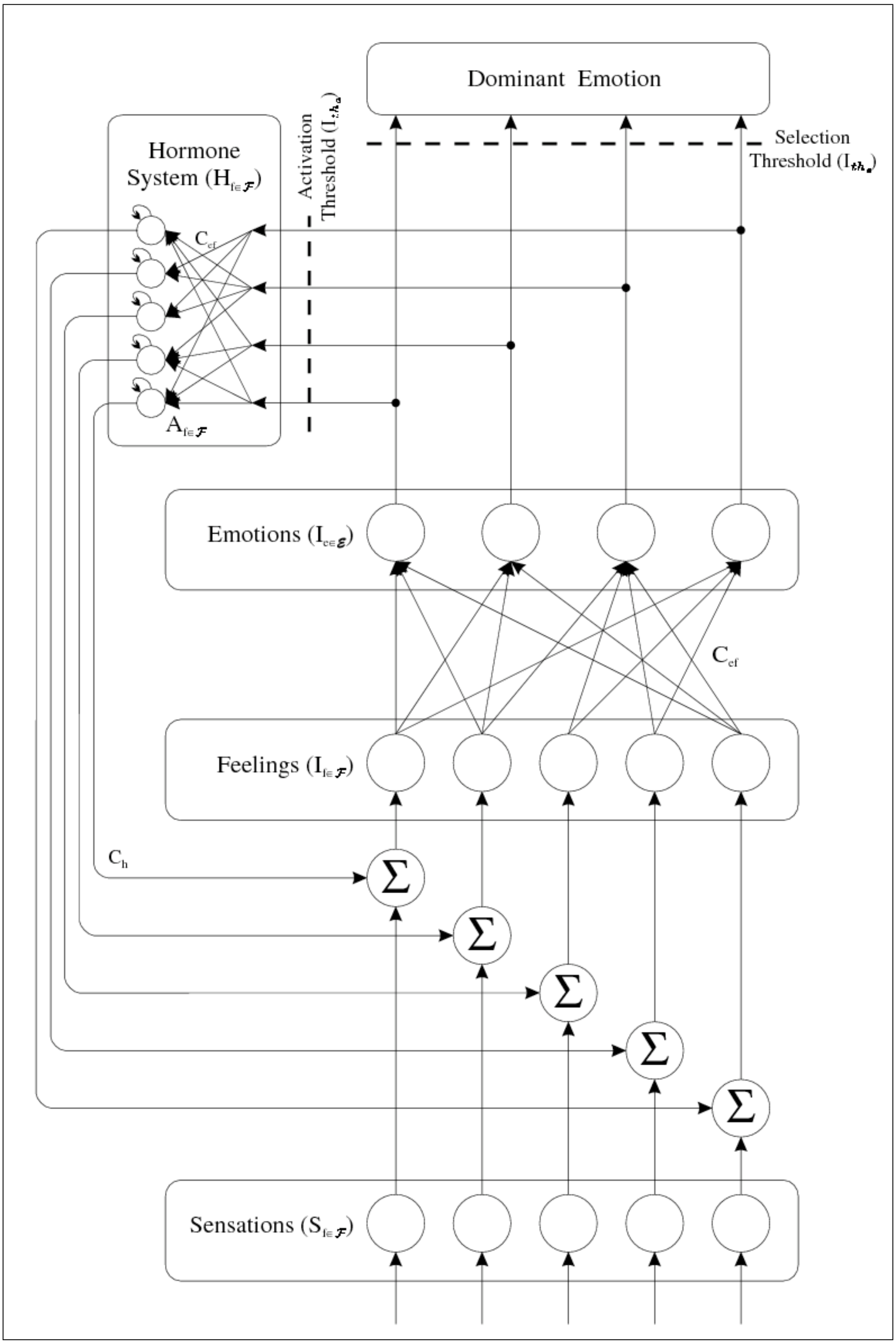

Figure 2.21: The emotion model. Image by Gadanho and Hallam (2001).

gain, α_{dn}, and its previous value). The latter produces the hormone system's memory effect. The hormone with the highest value defines the dominant emotion. See Figure 2.21 for an overview.

The intensity $I_{e_n} \in \mathbb{R}$ of emotion e's intensity at step n can be calculated as follows. First let us define a truncation function $Th_{[b_-,b_+]}(\cdot)$,

$$\mathrm{Th}_{[b_-,b_+]}(x) = \begin{cases} b_- & \text{if } x < b_- \\ b_+ & \text{if } x > b_+ \\ x & \text{otherwise} \end{cases} \tag{2.10}$$

where $b_-, b_+ \in \mathbb{R}$. Then, for all $e \in \mathcal{E}$ and all $n \in \mathbb{N}$,

$$I_{e_n} = \mathrm{Th}_{[0,1]}(B_e + \sum_{f\ \in\mathcal{F}} (C_{ef} I_{f_n})), \tag{2.11}$$

where $B_e \in \mathbb{R}$ is an emotion bias, $C_{ef} \in \mathbb{R}$ is a coupling coefficient between emotion $e \in \mathcal{E}$ and feeling $f \in \mathcal{F}$, and $I_{f_n} \in \mathbb{R}$ is the intensity of feeling f_n, For coupling the latter, let $H_{f_n} \in \mathbb{R}$ be the hormone level of f_n and S_{f_n} an external stimulus associated with f_n. Then, for all $f \in \mathcal{F}$ and all $n \in \mathbb{N}$,

$$I_{f_n} = \mathrm{Th}_{[0,1]}(C_h H_{f_n} + S_{f_n}), \tag{2.12}$$

$$H_{f_n} = \begin{cases} 0 & \text{if } n = 1, \\ \alpha_n H_{f_{n-1}} + (1 -\ \alpha_n) A_{f_{n-1}} & \text{if } n > 1, \end{cases} \tag{2.13}$$

where $A_{f_n} \in \mathbb{R}$ are emotion influences defined as

$$A_{f_n} = \sum_{e\in\mathcal{E}: I_{e_n} > I_{th_a}} C_{ef} I_{en} \tag{2.14}$$

and the attack or decay gain $\alpha_n \in [0, 1]$ is defined as

$$\alpha_n = \begin{cases} \alpha_{up} & \text{if } |A_{f_n}| > |H_{f_n}|, \\ \alpha_{dn} & \text{otherwise.} \end{cases} \tag{2.15}$$

The coupling coefficients C_{ef} can be understood as defining the personality of an agent equipped with this emotional system, regulating the intensity an emotion gains upon external and internal stimuli. For instance, a choleric character would react with much more anger upon the approach of an obstacle than a relaxed one. Similarly, the bias, B_e, further strenghtens certain characteristics.

The idea behind the coupling coefficients is that every stimuli can have a potential effect on every single emotion to a certain degree, e.g., approaching a source of food might make an agent

- happy, for there is food,
- fearful, for there might be other agents posing a threat,
- angry, for it might need to fight,

I_{th_a}	emotion activation threshold	0.2
I_{th_s}	emotion selection threshold	0.2
C_h	hormone coefficient	0.9
α_{up}	hormone attack gain	0.98
α_{dn}	hormone decay gain	0.996

Table 2.5: Constants as proposed by Gadanho and Hallam (2001).

- sad, for it might lose that fight.

As such, as depicted in Equation 2.11, the sum of all weighted stimuli feed into the emotion's intensity.

Besides the external stimuli, internal sensations add to the emotional experience. As mentioned above, this is grounded in the somatic markers hypothesis by Damasio (1994). For this reason, I_{f_n}, as described in Equation 2.13, the intensity of a feeling f in step n, is not just influenced by the sensation value S_{f_n}, but also by the current hormone level, H_{f_n}, of that feeling. The factor C_h defines the overall degree of influence the hormone system plays.

Equation 2.14 describes how the hormone level of a feeling, H_{f_n}, builds up over time, or dacays. Firstly, using Equation 2.15, it is determined whether the intensity of the feeling is rising or falling. Varying the values of α_{up} and α_{dn} the intensity increase or decay can be sped up or slowed down. Table 2.5 depicts values as proposed by Gadanho and Hallam (2001).

The hormone level is dependent on two factors: the hormone level itself at time $n-1$,[10] and the *emotional influence*, A_{f_n}, at time $n-1$. This emotional influece is the above mentioned back-coupling of emotions to feelings in the sense of somatic markers. Again, the emotions' dependencies, C_{e_f}, are used, this time backwards—that is, the values of the emotions are used to calculate the resulting feelings that feed into the hormone system.

However, this will only happen if the emotion is sufficiently vigorous, that is, its intensity is higher than the activation threshold I_{th_a}.

Finally, from all emotions whose intensities exceed the threshold I_{th_s}, the most dominant, i.e., the one with the highest intensity value, is selected.

These described mechanisms lead to a very simple but effective control system. For example, imagine a robot colliding with an obstacle. The collision produces a pain sensation which is captured by the pain feeling. Fear is defined as having a strong dependency on pain, thus the fear intensity rises. Gadanho and Hallam (2001) describe a scenario that pictures this behaviour quite well:

> If the intensity is high enough then fear will produce hormones. The hormone associated with pain will quickly build up during this collision. This will make the fear emotion grow stronger and possibly overtake other existing emotions. When the robot finally manages to cease the collision, it will still have pain not because the pain sensation is still there, but because the hormone associated with pain has a high value. So the fear emotion will persist while the hormone gradually decreases in value. This means that while the robot is gaining distance from the obstacle,

[10]This is depicted wrongly in the original paper by Gadanho and Hallam (2001) where H_{f_n} is said to depend on H_{f_n} instead of $H_{f_{n-1}}$.

> the fear will still be there. Nevertheless, it will usually fade away as soon as a short distance is gained and the risk of further collisions has diminished.

In contrast, a robot only equipped with a reactive control mechanism has to explicitly set a goal of gaining a certain distance from the object it has just collided with, as it continuously has to check for the fulfillment of that goal.

Chapter 3

Traffic Application

3.1 Introduction

In order to explore synthesized emotions in an artificial intelligence system that is used to simulate a "real world" problem, a traffic simulation called *SAD* (**S**imulation of **A**ffective **D**rivers) was implemented. The decision to utilize such a scenario is rooted in the two facts. Firstly, driving is a highly emotional task, and as current traffic models ignore this crucial aspect so far, a gap to be filled was found. For instance, Rothery (2002) does not mention the word *emotion* a single time, yet this publication is giving an in-depth state-of-the-art overview of car following models. Secondly, a plethora of demonstrations for emotional AI make use of "grid-worlds", where agents have to survive in a virtual two dimensional environment, provided with scarce resources while additionally threatened by dangerous creatures or other agents (Inoue et al. 1996). These scenarios seem to offer only questionable realism and are not particularly suited for comparison with real-world situations of everyday-life due to exceeded abstraction within a setup of overly discreet time and space. Contrary to that, a traffic simulation can reveal agent behaviour that almost everybody is very familiar with; annoyingly slow drivers, threatening aggressive behaviour by speeders, and the corresponding consequences like slow moving car columns and traffic jams.

Thus, the goal of this demonstration is not only to provide a dry comparison of reactive and emotional agent systems acting on a simulated freeway, but rather to provide a simulation where the agent behaviour can be seen and intuitively understood just by watching the traffic flow.

This chapter will describe the emotional architecture that is used to equip the drivers with emotions. Then, an introduction to traffic simulation will be given, followed by the description of the emotional traffic simulation. Further, empirical results are presented and discussed.

3.2 Traffic Simulations

The motivation to create (freeway) traffic simulations is mainly rooted in one question: How can the throughput of a freeway be increased by maintaining safety? In other words, how to avoid traffic jams and accidents? Traffic jams are usually caused by bottlenecks, e.g., on-ramps, lane closings, uphill gradients, accidents, overtaking trucks on a two lane highway; however, according to Treiber et al. (2000), the type of bottleneck seems not to be of importance.

There are two main methods to delineate traffic mathematically, namely *macroscopic* and *microscopic* models. While the former describe the dynamics in terms of aggregate quantities like density or flow, the latter model the motion and action of individual vehicles. Such microscopic simulations include continuous-time car-following models and cellular automata (Treiber et al. 2000).

The natural approach to explore emotionally capable agents is to choose a car-following model. Furthermore, as Treiber et al. (2000) states, car-following models are the preferred method for a direct comparison with empirical data as the position and velocity of each car is known, and thus can be used to reconstruct the way how data are obtained by usual induction-loop detectors (e.g., by spawning vehicles in accordance to recorded traffic data).

Gettman and Head (2003) state that microscopic simulations must model the key driver behaviours that produce opportunities for crashes. Those behaviours are:

- car following,
- gap acceptance, and
- lane changing.

However, most traffic simulations are accident free; nevertheless these are the behaviours that are explicitly modeled.

3.2.1 Basic properties of moving vehicles

The basic parameters to describe moving vehicles are the following (see also Figure 3.1):

- x_α: the position of vehicle α,
- L_α: its length,
- v_α: its velocity,
- s_α: the distance to vehicle $\alpha - 1$, the car in front,
- Δv_α: the relative velocity between α and $\alpha - 1$.

The velocity v_α of vehicle α is defined as the rate of change of position, that is the distance it covers over a specified time. Formally,

$$v_\alpha(t) = \dot{x}_\alpha = \frac{\mathrm{d}x_\alpha}{\mathrm{d}t}, \tag{3.1}$$

where $x_\alpha(t)$ denotes the longitudinal position of vehicle α at time t. The acceleration, a_α, of a vehicle α is defined as the rate of change of velocity, formally

$$a_\alpha(t) = \dot{v}_\alpha = \ddot{x}_\alpha = \frac{\mathrm{d}v_\alpha}{\mathrm{d}t}. \tag{3.2}$$

The minimum braking distance s for a vehicle to come to a complete standstill is calculated by

$$s = \frac{v^2}{2 \cdot b}, \tag{3.3}$$

where b is the maximum deceleration.

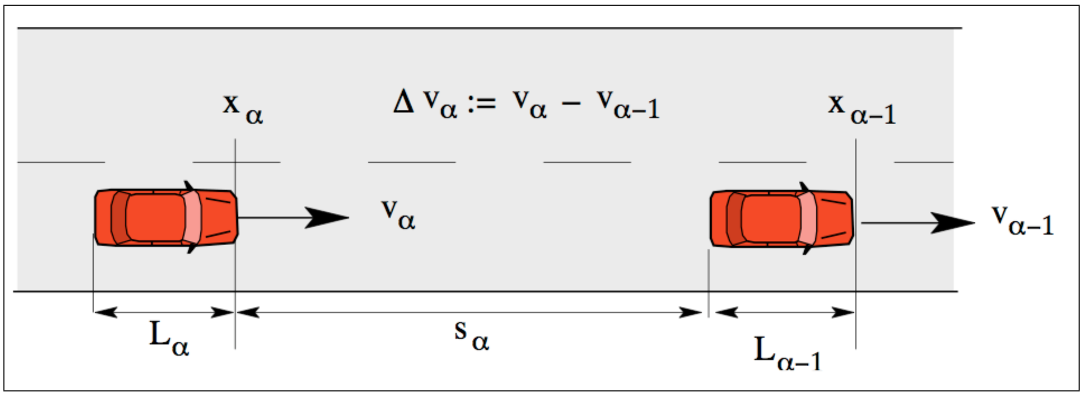

Figure 3.1: Car following. Picture from Treiber, Hennecke, and Helbing (2000).

3.2.2 Single-Lane Car Following

Single-lane car following models describe the behaviour of vehicles acting on a virtual lane in dependance to their respective leading vehicle, i.e., its velocity and the distance to each other. In dependance to these variables, the only parameter an imaginary driver is in control of is the velocity of his own vehicle. This is simply achieved by either increasing, keeping or decreasing the current speed. Simple car following models do not discriminate between actions that are used in a real car to perform these speed adjustments, such as hitting the brake or the throttle; in other words, the concept of an agent controlling a vehicle is not applied. Rather, these models directly calculate the velocity of the vehicle.

The basic idea behind some of these models is that a driver will always want to keep a certain distance to the leading vehicle. In its simplest form, this distance is the sum of three variables:

1. the length L_α of the own vehicle, as the minimum distance a driver should keep to a vehicle in front at complete standstill;

2. the distance s_r that a vehicle with speed v would cover during a reaction time T_r, that is, the time which a driver would both need to realize that braking is necessary and to act accordingly, e.g., start to brake, defined as

$$s_r = T_r v(t); \tag{3.4}$$

3. the distance a vehicle with velocity v would need to come to a complete stop, that is the braking distance s, as defined in Equation 3.3.

According to Rothery (2002), these speed-spacing relations can be represented by

$$s_\alpha(t) = L_\alpha + T_r v(t) + \gamma v(t)^2, \tag{3.5}$$

where γ denotes some factor influencing the braking distance.

The term γv^2 is intended to add sufficient spacing for the case that the leading vehicle comes to a full stop. In view of Equation 3.3, if the driver is supposed to act within a maximum of safety in mind, then

$$\gamma = \frac{1}{2 \cdot b_\alpha}, \tag{3.6}$$

where b_α is the maximum average deceleration.

According to Rothery (2002), a typical value empirically derived would be $\gamma \approx 0.075 s^2/m$. He further states that a less conservative interpretation would be

$$\gamma = \frac{1}{2 \cdot b_\alpha} - \frac{1}{2 \cdot b_{\alpha-1}}, \tag{3.7}$$

where only the relative deceleration of vehicle α and its leading vehicle $\alpha - 1$ is taken into account. This, however, ignores the possibility that $\alpha - 1$ can come to a sudden stop, e.g., by crashing into a stopped column of cars.

According to Treiber et al. (2000), simplified single-lane car following models are mainly defined by their acceleration function $\dot{v}_\alpha(t + T_r)$. The idea behind these models is that the acceleration an imaginary driver would choose at time $t + T_r$ depends on the relative velocity $\Delta v_\alpha(t)$ to a leading vehicle at time t. Jiménez et al. (2000) presents an acceleration function that is defined as

$$\dot{v}_\alpha(t + T_r) = -\lambda \cdot \Delta v_\alpha(t), \tag{3.8}$$

where the relative velocity, or approaching rate, is defined as

$$\Delta v_\alpha(t) = v_\alpha(t) - v_{\alpha-1}(t) \tag{3.9}$$

and λ is some factor which can be defined in multiple ways. Following Jiménez et al. (2000), λ could be defined as

$$\lambda = \frac{a_{l,m} \cdot \dot{x}_\alpha^m(t + T_r)}{s_\alpha^l}, \tag{3.10}$$

where

- l and m are speed and distance headway parameters,
- $a_{l,m}$ is a constant defining the characteristics of the driver, and
- $s_\alpha = x_{\alpha-1} - x_\alpha - L_\alpha$ is the "bumper-to-bumper" distance.

The deceleration is simply defined as the negative acceleration: $-\dot{v}_\alpha(t + T_r)$.

Due to the dependance on a leading vehicle to calculate a vehicle's acceleration, these models do not suffice for very low traffic densities. Thus, a vehicle's acceleration needs to be brought into correlation with a driver's desired velocity.

3.2.3 Intelligent Driver Model

The intelligent driver model (IDM) (Treiber et al. 2000) solves this shortcoming and adds a concept of a driver's desires. This IDM driver can be defined by two parameters:

- a tendency to accelerate on a free road, $a_f(v_\alpha)$, and

- a tendency to brake with deceleration, $b_{\text{int}}(s_\alpha, v_\alpha, \Delta v_\alpha)$.

The acceleration $\dot{v}_\alpha$ the driver of vehicle α will aim for is defined by the tendency to accelerate minus the tendency to brake. Let $a_f(v_\alpha)$ be the tendency to further accelerate the vehicle which has a current velocity of v_α, and $b_{int}(s_\alpha, v_\alpha, \Delta v_\alpha)$ the tendency to brake considering the distance to the front vehicle s_α, the current speed v_α, and the relative speed to the front vehicle Δv_α, then

$$\dot{v}_\alpha^{\text{IDM}}(v_\alpha, s_\alpha, \Delta v_\alpha) = a_f(v_\alpha) - b_{int}(s_\alpha, v_\alpha, \Delta v_\alpha), \tag{3.11}$$

where

$$a_f(v_\alpha) = a_{\text{ed},\alpha}\left[1 - \left(\frac{v_\alpha}{v_{\text{d},\alpha}}\right)^\delta\right], \tag{3.12}$$

and $a_{\text{ed},\alpha}$ denotes the acceleration of vehicle α in everyday traffic, and δ denotes an acceleration exponent. The ratio between current speed v_α and the desired speed $v_{\text{d},\alpha}$ defines the intensity of the acceleration. The closer v_α gets to $v_{\text{d},\alpha}$, the less the driver will need to accelerate. The acceleration exponent will further increase acceleration the farther the current speed is away from the one aimed for.

The tendency to brake depends on the ratio of the desired distance, and the actual distance to the front. Let s^* be the *desired minimum gap* and s_α the actual gap, then

$$b_{\text{int}}(s_\alpha, v_\alpha, \Delta v_\alpha) = a_{\text{ed},\alpha}\left(\frac{s^*(v, \Delta v)}{s_\alpha}\right)^2, \tag{3.13}$$

where, similar to Equation 3.5,

$$s^*(v, \Delta v) = s_{\text{g},\alpha} + T_{\text{h},\alpha} v + \frac{v \Delta v}{2\sqrt{a_{\text{ed},\alpha} b_{\text{ed},\alpha}}}, \tag{3.14}$$

and where

- $s_{\text{g},\alpha}$ denotes a constant safety gap,
- $T_{\text{h},\alpha}$ denotes the desired safety time headway when following other vehicles,
- $b_{\text{ed},\alpha}$ is the *comfortable* braking deceleration in everyday traffic.

Combining Equations 3.11, 3.12, and 3.13, the following equation, due to Treiber et al. (2000), we obtain:

$$\dot{v}_\alpha^{\text{IDM}}(v_\alpha, s_\alpha, \Delta v_\alpha) = a_{\text{ed},\alpha}\left[1 - \left(\frac{v_\alpha}{v_{\text{d},\alpha}}\right)^\delta - \left(\frac{s^*(v_\alpha, \Delta v_\alpha)}{s_\alpha}\right)^2\right]. \tag{3.15}$$

Different types of vehicles, or *driver-vehicle unit* can be defined by applying different values as parameters. Treiber (2006) states that, e.g.,

- trucks can be characterized by low values of v_d, a_ed, and b_ed,
- careful drivers keep a high safety time headway T_h,
- aggressive drivers could be characterized by a low T_h and high values of v_d, a_ed, and b_ed.

The standard parameters used in an IDM simulation are depicted in Table 3.1.

var	Parameter	Value Car	Value Truck	Comment
v_0	Desired velocity	$120km/h$	$80km/h$	
T	Time headway	$1.5s$	$1.7s$	Realistic values vary between $2s$ and $0.8s$ and even below
s_0	Minimum gap	$2.0m$	$2.0m$	Always kept, also in complete standstill
a	Acceleration	$0.3m/s^2$	$0.3m/s^2$	Very low values to enhance the formation of stop-and-go traffic. Realistic values are $1-2m/s^2$
b	Deceleration	$3.0m/s^2$	$2.0m/s^2$	Very high values to enhance the formation of stop-and-go traffic. Realistic values are $1-2m/s^2$

Table 3.1: Parameters for cars and trucks in IDM simulations, as proposed by Treiber (2006).

3.2.4 Multi-Lane Traffic with Lane Changing

While single-lane car following models are only concerned with the adjustments of the vehicle's velocity based on the velocity of a leading vehicle, multi-lane models are required to take additional factors into account, namely, drivers on the neighboring lanes and their velocities. Further, the driver has to perform lane changes when certain criteria are fulfilled. According to Treiber and Helbing (2002), a lane change occurs if

- it is safe to be performed, thus fulfilling the *safety criterion*, and
- an appeal for a change was given, that is, the target lane being more attractive than the current lane, thus satisfying the *incentive criterion*.

Let α, β be two consecutive vehicles, then the net distance, $s_{\alpha\beta}$, between these two vehicles is defined as

$$s_{\alpha\beta} = |x_\alpha - x_\beta|, \tag{3.16}$$

where x_α and x_β are the vehicles' positions as defined in Section 3.2.1. An acceleration function $a_{\alpha\beta}$ may then be given as

$$a_{\alpha\beta} = \dot{v}^{\mathrm{IDM}}(v_\alpha, s_{\alpha\beta}, v_\alpha - v_\beta). \tag{3.17}$$

Thus, the safety criterion checks that the brake delay $a_{\delta\alpha}$ of vehicle δ driving behind vehicle α after a fictive lane change does not exceed b_{save}:

$$a_{\delta\alpha} \geq -b_{\mathrm{save}}. \tag{3.18}$$

For the incentive criterion, as defined by Treiber and Helbing (2002), the accelerations are also considered: An appeal for a lane change is given, if, after a fictive lane change, the sum of a vehicle's own acceleration and accelerations of involved neighboring cars—weighted with a politeness bias p—is at minimum an ϵ higher than before. A right to left lane change, denoted as $R \rightarrow L$, occurs if the following criterion is fulfilled:

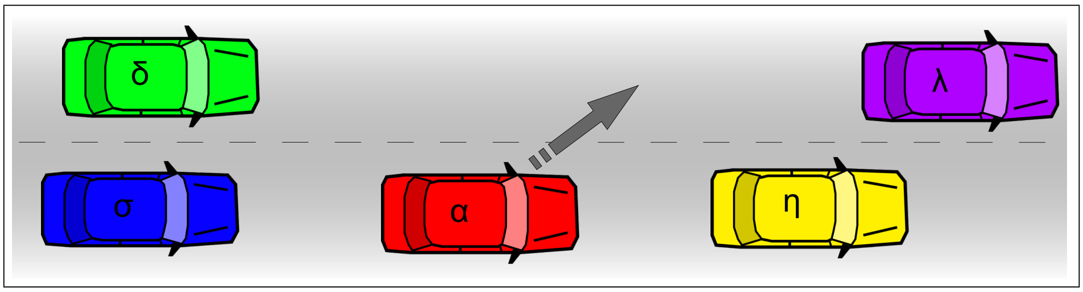

Figure 3.2: Lane changing cars.

var	Parameter	Realistic Values	Remarks
p	Politeness factor	$0 \ldots 0.5$	Atypical: $p > 1 \rightarrow$ altruistic behaviour $p = 0 \rightarrow$ purely selfish behaviour $p < 0 \rightarrow$ malicious behaviour
b_{save}	Maximum safe deceleration	$4m/s^2$	$b_{save} < 9m/s^2$
a_{thr}	Threshold	$0.2m/s^2$	$a_{thr} < a$

Table 3.2: Standard values used in IDM simulations, given by Treiber (2007).

$$\underbrace{a_{\alpha\lambda} + p \cdot (a_{\delta\alpha} + a_{\sigma\eta})}_{\text{after the change}} > \underbrace{a_{\alpha\eta} + p \cdot (a_{\sigma\alpha} + a_{\delta\lambda})}_{\text{before the change}} + \epsilon. \tag{3.19}$$

Again, individual driver characteristics can be defined by varying the given parameters. See Table 3.2 for values proposed by Treiber (2007).

3.3 Simulation of Affective Drivers—SAD

As previously mentioned, the motivation to implement a traffic simulator lay within the wish to explore the usefulness of agents augmented with emotional capabilities. Further, the simulation was supposed to be suitable for both empirical as well as intuitive analysis, and to provide a means to experiment with different environmental setups and different agent personalities.

The idea to simulate traffic situations is not a novel one (Lieberman and Rathi 2002). There are numerous implementations existing, e.g., by Treiber and Helbing (2002), Barcelí et al. (2005), or Ehlert and Rothkrantz (2001), using a wide variety of different models. Combining computational models of traffic with simulated emotions, is, however, as it seems a new approach.

For SAD, the only feasible method appeared to be to build upon a microscopic paradigm, that is, calculating each vehicles' actions and properties individually, including the simulation of a human driver with simulated emotions.

As opposed to macroscopic simulations, where the driver's psychology is disregarded insofar as the driver is merely seen as a physical component, e.g., in a granular floating process, microscopic simulations take—with varying degree—certain human aspects into account. As Saad (1993) points out, road users play an essential role in traffic systems and thus states that the study of their behaviour and the underlying psychological mechanisms is fundamental in any research aimed at increasing system reliability. This claim is supported by Vaya et al. (1993), describing the identification of the psychological factors and behaviours associated with driver behaviour as an important subject in traffic safety research.

However, it appears that affective AI has largely been ignored by the traffic science community so far, although it is an accepted theory that emotions play a very important role in human driving. `SAD` is an attempt to provide a virtual environment that, besides a classical reactive IDM architecture, incorporates an emotional system. This makes it possible, to, e.g., let purely reactive drivers and purely emotional drivers compete at the same time. `SAD` is an accident free simulation.

3.3.1 Don't Weep and Drive—Emotions and Traffic

As mentioned in Section 2.4.1, it is rather meaningless to define synthetic emotions and connect them with natural language words to achieve a certain effect. Thus, a simulation with affective drivers could be defined using emotions $E_1, E_2, \ldots, E_n$ that correlate with certain driving behaviours.

However, as stated in Section 3.1, one purpose of this simulation is to provide a demonstration that explains itself intuitively to a human observer, a demonstration for the usage of artificial emotion systems. As such, it is convenient to use familiar words to describe behaviour that is perceived accordingly. It is, however, important to keep in mind not to forget that it's not real human emotions that are being dealt with. As such, `SAD` does not yet claim to serve as a realistic model to real human driving behaviour.

Under the assumption that human emotions do have an important impact on driving behaviour (Saad 1993; Vaya et al. 1993; Carsten 1993), it is necessary to find out which these could be, and how they can be modeled. Anxiety (Vaya et al. 1993), aggression (Saad 1993) and happiness (Carsten 1993) were stated as influencing variables. To be in accord with emotional theories, these suggestions were transformed into a set of the presumably most important emotions in regard to driving:

Fear: Vaya et al. (1993) see anxiety as one of those variables which have a significant role in driving behaviour as research reveals. It is argued that

1. research would show that a significant number of drivers involved in traffic accidents suffer a high level of anxiety and insufficient level of self-confidence.
2. anxiety would be one of the emotional reactions most frequently related to driving behaviour in literature, and
3. anxiety and stress would be a factor which is included in most theoretical models explaining driver behaviour.

Anger: Other researcher see anger as one of the most important factors in driving behaviour, e.g., described as aggressiveness (Zuber et al. 2006). Lieberman and Rathi (2002) describe a randomly assigned "aggressiveness index" ranging from 1 (very aggressive) to 10 (very cautious) drawn from a uniform distribution to represent the range of human behaviour. Similarly, as described in Section 3.2.3, the IDM defines a politeness value which can be understood in the same fashion.

Lieberman and Rathi (2002) argue that increased aggressiveness results in the selection of higher speeds, as well as the acceptance of higher risks, i.e., bold lane-change maneuvers. Ardekani et al. (2002) state that the degree of driver aggressiveness manifests itself in speed and acceleration rates and thus has an impact on the fuel consumption rate.

Happiness: Contrary to intuition, happiness is described as an emotion with negative correlation to driving efficiency, in regards to safety(Carsten 1993) as well as milage (Mesken 2006). However, this latter negative correlation is shared with every other emotion, according to Mesken (2006).

Sadness: Unlike fear and anger, happiness is rarely mentioned in the traffic research literature. However, Mesken (2006) states that it has a significant negative correlation with milage.

3.3.2 Simulation Architecture

`SAD` simulates vehicles driving on an arbitrary number of cyclic lanes[1] in discrete time steps.

The lanes are abstractions insofar as they only allow continuous placement in one dimension, the driving direction. Contrary, lane changes are discrete events, where the vehicle "jumps" from the current to a neighboring lane.

A vehicle α is defined by properties such as its position s_α on the lane, its velocity v_α, and its length l_α. Additionally, it is possible to describe additional properties, like the vehicle's maximum speed, rate of acceleration, efficiency of the brakes, etc. The possibility to define these properties allows to experiment with different types of vehicles, like trucks, family vans, sports cars, etc.

A vehicle's movements are controlled by an agent which is associated with the vehicle when it is spawned into the simulation. The agents' emotional system is heavily based on the architecture of Gadanho and Hallam (2001)—see Section 2.4.2 for the detailed overview.

Agents are instantiated with definable personalities, which can be accomplished by assigning weightings, which map the agent's stimuli with emotional impetus. The simulation, i.e., the vehicles' movements and thereof resulting statistics, are calculated in discrete time steps, measured in seconds.

[1] The end of a lane is connected with its own beginning, as such agents cannot leave the virtual freeway once set into it.

3.3.3 Implementation Details and Data Structures

`SAD` is an application written in C++, optionally using OpenGL for visualization. It does not rely on any other external libraries, and is thus completely platform independent.[2] It can be controlled using the command line, but also provides means for interactive interaction during runtime. See the Appendix for a list of options and keymappings.

The data structures heavily rely on the C++ Standard Template Library (STL), and are defined as follows:

The Emotion System, `class CEmotion`, is a class containing a variety of `std::maps` for storing the emotion-, hormone-, and input values.

The Vehicle, `class CVehicle`, is an abstract class for the classes `CVehicleReactive` and `CVehicleEmotional`. All agent independent code is implemented within this class, such as routines for movement or for performing lane changes. It further keeps all state variables such as position or velocity inside arrays, storing values for n time steps; as such, access of a variable in time step t is accessing element $t \bmod n$. A Vehicle is given pointers to all relevant lanes; the lane it is driving on, the left and the right lane. These are used to make queries regarding the vehicle's surrounding, e.g., to "see" adjacent vehicles.

The Lane, `class CLane`, inherits from `std::vector<CVehicle*>`. A vehicle entering the lane is pushed to the back of the vector, as such, the order of vehicles is inherent to the data structure.

The Highway, `class CHighway`, inherits from `std::map<int,CLane*>`. As the highway is simply a map of lanes, it is possible to add an arbitrary amount of lanes.

In Procedure 3.3.1, the vehicle's states (and the vehicle's agent's states, respectively) are calculated. Each calculation in timestep t makes use of values (position, velocity, etc.) of timestep t, results are stored for timestep $t + 1$. Thus, the results of the calculations in timestep t do not take effect until all vehicles have been processed.

The architecture supports different types of agents, described next.

[2] As far as C++ with the STL library is available on the platform in question.

p	Behaviour	Comment
> 1	altruistic	
$(0, 0.5]$	realistic	Advantages of other drivers have lower priority, but are not neglected.
0	purely selfish	The safety criterion is not ignored, though.
< 0	malicious	Agent takes "pleasure" in thwarting other drivers, even at the cost of own disadvantages. Again, the safety criterion is not ignored.

Table 3.3: Effect of the politeness factor p on the driving behaviour (Treiber 2007).

Procedure 3.3.1: PROCESSHIGHWAY( )

```
comment: The simulation's main loop
main
 for each t ∈ timesteps
   for each λ ∈ lanes
     for each α ∈ vehicles_λ
       PerformLaneChange(α, λ, t)
   for each λ ∈ lanes
     for each α ∈ vehicles_λ
       do { ProcessVehicle(α, λ, t)
          { UpdateStatistics(α, λ, t)
   for each λ ∈ Lanes
     do UpdateLaneStatistics(λ, t)
   UpdateSimStatistics(t)
   Draw(t)
 WriteStatistics( )
```

Procedure 3.3.2: PROCESSVEHICLE[REACTIVE](t)

```
comment: A reactive Vehicle
α_la, α_lb ← GetNearestCarsOnLane(x_α, leftlane)
α_ra, α_rb ← GetNearestCarsOnLane(x_α, rightlane)
IDMAccleration(t, α_la, α_lb, α_ra, α_rb)
```

Procedure 3.3.3: PERFORMLANECHANGE[REACTIVE](t)

comment: LaneChange of a reactive Vehicle

if IDMLANECHANGE($LeftLane, t$)
then Change to left lane

else if IDMLANECHANGE($RightLane, t$)
then Change to right lane

3.3.4 IDM agent

This reactive agent's behaviour is determined by physically grounded calculations, which were presented in Sections 3.2.2 and Section 3.2.4. The most dominant way of altering the agent's behaviour and therefore its actions, is by changing the politeness factor p. The influence of the politeness factor on driving behaviour is depicted in Table 3.3.

3.3.5 Purely emotional agent

The purely emotional agent does not rely on physically grounded calculations to steer the vehicle, rather, every action in timestep t is based on the agent's most dominant emotion in timestep t. The emotions are influenced by certain stimuli, such as *speed* or *acceleration* and so forth. However, these stimuli are the result of the physically grounded perception of environmental properties. These stimuli and their calculation are elucidated below.

Acceleration: The subjective perception of acceleration, P_{acc}, is defined as

$$P_{\text{acc}} = \frac{v'}{\gamma}, \tag{3.20}$$

where

$$v' = v_\alpha(t) - v_\alpha(t-1) \tag{3.21}$$

is the difference in speed over the last time step, weighted with either the vehicle's maximum acceleration $a_{\max,\alpha}$ if the difference is positive, or with the maximum deceleration $b_{\max,\alpha}$ otherwise. This is denoted as

$$\gamma = \begin{cases} a_{\max,\alpha} & \text{if } v' > 0, \\ b_{\max,\alpha} & \text{otherwise.} \end{cases} \tag{3.22}$$

A higher value of P_{acc} yields a higher perceived acceleration.

Speed: The subjective perception of speed, P_{speed}, is the ratio of the current speed, and the speed the driver would like to achieve. This is given by

$$P_{\text{speed}} = 2 \cdot \frac{v_\alpha(t)}{\max(v_\alpha(t), v_{\text{d},\alpha})} - 1, \tag{3.23}$$

in order to be confined within an interval of $[-1, 1]$. For instance, a velocity of 0 results in the perceived value of -1, the maximum possible speed as well as the maximum desired speed result in a value of 1.

A higher P_{speed} yields a the higher perceived speed.

Approach of: The perception of an approaching vehicle $\alpha + 1$, $P_{\text{app}_{\text{of}}}$, is mainly influenced by the estimated braking distance of vehicle $\alpha + 1$ in relation to α.

To a driver who is moving himself, the relative braking distance s_{rel} is of interest, which, in view of Equation 3.3, is calculated by the formula

$$s_{\text{rel}} = \frac{(v_{\alpha+1} - v_\alpha)^2}{2 \cdot b_{\text{max},\alpha+1}}. \tag{3.24}$$

Then $P_{\text{app}_{\text{of}}}$ is given by

$$P_{\text{app}_{\text{of}}} = \text{Th}_{[-1,1]}(\theta_{\alpha+1} \cdot (s_\alpha - s_{\text{rel}} - s_{\text{g},\alpha})), \tag{3.25}$$

where $\theta_{\alpha+1}$ weights the result, influenced by the following vehicle, and $\text{Th}_{[-1,1]}$ is as defined in Equation 2.10. In the current implementation, θ is a constant ($\theta_{\alpha+1} = -\frac{1}{10}$).

With a higher value of $P_{\text{app}_{\text{of}}}$, the closer the following vehicle is subjectively felt.

Approach to: The perception of the approach to a leading vehicle $\alpha - 1$, $P_{\text{app}_{\text{to}}}$, is the single most important stimulus while driving. Similar to $P_{\text{app}_{\text{of}}}$, the perception is influenced by the estimated braking distance. However, this time, the driver is required to cope with the possibility of a complete standstill of the leading vehicle. As it is vital to be prepared for sudden standstill, in contrast to $P_{\text{app}_{\text{of}}}$, the relative speed is not to be taken into account, but rather the speed of α. To make foresighted driving possible at all, the perceived closeness is the ratio of the needed braking distance and the available braking distance, rather than the difference of both values as applied in Equation 3.25. This makes it possible for the driver to adjust the vehicle's speed relatively early. We calculate $P_{\text{app}_{\text{to}}}$ by

$$P_{\text{app}_{\text{to}}} = \text{Th}_{[-2,2]}(\frac{\phi \cdot s_{\text{abs}}}{s_\alpha} - 1), \tag{3.26}$$

where the braking distance s_{abs} is very conservatively defined as

$$s_{\text{abs}} = \frac{v_\alpha^2}{2 \cdot b_{\text{max},\alpha}} + s_{\text{g},\alpha}, \tag{3.27}$$

and where ϕ is a factor to enlarge the perception of the required braking distance, to further encourage foresighted driving. In the current implementation, $\phi = 3$.

With a higher value of $P_{\text{app}_{\text{to}}}$, the closer the leading vehicle is subjectively felt.

Unrestrictedness left: The perception of the spacial freedom on the left lane, $P_{\text{unrest}_{\text{l}}}$, has to take two factors into account:

- the braking distance to vehicle $\alpha - 1$ on the left lane (denoted as λ in Figure 3.2), and

- the braking distance s_{back} vehicle $\alpha+1$ on the left lane (denoted as δ in Figure 3.2) would require after a potential lane change.

As vehicle $\alpha-1$ might come to a sudden standstill right after the lane change, both distances will have to be calculated using the absolute velocity. Let the vehicle most restricting to α be determinant for the perception of unrestrictedness, then

$$P_{\text{unrest}_l} = \text{Th}_{[-1,1]}[\max(\frac{\phi \cdot s_{\text{abs}}}{s_\alpha}, \frac{\phi \cdot s_{\text{back}}}{s_{\alpha+1}}) - 1], \tag{3.28}$$

where s_{abs} and ϕ are as above and

$$s_{\text{back}} = \frac{v_{\alpha+1}^2}{2 \cdot b_{\max,\alpha+1}} + s_{g,\alpha}. \tag{3.29}$$

A higher value of P_{unrest_l} yields more freedom for the driver.

Unrestrictedness right: The perception of the spacial freedom on the right lane, P_{unrest_r} is calculated in exactly the same way as P_{unrest_l}, with $\alpha+1$ and $\alpha-1$ being cars driving on the right lane.

Success: The perception of success, P_{succ}, takes two aspects into account:

- whether the desire at time $t-1$ to change the lane, $D_{\text{LC}}(t) \in \{0,1\}$, has been successfully satisfied in time t, i.e., a lane change was performed, and
- whether the desired velocity $v_{d,\alpha}$ reached.

Let $\xi_\alpha(t) \in \{0,\dots,n\}$ be the lane number vehicle α is driving on at time t, 0 being the rightmost of $n \in \mathbb{N}$ ascendingly numbered lanes, then

$$P_{\text{succ}} = \frac{v_\alpha}{v_{d,\alpha}} - D_{\text{LC}}(t-1) + |\xi_\alpha(t-1) - \xi_\alpha(t)|. \tag{3.30}$$

A higher value of P_{succ} yields a more successful driver. We have $P_{\text{succ}} = -1$ if the vehicle is standing still and the plan to change the lane did not work out.

Law Abiding: The perception of behaving in a lawful manner, P_{law}, could be influenced by a variety of factors:

- the law to drive on the right,
- the law to not overtake on the right side, except when driving in convoy,
- adherence of speed limits,
- respecting other drivers, i.e., not to squeeze into insufficiently large gaps.

However, in the current implementation, only the adherence to the speed limit is taken into account. Let $v_{\max}$ be the maximum allowed speed on the lane, then

$$P_{\text{law}} = \text{Th}_{[-1,1]}(\frac{v_{\max} - v_{d,\alpha}}{v_{\max}} \cdot 10), \tag{3.31}$$

where the multiplication with 10 causes an exceeding of 10% of the allowed speed to be counted as maximum unlawful behaviour, i.e., $P_{\text{law}} = -1$. With a higher value of P_{law}, the more law abiding the driver is.

As mentioned before, the agent's actions are not direct results of physically grounded calculations, but rather its dominant emotion, as reflected by Procedure 3.3.4. Similar to the acceleration and deceleration actions, the agent's lane changing behaviour too relies on emotions exclusively, as expressed by Procedure 3.3.5. The actions which result from the elicitation of each of the defined emotions—happiness, sadness, fear, and anger—are described in Table 3.4 in greater detail.

Procedure 3.3.4: PROCESSVEHICLE[EMOTIONAL](t)

comment: A purely emotional Vehicle

$\textsc{ProcessEmotions}(P_{\text{acc}}, P_{\text{speed}}, P_{\text{app}_{\text{of}}}, P_{\text{app}_{\text{to}}}, P_{\text{unrest}_{\text{l}}}, P_{\text{unrest}_{\text{r}}}, P_{\text{succ}}, P_{\text{law}}, t)$
$\epsilon \leftarrow \textsc{GetDominantEmotion}(t)$
$\sigma \leftarrow \textsc{GetEmotionIntensity}(\epsilon, t)$

$$do \begin{cases} \textsc{SetSpeed}(min(v_{max}, v_\alpha(t) + a_{\text{ed},\alpha} \cdot \sigma)) & \text{if } \epsilon = \text{happiness} \\ \textsc{SetSpeed}(max(0, v_\alpha(t) - b_{\text{ed},\alpha} \cdot \sigma)) & \text{if } \epsilon = \text{fear} \\ \textsc{SetSpeed}(min(v_{vehmax}, v_\alpha(t) + a_{\text{ed},\alpha} \cdot \sigma)),\ D_{\text{LC}}(t) \leftarrow 1 & \text{if } \epsilon = \text{anger} \\ \textsc{SetSpeed}(max(0, v_\alpha(t) - b_{\text{ed},\alpha} \cdot \frac{\sigma}{2})),\ D_{\text{LC}}(t) \leftarrow 1 & \text{if } \epsilon = \text{sadness} \end{cases}$$

Procedure 3.3.5: PERFORMLANECHANGE[EMOTIONAL](t)

comment: Lane change of a purely emotional vehicle

$\epsilon \leftarrow \textsc{GetDominantEmotion}(t-1)$
if $D_{\text{LC}}(t-1) = 1$

$$\textbf{then } do \begin{cases} \textsc{ChangeLane}(left) & \text{if } \epsilon = \text{anger} \\ \textsc{ChangeLane}(right) & \text{if } \epsilon = \text{sadness} \end{cases}$$

3.3.6 Runtime Analysis

For each step and each lane, each of the n vehicles is making its move. Any queries regarding a vehicle's state, such as its position or velocity, are performed in O(1). Thus, measuring the position and velocity of the front and back vehicles, and adjusting its own speed, is done in O(1).

In each vehicle step, a search for the nearest neighbors on the left and right lane are performed. As the lane is a `std::vector`, element access is performed in O(1). Thus, using binary search, the upper bound to the vehicle's current position is found in a lane with m vehicles within $O(log(m))$, that is $O(2log(n))$ for both neighboring lanes considering that $m < n$.

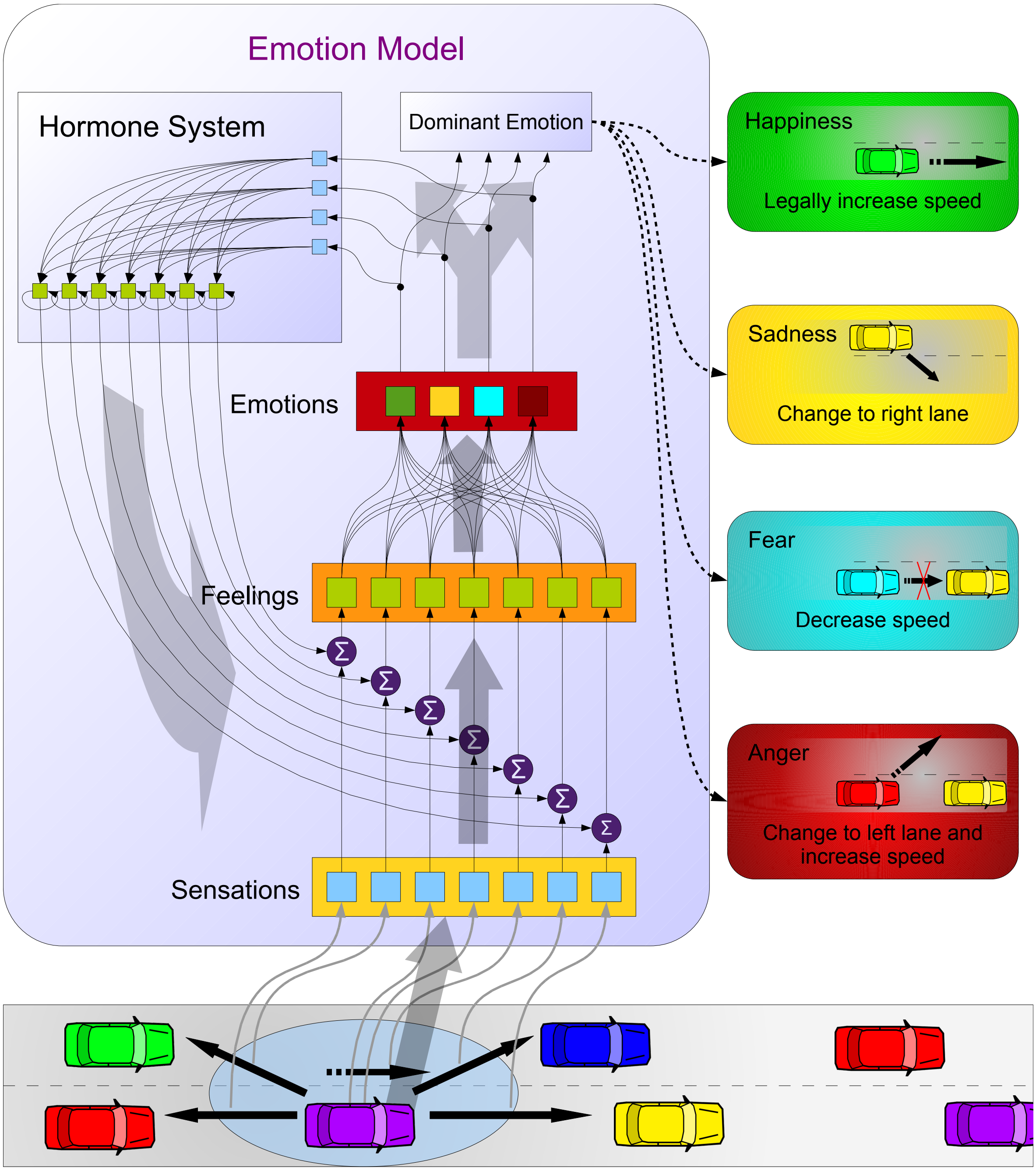

Figure 3.3: SAD driving

Emotion	Actions	Description
Happiness	Legally **increase speed**	A happy, satisfied driver is supposed to act within the law. A feeling of relaxedness and safety makes it possible to gain a desired traveling speed.
Sadness	Slightly **decrease speed**, change to **right lane**	Sadness is herein understood as a combination of guilt and boredom. It is usually invoked when the driver keeps on a lane, although the lane to the right side is free to go. This favors the law to drive on the right, as a sad driver will change to the right lane, whenever possible. Further, speed is reduced, as this seems a natural reaction while being in such a mood. Further research is needed to see whether this speed reduction really serves as a useful control mechanism.
Fear	**Reduce speed**	Fear, an emotion with a broad diversity of reaction in living creatures, is reduced to a single action in this simulation: reduction of speed. It is thus an important means to avoid rear-ending and makes it possible to slow down or even stop when the front traffic makes it necessary.
Anger	Change to **right lane**, **increase speed** while potentially ignoring laws	Anger leads to aggressive driving (Mesken 2006), and thus in increased speed and reckless lane changing behaviour. However, it serves as a vital control mechanism, as it enables the overtaking of other drivers. Moderate anger is thus important for a steady traffic flow. Entirely happy, sad, or frightened drivers would not feel the inclination to perform such a maneuver.

Table 3.4: Emotional control in `SAD`. This table describes the used emotions and their corresponding behaviour. It should be noted that these emotion names are not meant to reflect their natural meaning in their entirety.

When a vehicle crosses the line where the lane's end and beginning are connected, it is removed from the beginning of the vector using `erase` and pushed to the back of the vector with `push_back`. `erase` of one element performs linear to the amount of elements after deleted elements. Thus, removal of the first element is performed in $O(n)$, and `push_back` in constant time. At an average speed of v_{avg} and a lane length of L_Π, a vehicle crosses the border every $\frac{L_\Pi}{v_{avg}}$ steps, thus an average of $\frac{v_{avg}n}{L_\Pi}$ vehicles need to be erased with $O(n)$ complexity every step, leading to an unfortunate theoretical complexity of $O(\frac{v_{avg}}{L_\Pi}n^2)$. It should be noted though, that in realistic simulation runs[3] $\frac{v_{avg}n}{L_\Pi} < 1$, which practically equals to $O(n)$. As such, the movements of n vehicles have a theoretical worst case complexity of $O(n\ (log(n) + n))$, and

[3] A simulation run with 120 IDM drivers on a length of 2km for 10000 seconds results in a mean of 0.163 border crossings per second per lane. A test run with emotional drivers of personality *disciplined* exhibited a mean of 0.135 border crossings per second per lane.

best complexity of $O(n\ (log(n) + 1))$.

In each step, a vehicle can perform a lane change. A lane change requires the vehicle to be erased from its source lane, and inserted into the target lane with `insert`. As stated above, in the worst case, `erase` is $O(n)$, and `insert` behaves similarly, resulting in $O(2n)$ for the lane change of one vehicle. In the unlikely scenario that all vehicles change their lane in one step, the overall theoretical complexity would be $O(n^2)$. However, the average amount of lane changes per step in realistic simulation runs, as evident in Table 3.9, exhibit a mean of around 0.9 lane changes per second, reducing the complexity to $O(n)$ per step.

The worst case complexity for the simulation with n vehicles therefore is:

$$O(n\ (log(n) + n) + n^2) = O(2n^2 + n\ log(n)) = O(n^2 + n\ log(n)) = O(n^2), \tag{3.32}$$

while the best case complexity is

$$O(n\ (log(n) + 1) + n) = O(n\ log(n) + 2n) = O(n\ (log(n) + 1)) = O(n\ log(n)). \tag{3.33}$$

3.4 Empirical Evaluation

3.4.1 Used Measurements

For comparing the IDM and emotional agent architectures, including the differing personalities within the emotional domain, a variety of quality measurements were taken into account. These are described in what follows.

Traffic Flow Rate: According to Smith and Ulmer (2003), the traffic flow rate, the equivalent hourly rate of vehicles passing a point during a given time interval $\mathcal{I}$, is a fundamental measure of traffic status that plays a critical role in many aspects of transportation engineering. Thus, this rate is taken as one of the main criteria for the ranking procedure. In the literature, a minimum of 15 minute measurement intervals is recommended. Smith and Ulmer (2003), however, reveal that intervals as short as 10 minutes suffice for the calculation of flow rates. On real freeways, this rate is measured using induction loops detectors. Double induction loop detectors are also capable of measuring the vehicles' speed (Treiber et al. 2000).

Let $\mathit{outflow}_{\lambda,t}$ denote the number of vehicles that have passed the checkpoint, e.g., a loop detector, on lane λ during the time interval $[t-1, t]$. The traffic flow rate R_f, determines the number of vehicles that passed the point of measurement during the specified interval $\mathcal{I}$, can be defined as

$$R_f = \frac{1}{\mathcal{I}} \sum_{t=1}^{\mathcal{I}} \sum_{\lambda=1}^{\mathcal{L}} \mathit{outflow}_{\lambda,t}, \tag{3.34}$$

where $\mathcal{L}$ denotes the number of lanes on the highway.

Knowing the length L_Π of the highway, the sum of distances driven is given by

$$d_\Sigma = R_f \cdot L_\Pi + \epsilon. \tag{3.35}$$

The distances of vehicles not yet having passed that point are disregarded by that measurement technique. These distances are denoted as ϵ.

As it is possible to query the exact distances driven by every vehicle in a computer simulation, this measurement can be replaced by more efficient ones, as described below.

Summation of covered distances: The sum of the distance driven by all cars over a certain time span gives insight over the throughput of the highway. The higher the throughput, the more efficient the way of driving. Let d_Σ be the sum of driven distances, then

$$d_\Sigma = \sum_{i=1}^{\mathcal{V}} d_{\alpha_i}, \tag{3.36}$$

where $\mathcal{V}$ denotes the total number of vehicles on the highway. Thus, the average distance a vehicle is driving is given by

$$\bar{d} = \frac{d_\Sigma}{\mathcal{V}}. \tag{3.37}$$

Average Speed: The average speed of all cars (using the arithmetic mean) during a specified time interval $\mathcal{I}$. $\bar{v}$ gives an additional insight over the throughput of the highway. This parameter is defined as

$$\bar{v} = \frac{1}{\mathcal{V}} \sum_{i=1}^{\mathcal{V}} \left[\frac{1}{\mathcal{I} - t_i} \sum_{t=t_i}^{\mathcal{I}} v_{\alpha_i}(t) \right], \tag{3.38}$$

where t_i denotes the point in time where vehicle i was entering the highway.

Δ Speed: The average sum of all changes in speed over a specified time interval, $\overline{\Delta v}$ reveals how often the drivers are forced to change their speed. Under the assumption that acceleration results in higher fuel consumption, and that braking has a certain negative impact on the tires and thus leading to the emission of health threatening particulate matter, it can be assumed that the less this value is, the higher the ecological efficiency of the collective driving behaviour is. This assumption is backed up by Ardekani et al. (2002), who state that the main variables in fuel consumption include speed, number of stops, speed noise, and acceleration noise. $\overline{\Delta v}$ can then be computed as follows:

$$\overline{\Delta v} = \frac{1}{\mathcal{V}} \sum_{i=1}^{\mathcal{V}} \left[\frac{1}{\mathcal{I} - t_i} \sum_{t=t_i}^{\mathcal{I}} |v_{\alpha_i}(t+1) - v_{\alpha_i}(t)| \right]. \tag{3.39}$$

Number of Lane Changes: As defined in Section 3.3.5, let $\xi_\alpha(t) \in \{0, \ldots, n\}$ be the lane number vehicle α is driving on at time t and let 0 be the rightmost of $n \in \mathbb{N}$ ascendingly numbered lanes. Then the number of lane changes is given by

$$C_\Sigma = \sum_{t=1}^{\mathcal{I}} \sum_{i=1}^{\mathcal{V}} |\xi_i(t-1) - \xi_i(t)|. \tag{3.40}$$

For a better understanding of the lane changing behaviour of different agent types, the average amount of lane changes per simulation run, $\overline{C}$, defined as

$$\overline{C} = \frac{1}{\mathcal{V}} \sum_{i=1}^{\mathcal{V}} \left[\frac{1}{\mathcal{I} - t_i} \sum_{t=t_i}^{\mathcal{I}} |\xi_i(t-1) - \xi_i(t)| \right], \tag{3.41}$$

can be utilized.

Average Lane Height: The average lane height, $\overline{\xi}$, gives, among other things, insight into the right driving rule behaviour of the drivers, and is defined as

$$\overline{\xi} = \frac{1}{\mathcal{V}} \sum_{i=1}^{\mathcal{V}} \left[\frac{1}{\mathcal{I} - t_i} \sum_{t=t_i}^{\mathcal{I}} \xi_i(t) \right]. \tag{3.42}$$

3.4.2 Tests

As previously mentioned, `SAD` was developed in order to explore synthetic emotions in an intuitive way. These were supposed to be tested in mainly three regards. The first goal was to find out if emotions alone suffice to drive a vehicle, without the help of deductions or reasoning. This found its motivation in various insights delivered by various researchers, such as Minsky (1988) and Brooks (1991). Technical inspiration came from a variety of sources, mainly from the field of affective agents in surviving tasks, e.g., by Gadanho and Hallam (2001), Velásquez (1997), or Scheutz (2004). Gadanho and Hallam (2001) state that their "research led to the conclusion that artificial emotions are a useful construct to have in the domain of behaviour based autonomous agents with multiple goals and faced with an unstructured environment". This is certainly not the case within a highway domain; it is highly structured and an agent only has a limited set of goals: advance forward in a timely manner and do not collide with other vehicles. Thus, this domain promised to be an interesting new field for research.

The second goal was to see whether artificial emotions increase the realism of agent behaviour. In simulations relying on the classical IDM, agents show extremely predictable behaviour. An IDM agent performs an action in time t based on various parameters in time $t-1$, and does not take into account events that occured prior to that. Resulting thereof are perfect patterns, which are unlikely to be found in real life. The `SAD` agents however, with their hormonal system, slowly build up certain moods which affect their driving behaviour. It was to be tested whether this leads to behaviour more realistic than the one emerging from the IDM agents.

The third goal was to determine the influence of emotions in regards to driving efficiency. Scheutz and Logan (2001) were able to show that in a setup of reactive and emotional agents competing for natural resources in a dangerous gridworld, emotional agents could outperform reactive agents by far. Thus, the question was whether emotional agents can outperform reactive agents in some of the stated measurements, e.g., whether the driving behaviour results in a higher overall throughput, produces less traffic jams, or leads to a cleaner virtual environment by not braking and accelerating in an exceeding manner.

In order to evaluate these questions, a series of tests were performed, where the IDM drivers act as a reference. The IDM model is the best driving model in traffic research as of now, as such it should suffice to compare the emotional drivers with the IDM drivers in order to gain insights regarding their efficiency. For the comparison of agent types, two different setups were built: a homogenous setup, Sim_{hom}, and a heterogenous one, Sim_{het}. Figure 3.4 depicts screenshots of two homogenous and one heterogenous simulation run with `SAD`. Sim_{hom} is used to evaluate the stand-alone quality of a type of agent in comparison to the stand-alone quality of another agent type. Sim_{het} is used to compare how a number of different agent types behave in direct competition to others. All setups start with an empty highway of 2000 meters, n agents enter the highway with a delay of d seconds. The simulations are run for $N = 10000$ s ≈ 166 min. The homogenous tests were performed with $n = 100$ agents, using an entrance delay of $d = 80$. The heterogenous tests were performed with $n = 80$ vehicles. The IDM driver and the four personalities are represented with 15 vehicles each. 50 drivers enter the road with a delay of $d = 16$, the remaining 30 are released in groups of 5 every 1000 seconds starting from $t = 1500$.

3.4.3 Characters

Four hand-crafted characters were tested in isolation as well as competing with the reactive IDM agents. These characters are defined through the coupling coefficients of their emotional system, as defined in Section 2.4.2. These coefficients are understood as their personality, and they define how and in which intensity the agents react to certain stimuli. These four hand-crafted personalities are the *normal driver*, the *aggressive driver*, the *fearful driver*, and the *disciplined driver*, and they were designed with the human counterparts in mind. A normal driver is supposed to drive like an average person without attracting attention. The aggressive driver is supposed to be that pushy driver who seems to believe that he owns the highway. The fearful driver is intended to simulate a typical sunday driver, who is scared to overtake on a three-lane highway on his way to church. The disciplined driver is supposed to be the professional driver, behaving in the most effective way with a balance of egoistic and cooperative driving.

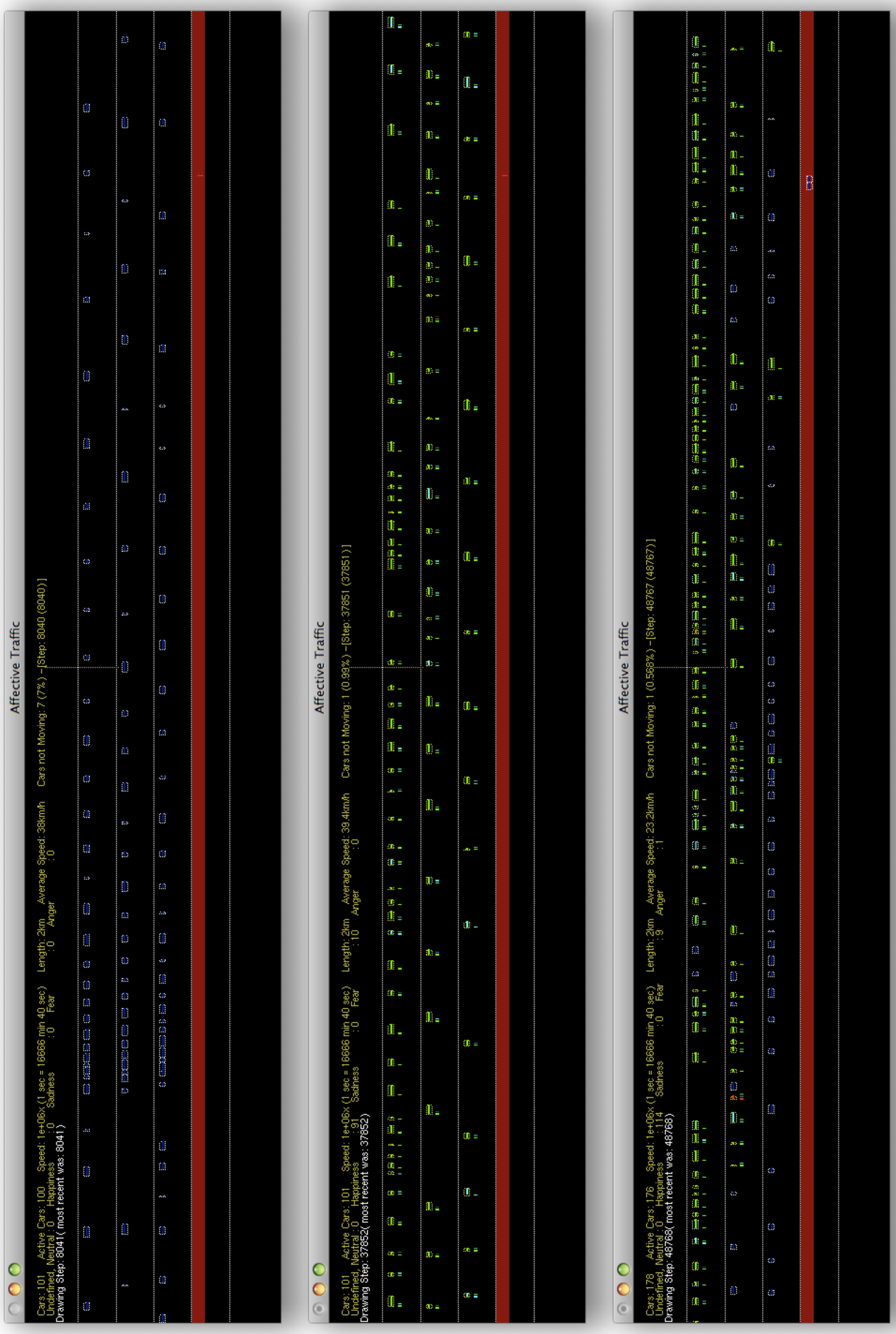

Figure 3.4: Simulation of (1) IDM drivers, (2) emotional drivers and (3) all together.

Feeling	Happiness	Sadness	Fear	Anger
Bias	0.1	0.1	0.1	0.2
Acceleration	0.1	-0.1	0.1	-0.1
Speed	0.3	-0.2	0.2	-0.6
Approach Of	-0.3	0.2	0.1	0.4
Approach To	-2.0	0.0	0.6	1.0
Unrestricted Left	0.0	-0.3	-0.7	0.4
Unrestricted Right	-0.3	1.0	1.0	-0.2
Success	0.6	-0.3	-0.1	-0.3
Lawabiding	0.5	-0.2	-0.3	0.0

Table 3.5: Emotion dependencies of a normal driver.

3.4.3.1 The normal driver

The normal driver's personality was constructed with a balanced emotional behaviour in mind. It responds rather evanescent to acceleration stimuli, is mildly gaining a good mood when keeping a steady speed and can further reduce anger hereby. This is supposed to lead to a moderate driving behaviour, preventing constant left-driving or unnecessary speed enhancement. The agent responds only to a minor degree to vehicles approaching from the back, but acts rather distinct when approaching another vehicle by gaining fear and anger while losing happiness. It loses fear when the left lane is free, as such it has a tendency to overtake—if possible–when an obstacle is in sight. Success and an law-abiding behaviour both lead to happiness. The exact emotion dependencies and the emotional bias are depicted in Table 3.5.

Figure 3.5 shows the dominant emotion of each step of a simulation with 100 normal drivers over a period of 10000 seconds. Beginning from time 0, a new car is entering the lane every 80 seconds. Figure 3.6 depicts the distribution of the driver's emotions over the whole simulation. It can be seen that the driver keeps distance to front vehicles which can be seen by high amount

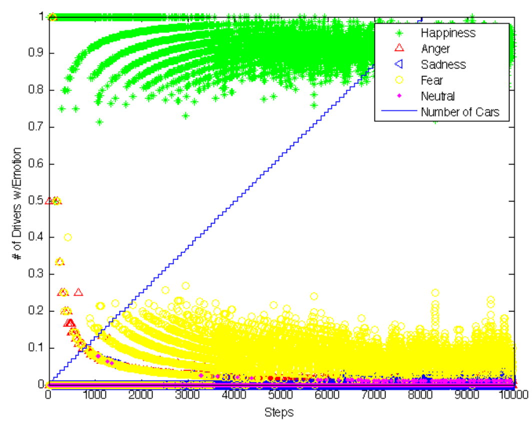

Figure 3.5: Emotional changes of a normal driver during the simulation.

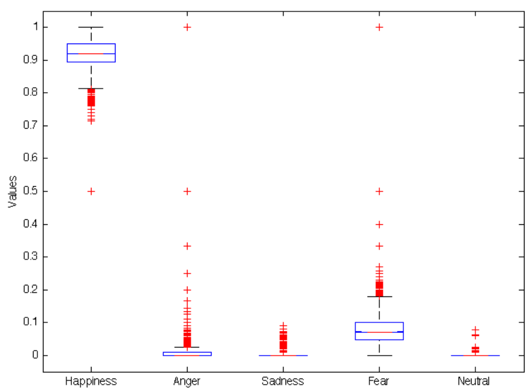

Figure 3.6: Distribution of all emotions of a normal driver during the simuluation.

	Happiness	Sadness	Fear	Anger
Bias	0.1	0.1	0.1	0.1
Acceleration	0.9	0.0	0.0	0.0
Speed	0.7	0.0	0.0	-0.2
Approach Of	0.5	0.4	0.0	0.5
Approach To	-0.5	0.1	1.0	2.0
Unrestricted Left	0.5	-1.0	-1.0	2.0
Unrestricted Right	0.2	0.4	0.0	-0.3
Success	1.0	0.0	0.0	0.0
Lawabiding	0.0	0.0	0.0	0.0

Table 3.6: Emotion dependencies of an aggressive driver.

of fear experienced. The x-axis in Figure 3.5 represents the number of steps performed. The y-axis depicts the percentage of drivers experiencing a certain emotion, where 1 = 100%.

3.4.3.2 The aggressive driver

Table 3.6 depicts the aggressive driver's personality, which was constructed with an unsocial and unlawful behaviour in mind. It responds amplifying to acceleration stimuli and "enjoys" speed. It responds with anger to drivers in front, and is thus inclined to overtake rather than brake, if possible. His mood is not affected by an empty right-hand lane, as such the chance of adhering to the right-driving rule is small, yet existant. The agent responds aggressively to vehicles approaching from the back, leading to a further increase in the wish for speed. Law-abiding behaviour is a non-issue to the aggressive driver.

Figure 3.7 shows the dominant emotion of each step of the simulation. As clearly visible, high traffic leads to a strong increase in anger, which is also visible in Figure 3.8. This is

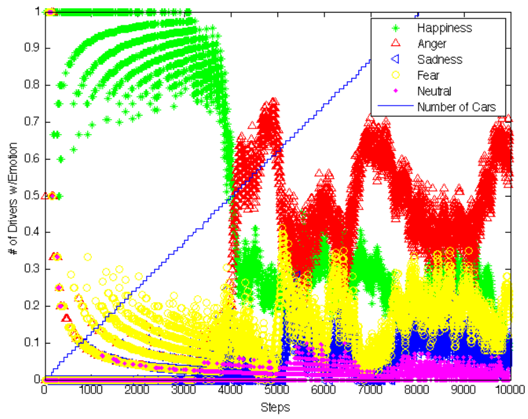

Figure 3.7: Emotional changes of an aggressive driver during the simulation.

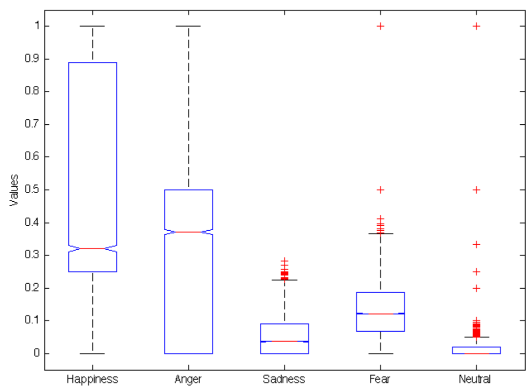

Figure 3.8: Distribution of all emotions of an aggressive driver during the simulua-tion.

	Happiness	Sadness	Fear	Anger
Bias	0.1	0.2	0.1	0.1
Acceleration	0.1	-0.1	0.5	-0.1
Speed	0.1	-0.1	0.3	-0.3
Approach Of	-0.2	1.0	1.0	0.1
Approach To	-1.0	0.2	2.0	1.0
Unrestricted Left	-0.5	0.1	0.0	0.5
Unrestricted Right	-0.1	2.0	0.3	0.0
Success	0.5	-0.4	-0.1	-0.1
Lawabiding	0.7	-0.2	-0.1	-0.1

Table 3.7: Emotion dependencies of fearful driver.

evident beginning at step $t = 4000$ when the traffic collapses, which is can be clearly seen by the high amount of negative emotions present. The aggressive driver behaves as one would expect: uncooperative, and as such incompetent to cope with situations where everybody behaves equivalently ignorant.

3.4.3.3 The fearful driver

The fearful driver's personality, as depicted in Table 3.7, is constructed to model an overcautious, frightened driver who respects the law more than necessary. It responds with fear to speed and acceleration, and rather brakes than overtake when approaching another car. It obeys the right-driving rule very strictly by having a strong bias on sadness in regards to an empty right-hand lane. Accordingly, it will reduce speed and try to change to the right lane when a vehicle is approaching from behind. Further, law abiding behaviour is a rewarding factor to this personality. Figure 3.9 and Figure 3.10 show that the fearful driver suits his name. He performs better than the aggressive driver—the traffic collapses not until $t = 6000$.

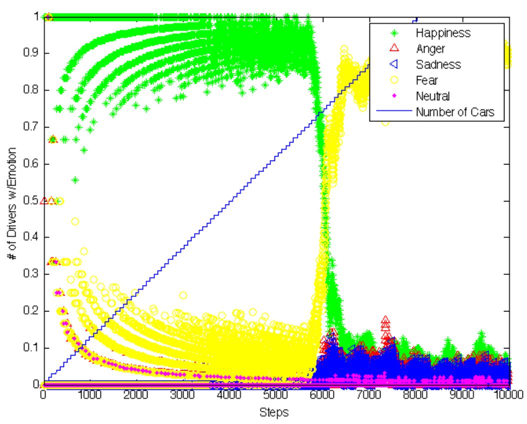

Figure 3.9: Emotional changes of a fearful driver during the simulation.

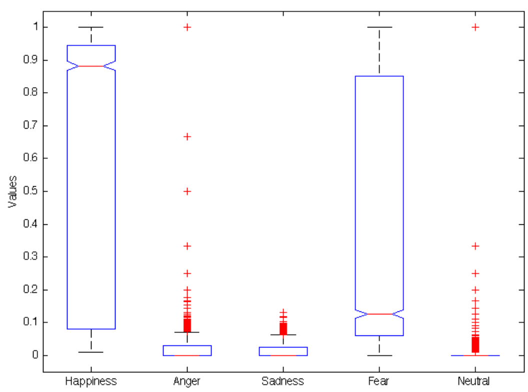

Figure 3.10: Distribution of all emotions of a fearful driver during the simuation.

	Happiness	Sadness	Fear	Anger
Bias	0.1	0.1	0.1	0.2
Acceleration	0.1	-0.1	0.1	-0.1
Speed	0.3	-0.2	0.1	-0.6
Approach Of	-0.3	0.5	0.0	0.4
Approach To	-2.0	0.0	0.6	1.0
Unrestricted Left	0.0	-0.3	-0.3	0.9
Unrestricted Right	-0.3	1.0	-0.3	-0.2
Success	0.6	-0.3	-0.0	-0.3
Lawabiding	0.5	-0.2	-0.0	0.0

Table 3.8: Emotion dependencies of a disciplined driver.

However, it is nevertheless evident, that overly fearful driving is similarly inefficient as overly aggressive driving. As clearly visible, the fearful driver responds with immense fear in high traffic densities.

3.4.3.4 The disciplined driver

The disciplined driver's personality was constructed with optimal driving behaviour in mind, which is depicted in Table 3.8. A driver with this personality appears emotionally to be more mature than the other drivers, as he responds in a reasonable way to most situations. He responds with sadness to approaching, i.e., faster vehicles, which encourages a better overall traffic flow. A free left lane increases anger, but not as much as an empty right lane increases sadness. Combined with the tendency to gain anger when approaching another car, this leads to a tendency of overtaking only when really necessary, and inhibits a lane change to the left without the need to do so. This, too, facilitates a better traffic flow.

Emotional changes during the simulation run and the emotion distribution are depicted

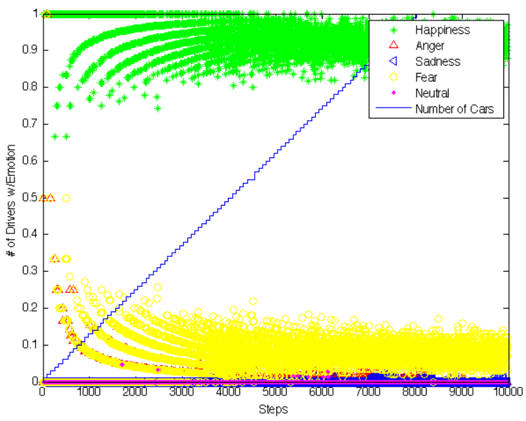

Figure 3.11: Emotional changes of a disciplined driver during the simulation.

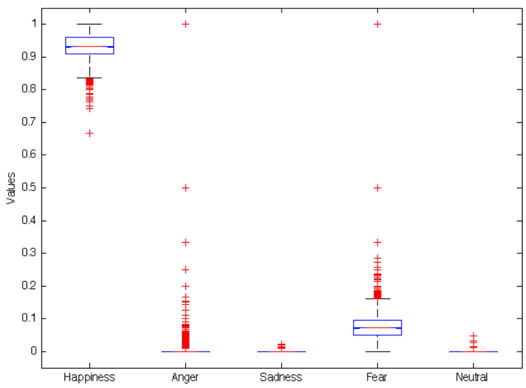

Figure 3.12: Distribution of all emotions of a disciplined driver during the sim.

Feature	Unit	IDM	Normal	Aggressive	Fearful	Disciplined
d_Σ	m	9838048	8142342	2646590	3169364	7827627
$\bar{v}$	m/s	18.646	17.492	8.938	9.359	15.974
$\overline{C}$	times	0.929	0.904	0.765	0.618	0.922
$\overline{\Delta v}$	m/s^2	0.155	0.298	0.276	0.316	0.325
$\bar{\xi}$	lanes	0.792	1.176	1.437	0.861	0.915

Table 3.9: Some results of Sim$_{hom}$.

in Figure 3.11 and Figure 3.12. It can be seen that the disciplined driver is emotionally stable regardless of traffic density. The high amount of fear indicates that he adapts his speed continuously, which is visible during simulation.

3.4.4 Discussion

To better visualize the course of events, the amount of cars is plotted into Figures 3.14 and 3.18. Note that it does not correspond to the scale on the y-axis.

3.4.4.1 Homogenous Setup

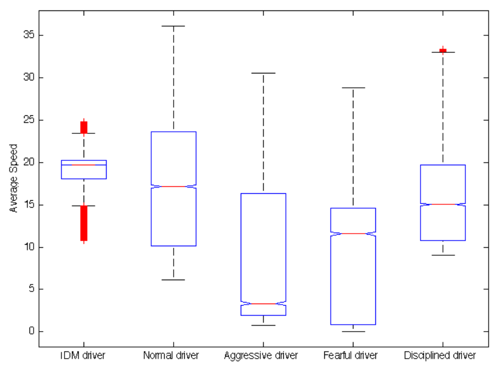

Figure 3.13: Boxplot of average speeds in Sim$_{hom}$.

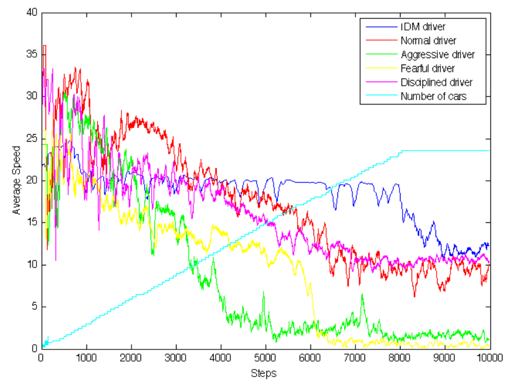

Figure 3.14: Average speeds over time in Sim$_{hom}$.

Figure 3.13 shows average speeds of the drivers within the homogenous setup. It indicates that the emotional drivers exhibit a notable variance in speed over the whole simulation run, while the IDM driver maintains a constant velocity, until the first traffic jam occurs, which is clearly visible in Figure 3.14 at around $t = 8000$. In regards to realism, this clearly speaks for the emotional agents. As depicted in Table 3.9, the overall driven distance, d_Σ, is $9838km$ for IDM agents, and $8142km$ for the—in this category—best emotional driver, almost 20% less. In combination with insights derived from Figure 3.14, namely that drivers with the normal personality keep a higher velocity than the IDM drivers until $t = 4000$, it can be stated that emotional drivers do not cope well with crowded highways. This is especially true for aggressive

Feature	Unit	IDM	Normal	Aggressive	Fearful	Disciplined
d_Σ	m	1447538	1102902	1032045	1215339	1222482
$\bar{v}$	m/s	12.328	9.898	9.334	10.411	10.578
$\overline{C}$	times	0.979	0.883	0.831	0.928	0.917
$\overline{\Delta v}$	m/s^2	0.064	0.317	0.251	0.215	0.334
$\bar{\xi}$	lanes	0.498	1.607	1.722	0.574	0.962

Table 3.10: Some results of Sim$_{het}$.

and fearful drivers, both cause traffic jams very early. While aggressive drivers block the road with merely around 50 drivers, fearful driving comes to an end with 70 vehicles on the road.

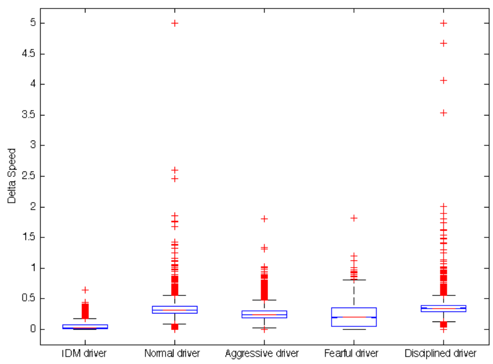

Figure 3.15: Average intensity of speed changes in Sim$_{hom}$.

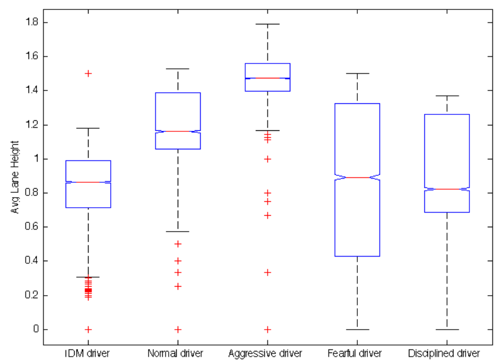

Figure 3.16: Average lane height of all drivers in Sim$_{hom}$.

In regards to ecological driving, Figure 3.15 shows that, again, IDM drivers perform better than their emotional competitors. Observations show that the emotional drivers utilize the maximum acceleration possible, and similarly brake with the maximum possible rate.

Figure 3.16 shows the drivers preferences in regards to lane height. Whereas IDM drivers seem to follow the right driving rule quite strictly, the normal and especially aggressive drivers do not seem to have a tendency to follow this regulation. Fearful drivers, on the other hand, do follow this law, and so does the disciplined driver, albeit less vigorously.

3.4.4.2 Heterogenous Setup

The most obvious conclusion from Figure 3.18 is, that the worst drivers dictate the performance of the highway. One can clearly see the shock effects taking place at $t = 1500$, $t = 2500$ and so forth, from which the drivers recover quite quickly, up until to that point, where the traffic collapses, at $t = 6000$, when the aforementioned 80 drivers have all entered the highway. Due to being restricted by the other drivers, the average speeds visible in Figure 3.17 do not show any significant differences, similarly only a small tendency can be seen in Figure 3.19.

However, the lane preferences depicted in Figure 3.20 show the same characteristics as were visible in Figure 3.16 of the homogenous setup.

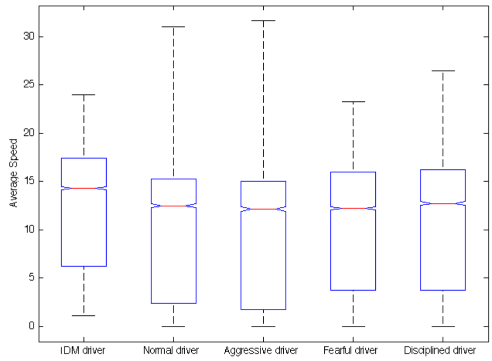

Figure 3.17: Boxplot of average speeds in Sim_{het}.

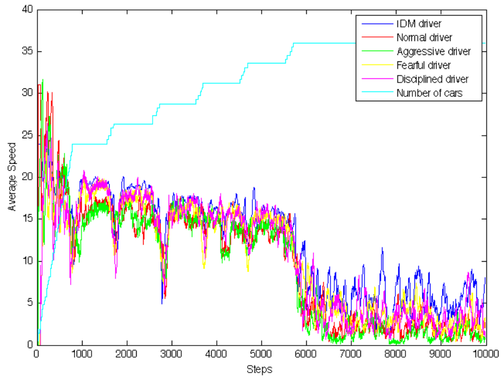

Figure 3.18: Average speeds over time in Sim_{het}.

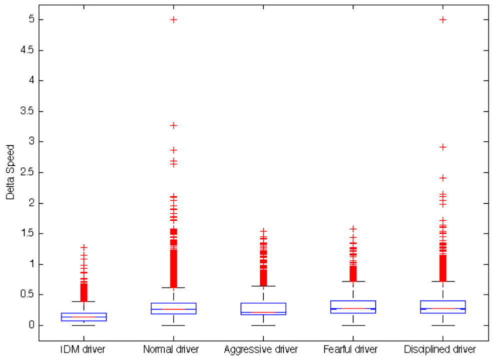

Figure 3.19: Average intensity of speed changes in Sim_{het}.

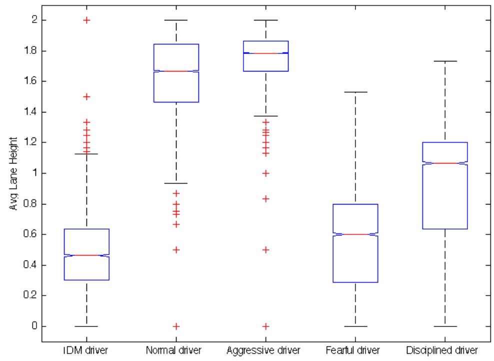

Figure 3.20: Average lane height of all drivers in Sim_{het}.

The results show that driving controlled only by emotions is indeed possible, it is even surprisingly efficient. Moreover, even though the level of abstraction of human emotions is so high, the observed behaviour—especially in regard to the different personalities—is surprisingly realistic. Further, the presented results are in accordance to findings in traffic psychology, which has shown that emotions, regardless of the type, have a negative effect on driving (Mesken 2006).

Chapter 4

Conclusion

Two questions are of main interest in regards to emotions and traffic: how are emotions elicited, and how do these emotions affect the driving behavior of an agent. Chapter 2 of this thesis provided some clues to answer the first question; it was defined how the word *emotion* is currently understood, and how underlying cognitive concepts are perceived by different fields of research, such as neuroscience and psychology. Indications were presented, that neither origin nor functionality of human emotions are yet fully understood, and the question of their benefits is still an ongoing dispute.

A variety of different pragmatic approaches were presented, which try, to some extent very successfully, to reach results without an extensive understanding of neurocognitive foundations. It has been shown that between mainly two approaches can be distinguished: phenomenological models and abstractions of biological systems. The latter approach was chosen to experimentally answer the second question asked above, namely how emotions affect driving behavior. Following recent results from neuroscience, specifically the somatic marker hypothesis, this approach is based on the assumption that emotions an agent experiences are elicited by sensorial stimuli as well as internal stimuli. A tested model was chosen and integrated into a traffic simulation, which was developed in the course of this thesis. This simulation, `SAD`, was developed to let driver agents compete with each other.

In the course of development and analysis of simulation runs, insights in regard to possible enhancements and refinements were gained. These could prove worthwhile in order to increase the realism of the agent behavior and thus the usefulness of the simulation.

One of the first insights gained was that it is a highly tedious task to find useful personality weightings that result in realistic driving behavior. The modification of only a single value can have a severe impact on the agent. I.e., if the stimulus of upcoming vehicles is weighted too strong towards fear, the agent might refuse to drive at all. This might also happen when these specific weightings fit, but other stimuli are not weighted sufficiently high, e.g., if the agent has no chance to build up sufficient happiness, although everything being fine. Then, a minor level of fear might be selected as dominant emotion and thus trigger a brake event, even though there was no danger on the lane. To solve this problem of unrealistic weightings automatically, a training function might be used to produce optimal driving personalities. An earlier approach to automatically generate personalities using a random approach failed: of 120.000 randomly generated characters, not a single one was driving similarly efficient as the hand-crafted ones. This training function could be constructed to utilize techniques known from neural networks, such as back-propagation.

The source of these undesirable phenomena might also lie within the fact that agents acting according to an emotion having no information as to why this emotion was elicited. E.g., an agent is supposed to feel fearful when the left lane is not free, respectively it already drives on the leftmost lane, and it is approaching another vehicle. This is due to the fact that—in this case—there are no other possible (legal) actions other than hitting the brake. If, however, the agent is fearful only because of the fact that the left lane is not a possibility to evade to in the thinkable scenario of another vehicle blocking the current lane, and even though there is no vehicle in sight and no other stimuli present, it might notwithstanding hit the brake. This, e.g., is the case with the fearful driver, when it comes to a standstill on the leftmost lane as a result of a traffic congestion. It will refuse to move forward due to being too frightened, even when the lane ahead is free. This is further amplified by drivers standing behind, which further adds to the agent being fearful, as it is frightened by approaching vehicles as well. Aspects of the OCC model, as discussed in Section 2.3.1.2, could be taken into account to solve this problem, and introduce a level of cognitive appraisal. This might support the perception of stimuli, by emphasizing objects and events that are of importance to the agent. On the other hand, stimuli of no importance could be ignored for the sake of improving the driving behavior, such as the still standing vehicle behind the agent's own vehicle. In the current implementation, this is not yet the case, as depicted in Equation 3.25. Further, the IDM could be incorporated into the emotional agent's perception, or vice-versa, the emotion system could serve as a bias to the IDM. Besides the potential positive effect this could have, it would further lead to a better comparability of both models. As of now, differences in driving behavior also arise from differences in calculation of distances and required gaps.

A shortcoming unsolved by both models, the emotional agents and the IDM—also in the implementation by Treiber (2006)—is that the drivers ignore the fact that it is forbidden to overtake another vehicle on the right side. The law-abiding feeling could be utilized for this in combination with a perception that checks whether a leading vehicle on the left lane is about to be overtaken.

Some added features might be interesting. E.g., a wider variety of vehicle parameters, such as fuel consumption could be simulated. This would be useful to measure the ecological differences in driving styles. Additionally, a model of damage would make accidents possible and thus further increase the realism and therefore the usefulness of the simulation. A wider range of insights could therewith be drawn from the simulation, i.e., the effects of particular emotions on accidents.

Bibliography

Hyungil Ahn and Rosalind W. Picard. Affective Cognitive Learning and Decision Making: The Role of Emotions. In *The 18th European Meeting on Cybernetics and Systems Research (EMCSR 2006). April 18-19, 2006, Vienna, Austria*, 2005.

Giambattista Amati, Claudio Carpineto, and Giovanni Romano, editors. *Advances in Information Retrieval, 29th European Conference on Ir Research, Ecir 2007, Rome, Italy, April 2-5, 2007, Proceedings*, volume 4425 of *Lecture Notes in Computer Science*, 2007. Springer.

Elisabeth André, Laila Dybkjær, Wolfgang Minker, and Paul Heisterkamp, editors. *Affective Dialogue Systems, Tutorial and Research Workshop, Ads 2004, Kloster Irsee, Germany, June 14-16, 2004, Proceedings*, volume 3068 of *Lecture Notes in Computer Science*, 2004. Springer.

Sia Ardekani, Ezra Hauer, and Bahram Jamei. Traffic Impact Models. In Nathan H. Gartner, Carrol J. Messer, and Ajay K. Rathi, editors, *Traffic Flow Theory - A State-of-the-Art Report: Revised Monograph on Traffic Flow Theory*. Turner Fairbank Highway Research Center (TFHRC) U.S. Department of Transportation, April 2002.

Ruth Aylett. Agents and Affect: Why Embodied Agents Need Affective Systems. In Carlos Delgado-Mata and Jesús Ibáñez, editors, *IVEVA*, volume 97 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2004.

Ruth Aylett, Joao Dias, and Ana Paiva. An Affectively Driven Planner for Synthetic Characters. In *Proceedings of the Sixteenth International Conference on Automated Planning and Scheduling (ICAPS 2006)*. AAAI, 2006.

Joschua Bach, Dietrich Dörner, and Ronnie Vuine. Psi and MicroPsi - a Novel Approach to Modeling Emotion and Cognition in a Cognitive Architecture. In *Seventh International Conference on Cognitive Modeling (ICCM 2006)*, Tutorial Notes, Trieste, Italy, EU, 2006. University of Trieste.

Eugene Ball. *A Bayesian Heart: Computer Recognition and Simulation of Emotion*, volume 1, pages 303–332. MIT Press, Cambridge, MA, USA, 2003.

J. Barcelí, E. Codina, J. Casas, J. L. Ferrer, and D. García. Microscopic Traffic Simulation: A Tool for the Design, Analysis and Evaluation of Intelligent Transport Systems. *J. Intell. Robotics Syst.*, 41(2-3):173–203, 2005.

Christoph Bartneck. Integrating the OCC Model of Emotions in Embodied Characters. In *Workshop on Virtual Conversational Characters: Applications, Methods, and Research Challenges*, 2002.

Luc P. Beaudoin and Aaron Sloman. A Study of Motive Processing and Attention. In Aaron Sloman, D. Hogg, G. Humphreys, A. Ramsay, and D. Partridge, editors, *Prospects for Artificial Intelligence: Proceedings of AISB-93*, pages 229–238. IOS Press, Amsterdam, 1993.

Antoine Bechara and Antonio R. Damasio. The Somatic Marker Hypothesis: A Neural Theory of Economic Decision. *Games and Economic Behavior*, 52(2):336–372, 2005.

Christian Becker, Stefan Kopp, and Ipke Wachsmuth. Simulating the Emotion Dynamics of a Multimodal Conversational Agent. In Elisabeth André, Laila Dybkjær, Wolfgang Minker, and Paul Heisterkamp, editors, *ADS*, volume 3068 of *Lecture Notes in Computer Science*, pages 154–165. Springer, 2004.

Kirstie L. Bellman. *Emotions: Meaningful Mappings between the Individual and Its World*, volume 1, pages 149–188. MIT Press, Cambridge, MA, USA, 2003.

V. Braitenberg. *Vehicles: Experiments in Synthetic Psychology.* MIT Press, Cambridge, MA, USA, 1984. A classic book on the synthesis of complex behavior from the interactions of simple components. See Dewdney, 1987.

Cynthia Breazeal. Emotion and Sociable Humanoid Robots. *International Journal Human-Computer Studies*, 59(1-2):119–155, 2003.

Rodney A. Brooks. Intelligence without Reason. In John Mylopoulos and Raymond Reiter, editors, *Proceedings of the 12th International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 569–595. Morgan Kaufmann, 1991.

Larry Cahill and James L. McGaugh. Mechanisms of Emotional Arousal and Lasting Declarative Memory. *Trends in Neurosciences*, 21:294–299, 1998.

Lola Dolores Cañamero. *Designing Emotions for Activity Selection in Autonomous Agents*, volume 1, pages 115–148. MIT Press, Cambridge, MA, USA, 2003.

Oliver Carsten. Safety is Not Necessarily Pleasant and Pleasantness is Not Necessarily Safe. In Wolf Dietrich Zuzan, editor, *Safety Evaluation of Traffic Systems: Traffic Conflicts and Other Measures, Proceedings of the 6th ICTCT Workshop*, page 123, Salzburg, Austria, October 1993. Kuratorium für Verkehrssicherheit.

David Charles. Emotion, Cognition and Action. In John Hyman and Helen Steward, editors, *Agency and Action*, pages 105–136. Cambridge University Press, 2004.

Philippe Chassy and Fernand Gobet. A Model of Emotional Influence on Memory Processing. In Lola Dolores Cañamero, editor, *Symposium on Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action: Proceedings of the AISB 2005 Convention. University of Hertforshire*, UK, 2005.

Robert E. Cochran, Frank J. Lee, and Eric Chown. Modeling Emotion: Arousal's Impact on Memory. In *proceedings of the 28th Annual Conference of the Cognitive Science Society*, pages 1133–1138, Vancouver, British Columbia, Canada., 2006.

E. Coutinho, E. Miranda, and A. Cangelosi. Towards a Model for Embodied Emotions. In Carlos Bento, Amílcar Cardoso, and Gaël Dias, editors, *Proceedings of EPIA2005, 2005 Portuguese Conference on Artificial Intelligence*, pages 54–63, Portugal, 2005. IEEE Press.

Antonio R. Damasio. A Second Chance for Emotion. In *Cognitive Neuroscience of Emotion (Series in Affective Science)*. Oxford University Press, USA, April 2002.

Antonio R. Damasio. *Descartes Error: Emotion, Reason and the Human Brain.* Gosset/Putnam Press, New York, 1994.

Charles Robert Darwin. *The Expression of the Emotions in Man and Animals.* John Murray, London, 1^{st} edition, 1872.

Richard J. Davidson, editor. *Handbook of Affective Sciences (Series in Affective Science).* Oxford University Press, USA, December 2002.

Darryl N. Davis. Minds Have Personalities - Emotion is the Core. In *Designing a Functioning Mind, AISB2000 - AI and Society.* University of Birmingham, 2000.

Carlos Delgado-Mata and Ruth Aylett. Do Virtual Sheep Smell Emotion? In *Proccedings of the Intelligent Virtual Agents Conference, Madrid*, 2001.

J. M. Digman. Personality Structure: Emergence of the Five Factor Model. *Annual Review of Psychology*, 41:417–440, 1990.

P.A.M. Ehlert and L.J.M. Rothkrantz. Microscopic Traffic Simulation with Reactive Driving Agents. *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, pages 860–865, 2001.

Paul Ekman. Facial Expression and Emotion. *American Psychologist*, 48(4):384–392, 1993.

Paul Ekman. Basic Emotions. In Tim Dalgleish and M. Power, editors, *Handbook of Cognition and Emotion*, chapter 3. John Wiley & Sons, Ltd., Sussex, U.K., 1999.

Clark Davidson Elliott. *The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System.* PhD thesis, Northwestern University, Evanston, IL, USA, 1992.

H. J. Eysenck and M. W. Eysenck. *Personality and Individual Differences: A Natural Science Approach.* Plenum Press, New York, 1985.

E. A. Feinberg and A. Shwartz, editors. *Handbook of Markov Decision Processes - Methods and Applications.* Kluwer International Series, 2002.

Jerry A. Fodor. *The Modularity of Mind.* MIT Press, Cambridge, MA, USA, 1983.

Sandra Clara Gadanho and John Hallam. Robot Learning Driven by Emotions. *Adaptive Behavior*, 9(1):42–64, 2001.

Douglas Gettman and Larry Head. *Surrogate Safety Measures from Traffic Simulation Models.* U.S. Department of Transportation, Federal Highway Administration, 2003.

Jonathan Gratch. Why You Should Buy an Emotional Planner. In Juan D. Velásquez, editor, *Proceedings of the Agents'99 Workshop on Emotion-based Agent Architectures (EBAA'99)*, 1999.

Jonathan Gratch and Stacy Marsella. Evaluating a Computational Model of Emotion. *Autonomous Agents and Multi-Agent Systems*, 11(1):23–43, 2005.

Michael Grimm and Kristian Kroschel. Emotion Estimation in Speech Using a 3D Emotion Space Concept. In Michael Grimm and Kristian Kroschel, editors, *Robust Speech Recognition and Understanding.* I-Tech Education and Publishing, Vienna, Austria, June 2007.

Michael Grimm, Kristian Kroschel, and Shrikanth Narayanan. Speaker and Listener Variations in Emotion Assessment. In *Proceedings of DAGA*, Braunschweig, Germany, March 2006.

Amy E. Henninger, Randolph M. Jones, and Eric Chown. Behaviors that Emerge from Emotion and Cognition: A First Evaluation. In *Proceedings of the '02 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, 2002.

Fenghsiung Hsu. Ibm's Deep Blue Chess Grandmaster Chips. *IEEE Micro*, 19(2):70–81, 1999.

Kakyu Inoue, Kuniaki Kawabata, and Hisato Kobayashi. On a Decision Making System with Emotion. In *5th IEEE International Workshop on Robot and Human Communication, 1996.* IEEE, 1996.

Tania Jiménez, Philippe Mussi, and Günther Siegel. A Road Traffic Simulator: Car-Following and Lane-Changing. In Rik Van Landeghem, editor, *ESM*, pages 241–245. SCS Europe, 2000.

Susanne Kaiser and Thomas Wehrle. Some Studies for Modeling the Expression of Emotion. In Robert Trappl, editor, *Cybernetics and Systems 2006, Proceedings of the Eighteenth Meeting on Cybernetics and Systems Research*, volume 1 and 2, pages 619–624, Austrian Society for Cybernetic Studies, Vienna, 2006.

Rana El Kaliouby and Peter Robinson. Generalization of a Vision-Based Computational Model of Mind-Reading. In Jianhua Tao, Tieniu Tan, and Rosalind W. Picard, editors, *ACII*, volume 3784 of *Lecture Notes in Computer Science*, pages 582–589. Springer, 2005.

Eric R. Kandel, James H. Schwartz, and Thomas M. Jessell. *Principles of Neural Science.* McGraw-Hill Medical, January 2000.

Stephen Kaplan. Beyond Rationality: Clarity Based Decision Making. In Tommy Gärling and Gary W. Evans, editors, *Environment, cognition and action: An integrative multidisciplinary approach*, pages 171–190. Oxford University Press, New York, 1991.

Elizabeth A. Kensinger. Negative Emotion Enhances Memory Accuracy: Behavioral and Neuroimaging Evidence. *Current Directions in Psychological Science*, 16(4):213–218, August 2007.

P. Kleinginna and A. Kleinginna. A Categorized List of Emotion Definitions, with Suggestions for a Consensual Definition. *Motivation and Emotion*, 5:345–379, 1981.

G Klösch, B Kemp, T Penzel, A Schlögl, P Rappelsberger, E Trenker, G Gruber, J Zeitelhofer, B Saletu, W M Herrmaann, S L Himanen, D Kunz, M J Barbanoj, J Röschke, A Värri, and G Dorffner. The SIESTA Project Polygraphic and Clinical Database. *IEEE Engineering in Med. and Bio Mag 20(3)*, pages 51–7, 2001.

Zornitsa Kozareva, Borja Navarro, Sonia Vázquez, and Andrés Montoyo. A Headline Emotion Classification through Web Information. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic, 2007. Association for Computational Linguistics.

Sumedha Kshirsagar. A Multilayer Personality Model. In *SMARTGRAPH '02: Proceedings of the 2nd international symposium on Smart graphics*, volume 24 of *ACM International Conference Proceeding Series*, pages 107–115, New York, 2002. ACM Press.

Richard D. Lane and Lynn Nadel. *Cognitive Neuroscience of Emotion (Series in Affective Science)*. Oxford University Press, USA, April 2002.

Linda J. Levine and David A. Pizarro. Emotion and Memory Research: A Grumpy Overview. *Social Cognition*, 22(5):530–544, 2004.

Edward Lieberman and Ajay K. Rathi. Traffic Simulation. In Nathan H. Gartner, Carrol J. Messer, and Ajay K. Rathi, editors, *Traffic Flow Theory - A State-of-the-Art Report: Revised Monograph on Traffic Flow Theory*. Turner Fairbank Highway Research Center (TFHRC) U.S. Department of Transportation, April 2002.

Mei Yii Lim, Ruth Aylett, and Christian Martyn Jones. Affective Guide with Attitude. In Jianhua Tao, Tieniu Tan, and Rosalind W. Picard, editors, *ACII*, volume 3784 of *Lecture Notes in Computer Science*, pages 772–779. Springer, 2005.

Robert R. McCrae and Oliver P. John. An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, 60(2):175–215, 1992.

Albert Mehrabian. Framework for a Comprehensive Description and Measurement of Emotional States. *Genetic, Social & General Psychology Monographs*, 121:339–361, August 1995.

Jolieke Mesken. *Determinants and Consequences of Drivers' Emotions*. PhD thesis, Rijksuniversiteit Groningen, Leidschendam, Nederland, November 2006.

C.H. Messom, A. Sarrafzadeh, M.J. Johnson, and F. Chao. Affective State Estimation from Facial Images Using Neural Networks and Fuzzy Logic. In Dianhui Wang, editor, *Neural Networks Applications in Information Technology and Web Engineering*, 2005.

Wulf-Uwe Meyer, Rainer Reisenzein, and Achim Schützwohl. *Einführung in Die Emotionspsychologie*, volume 3. Huber Hans, Bern, 2003.

Marvin Minsky. *Society of Mind*. Simon & Schuster, New York, NY, USA, 1988.

Marvin Minsky. *The Emotion Machine.* Simon & Schuster, New York, NY, USA, 2006.

Luís Morgado and Graça Gaspar. Adaptation and Decision-Making Driven by Emotional Memories. In Carlos Bento, Amílcar Cardoso, and Gaël Dias, editors, *EPIA*, volume 3808 of *Lecture Notes in Computer Science*, pages 102–114. Springer, 2005.

Keith Oatley. *Emotions: A Brief History.* Blackwell Publishing, Oxford, UK, 2004.

Keith Oatley and Jennifer M. Jenkins. *Understanding Emotions.* Blackwell Publishers, Cambridge, MA, USA, 1996.

Andrew Ortony. *On Making Believable Emotional Agents Believable*, volume 1, pages 189–211. MIT Press, Cambridge, MA, USA, 2003.

Andrew Ortony, Gerald L. Clore, and Mark A. Foss. The Referential Structure of the Affective Lexicon. *Cognitive Science*, 11(3):341–364, 1987.

Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotion.* Cambridge University Press, Cambridge, UK, 1988.

Ana Paiva, Rui Prada, and Rosalind W. Picard, editors. *Affective Computing and Intelligent Interaction, Second International Conference, Acii 2007, Lisbon, Portugal, September 12-14, 2007, Proceedings*, volume 4738 of *Lecture Notes in Computer Science*, 2007. Springer.

Christian Peter, Eric Ebert, and Helmut Beikirch. A Wearable Multi-Sensor System for Mobile Acquisition of Emotion-Related Physiological Data. In Jianhua Tao, Tieniu Tan, and Rosalind W. Picard, editors, *ACII*, volume 3784 of *Lecture Notes in Computer Science*, pages 691–698. Springer, 2005.

Rosalind W. Picard. *Affective Computing.* The MIT Press, September 1997.

Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, 2001.

Stefan Rank, Paolo Petta, and Robert Trappl. Features of Emotional Planning in Software Agents. In Giacomo Della Riccia, Didier Dubois, Rudolf Kruse, and Hans-Joachim Lenz, editors, *Decision Theory and Multi-Agent Planning (CISM International Centre for Mechanical Sciences).* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

Giacomo Della Riccia, Didier Dubois, Rudolf Kruse, and Hans-Joachim Lenz, editors. *Decision Theory and Multi-Agent Planning (Cism International Centre for Mechanical Sciences).* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

Edmund T. Rolls. Memory Systems in the Brain. *Annual Review of Psychology*, 51:599–630, 2000.

Edmund T. Rolls. *A Theory of Emotions, Its Functions, and Its Adaptive Value*, volume 1, pages 11–34. MIT Press, Cambridge, MA, USA, 2003.

Edmund T. Rolls. *The Brain and Emotion.* Oxford University Press, Oxford, UK, 1999.

Ira J. Roseman. Cognitive Aspects of Emotion and Emotional Behavior. Paper presented at the 87th Annual Convention of the American Psychological Association, 1979.

Ira J. Roseman, Ann Aliki Antoniou, and Paul E. Jose. Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. *Cognition and Emotion*, 10(38): 241–278, May 1996.

Richard W. Rothery. Car Following Models. In Nathan H. Gartner, Carrol J. Messer, and Ajay K. Rathi, editors, *Traffic Flow Theory - A State-of-the-Art Report: Revised Monograph on Traffic Flow Theory.* Turner Fairbank Highway Research Center (TFHRC) U.S. Department of Transportation, April 2002.

Farida Saad. Theories on Human Behavior and Interaction in Traffic. In Wolf Dietrich Zuzan, editor, *Safety Evaluation of Traffic Systems: Traffic Conflicts and Other Measures, Proceedings of the 6th ICTCT Workshop*, pages 103–106, Salzburg, Austria, October 1993. Kuratorium für Verkehrssicherheit.

Aristide Saggino. The Big Three or the Big Five? A Replication Study. *Personality and Individual Differences*, 28(5):879–886, 2000.

Marina Santini. State-of-the-Art on Automatic Genre Identification. Technical Report ITRI-04-03, Information Technology Research Institute, University of Brighton, 2003.

Klaus R. Scherer. Emotions as Episodes of Subsystem Synchronization Driven by Nonlinear Appraisal Processes. In Marc D. Lewis and Isabela Granic, editors, *Emotion, Development, and Self-Organization*, pages 70–99. Cambridge University Press, 2000.

Klaus R. Scherer. The Nature and Study of Appraisal: A Review of the Issues. In Angela Schorr and Klaus R. Scherer, editors, *Appraisal processes in emotion: Theory, methods, research*, pages 369–391. Oxford University Press, London, 2001.

Klaus R. Scherer and Martin Peper. Psychological Theories of Emotion and Neuropsychological Research. In F. Boller and J. Grafman, editors, *Handbook of Neuropsychology*, volume 5. Elsevier, Amsterdam, 2001.

Matthias Scheutz. Useful Roles of Emotions in Artificial Agents: A Case Study from Artificial Life. In Deborah L. McGuinness and George Ferguson, editors, *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence, July 25-29, 2004, San Jose, California, USA*, pages 42–48. AAAI Press / The MIT Press, 2004.

Matthias Scheutz and Brian Logan. Affective Vs. Deliberative Agent Control. In *Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing*, pages 1–10. AISB, The Society for the Study of Artificial Intelligence and, March 2001.

Matthias Scheutz and Aaron Sloman. Affect and Agent Control: Experiments with Simple Affective States. In Ning Zhong, Jiming Liu, Setsuo Ohsuga, and Jeffrey Bradshaw, editors, *Intelligent Agent Technology: Research and Development*, pages 200–209, New Jersey, 2001. World Scientific Publisher.

Marc Schröder and Roddy Cowie. Issues in Emotion-Oriented Computing – Towards a Shared Understanding. In *Workshop on Emotion and Computing at KI 2006*, Bremen, Germany, 2006.

Herbert A. Simon. Motivational and Emotional Controls of Cognition. *Psychological Review*, 74(1):29–39, 1967.

Aaron Sloman. *How Many Separately Evolved Emotional Beasties Live within Us?*, volume 1, pages 35–114. MIT Press, Cambridge, MA, USA, 2003.

Aaron Sloman. What Are Emotion Theories about? In Eva Hudlicka and Lola Dolores Cañamero, editors, *Architectures for Modeling Emotion: Cross-Disciplinary Foundations*, pages 128–134. AAAI Press, 2004.

Aaron Sloman. Prospects for AI as the General Science of Intelligence. In Aaron Sloman, D. Hogg, G. Humphreys, A. Ramsay, and D. Partridge, editors, *Prospects for Artificial Intelligence: Proc. of AISB-93*, pages 1–10. IOS Press, Amsterdam, 1993.

Aaron Sloman. Beyond Shallow Models of Emotion. In Elisabeth André, editor, *Behaviour planning for life-like avatars*, pages 35–42. Sitges, Spain, 1999a. Proceedings I3 Spring Days Workshop March 9th–10th 1999.

Aaron Sloman. Review of Affective Computing. *AI Magazine*, 20(1):127–133, 1999b.

Aaron Sloman and Monica Croucher. Why Robots Will Have Emotions. In *IJCAI*, pages 197–202, 1981.

Brian L. Smith and Jared M. Ulmer. Freeway Traffic Flow Rate Measurement: Investigation into Impact of Measurement Time Interval. *Journal of Transportation Engineering*, 129(3): 223–229, 2003.

Craig A. Smith and Richard S. Lazarus. Emotion and Adaptation. In L. A. Pervin, editor, *Handbook of personality: Theory and research*, chapter 23, pages 609–637. Guilford, New York, 1990.

Carlo Strapparava and Rada Mihalcea. Affective Text. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic, 2007. Association for Computational Linguistics.

Marc Strauss, Carson Reynolds, Stephen Hughes, Kyoung Park, Gary McDarby, and Rosalind W. Picard. The Handwave Bluetooth Skin Conductance Sensor. In Jianhua Tao, Tieniu Tan, and Rosalind W. Picard, editors, *ACII*, volume 3784 of *Lecture Notes in Computer Science*, pages 699–706. Springer, 2005.

Michael Streit, Anton Batliner, and Thomas Portele. Cognitive-Model-Based Interpretation of Emotions in a Multi-Modal Dialog System. In Elisabeth André, Laila Dybkjær, Wolfgang Minker, and Paul Heisterkamp, editors, *ADS*, volume 3068 of *Lecture Notes in Computer Science*, pages 65–76. Springer, 2004.

Pero Subasic and Alison Huettner. Affect Analysis of Text Using Fuzzy Semantic Typing. In John Yen and Reza Langari, editors, *Proceedings of the 9th IEEE International Conference on Fuzzy Systems*, volume 2, pages 647–652. IEEE Neural Networks Council, May 2000.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning)*. The MIT Press, March 1998.

Jianhua Tao, Tieniu Tan, and Rosalind W. Picard, editors. *Affective Computing and Intelligent Interaction, First International Conference, Acii 2005, Beijing, China, October 22-24, 2005, Proceedings*, volume 3784 of *Lecture Notes in Computer Science*, 2005. Springer.

Martin Treiber. Longitudinal Traffic Model: The IDM, 2006. URL `http://vwisb7.vkw.tu-dresden.de/~treiber/MicroApplet/IDM.html`.

Martin Treiber. The Lane-Change Model MOBIL, 2007. URL `http://www.vwi.tu-dresden.de/~treiber/MicroApplet/MOBIL.html`.

Martin Treiber and Dirk Helbing. Realistische Mikrosimulation Von Straßenverkehr Mit Einem Einfachen Modell. In *Tagungsband 16. Symposium Simulationstechnik, Rostock*, pages 514–520, 2002.

Martin Treiber, Ansgar Hennecke, and Dirk Helbing. CongestedTTraffic SStatesïn Empirical Observations and Microscopic Simulations. *Phys. Rev. E*, 62(2):1805–1824, Aug 2000.

Maurits van den Noort, Kenneth Hugdahl, and Peggy Bosch. Human Machine Interaction: The Special Role for Human Unconscious Emotional Information Processing. In Jianhua Tao, Tieniu Tan, and Rosalind W. Picard, editors, *ACII*, volume 3784 of *Lecture Notes in Computer Science*, pages 598–605. Springer, 2005.

Wilco W. van Dijk and Marcel Zeelenberg. Investigating the Appraisal Patterns of Regret and Disappointment. *Motivation and Emotion*, 26(4):321–331, 2002.

Enrique Carbonell Vaya, Rosa Bañuls Egeda, and J. J. Miguel Tobal. Anxiety and Traffic Safety. In Wolf Dietrich Zuzan, editor, *Safety Evaluation of Traffic Systems: Traffic Conflicts and Other Measures, Proceedings of the 6th ICTCT Workshop*, pages 139–155, Salzburg, Austria, October 1993. Kuratorium für Verkehrssicherheit.

Juan D. Velásquez. Modeling Emotions and Other Motivations in Synthetic Agents. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Innovative Applications of Artificial Intelligence Conference, AAAI 97, IAAI 97, July 27-31, 1997, Providence, Rhode Island*, pages 10–15. AAAI, AAAI Press / The MIT Press, 1997.

Rodrigo Ventura and Carlos Pinto-Ferreira. Emotion-Based Agents: Three Approaches to Implementation. In Juan D. Velásquez, editor, *Workshop on Emotion-Based Agent Architectures of the Third International Conference on Autonomous Agents.*, 1999.

Thomas Wehrle. The Grounding Problem of Modeling Emotions in Adaptive Artifacts. *Cybernetics and Systems*, 32:561–580(20), July 2001.

Thomas Wehrle and Klaus R. Scherer. Potential Pitfalls in Computational Modeling of Appraisal Processes: A Reply to Chwelos & Oatley. *Cognition and Emotion*, 9:599–616, 1995.

Wilhelm Wundt. *Vorlesungen Über Die Menschen- Und Tierseele*. Verlag von Leopold Voss, Leipzig, 1919.

N. Zuber, Z. Bockal, and Z. Kavran. Drivers' Aggressiveness and Road Traffic Safety. In B. Katalinic, editor, *Intelligent Manufactoring and Automation, Proceedings of the 17th International DAAAM Symposium*, pages 459–460, Vienna, Austria, 2006.

Index

Appendix

SAD Command Line Options

--help	show the list of options
--small	small window
--large	large window
--length	lane length
--debug[=n]	enable debug output [optional: set level]
--timetag	add timestamp to filenames
--prefix=<PREFIX>	prefix for output files
--lanes=<0..9>	initial number of lanes
--inputfile=<FILENAME>	car position snapshot file
--overwrite=<0..p>	use personality x for vehicles in file
--nopause	no pause after reading snapshot
--cycles=n	run the simulation for n steps
--autospawn=n	automatically spawn n cars
--spawnrate=n	spawn in an interval of every n seconds
--spawntype=<0..p>	automatically spawn with personality x
--nowait	run simulation at full speed
--sectorsize=n	size for sector measures
--sectordensity	write sector density
--personality=<"f f f f...">	personality definition
--nographics	no graphics
--startcounterat=n	start counter when n cars are driving
--countflowrate	count flow rate
--runcounter=s	run counter for s seconds
--counterinterval=s	measure interval in s
--startstatisticsat=s	start to measure statistics at time s
--writeeverystep	write a variety of data
--alpha_up=n	set attack gain, α_{up} to $n/1000$
--alpha_dn=n	set decay gain, α_{dn} to $n/1000$

`SAD` Key Mappings in Graphical Mode

esc, `q`	exit simulation
space	pause simulation
`+`	increase the speed of the simulation
`-`	decrease the speed of the simulation
`*`	set speed of simulation to maximum
`_`	default speed
`.`	step forward in single time steps (interrupts normal operation)
`:`	continue normal operation
`n`	show numbers of the vehicles
`b`	step backwards in time
`l`	add a lane
`L`	remove a lane
`e`	show the driver's emotions
`a`	start autospawning of vehicles
`c`	increase the rate of spawning vehicles
`C`	decrease the speed of spawning vehicles
`s`	make a snapshot of all vehicles
`!`	insert an obstacle on the rightmost lane
`"`	insert an obstacle on the middle lane
`$`	insert an obstacle on the leftmost lane
`0`	insert an IDM driver
`1`	insert a pseudo-emotional IDM driver
`2`	insert an emotional driver: normal
`3`	insert an emotional driver: aggressive
`4`	insert an emotional driver: fearful
`t`	insert an emotional driver: disciplined
`f`	insert an emotional driver with a personality specified using `--personality` on the command line
`5`	display which vehicle a driver is seeing at his front
`6`	display which vehicle a driver is seeing at his left front
`7`	display which vehicle a driver is seeing at his left back
`8`	display which vehicle a driver is seeing at his right front
`9`	display which vehicle a driver is seeing at his right back
`m`	switch to colored lines