

**Emotionale Agenten:
Darstellung der emotionstheoretischen
Grundlagen und Entwicklung eines
Referenzmodells auf Basis einer
petrinetz-basierten Modellierungstechnik**

Dissertation

zur Erlangung des Doktorgrades
an der MIN-Fakultät
Fachbereich Informatik der
Universität Hamburg

vorgelegt von

Julia Fix

aus St. Petersburg (Russland)

Hamburg, 2012

Meinen Eltern gewidmet

Genehmigt von der MIN-Fakultät, Fachbereich Informatik der Universität Hamburg auf
Antrag von

Dr. Daniel Moldt,
Prof. Dr. Rüdiger Valk,
Prof. Dr. Bernd Page,
Prof. Dr. Rolf von Lüde und
Prof. Dr. Tilo Böhmann

Hamburg, den 6. August 2012 (Tag der Disputation)

Prof. Dr. Ingrid Schirmer,
Leiterin des Fachbereichs Informatik

Zusammenfassung

Seit einigen Jahrzehnten stellt die Entwicklung von *Emotionalen Agenten* ein aktives und ständig wachsendes Forschungsgebiet der Informatik dar. Trotz (oder gerade wegen) dieses kontinuierlichen Wachstums ist das Forschungsfeld zurzeit sehr fragmentiert: es gibt zahlreiche Emotionstheorien, -konzepte und -modelle, ohne dass die Integration der emotionstheoretischen Grundlagen für Agenten- oder Multiagentensysteme bisher erfolgt ist. Hinsichtlich ihrer grundlegenden Annahmen, ihrer Terminologie, ihrer Gültigkeitsgrenzen, ihres Erklärungspotenzials und ihrer empirischen Überprüfung divergieren die informatischen Emotionsmodelle daher sehr stark. Die intuitive Vorgehensweise und die mangelnde theoretische Fundierung vieler Modellierungsansätze ist vor allem auf fehlende Grundlagenforschung und mangelnde theoretische Verankerung der Emotionsproblematik in der Informatik zurückzuführen. In diesem Zusammenhang ist eine Konsolidierung wünschenswert, für die in dieser Arbeit eine erste Grundlage geschaffen werden soll.

Die allgemeine Zielsetzung der vorliegenden Arbeit besteht in der Aufbereitung theoretischer Grundlagen für Emotionsmodellierung in der Informatik durch die Untersuchung und modellhafte Umsetzung von psychologischen, soziologischen und neurowissenschaftlichen Theorien zu Emotion mit einer informatischen Darstellungstechnik. Als Grundlage der Modellierung dient der höhere Petrinetzformalismus der Referenznetze in Verbindung mit dem sozionischen Modellierungsansatz. Für die Auswertung von existierenden Emotionstheorien im Hinblick auf ihre Verwendbarkeit als Lösungsangebot für spezifische informatische Fragestellungen bzw. Anwendungen werden Theorien, Definitionen und Konzepte strukturiert zusammengestellt und in einen Bezug zu entsprechenden Anwendungsgebieten für Emotionen in der Informatik gesetzt.

Diverse Probleme der Emotionsmodellierung, insbesondere die zunehmende Forschungsfragmentierung, sind auf das Fehlen einer gemeinsamen Beschreibungssprache für emotionale Prozesse / Architekturen zurückzuführen. In dieser Arbeit wird ein Vorschlag für ein Referenzmodell für Emotionsmodellierung unterbreitet, das für die Analyse, Klassifizierung und den Vergleich verschiedener emotionaler Systemarchitekturen und Prozesse eingesetzt werden kann.

Das Referenzmodell E-MULAN wird auf Basis der referenznetz-basierten Multiagentenarchitektur MULAN entwickelt und so konzipiert, dass es als generisches Modell eine explizite Repräsentation sowohl rationaler / kognitiver als auch emotionaler Prozesse innerhalb desselben Rahmenwerks ermöglicht, sowie insbesondere die Wechselwirkungen zwischen diesen Prozessen darstellbar macht. Damit wird einerseits eine Modellierungsbasis für die Darstellung emotionaler Prozesse für Agentenanwendungen geschaffen, andererseits aber auch ein Referenzmodell zur Beschreibung und Analyse von emotionalen Prozessen in Emotionsarchitekturen. Das E-MULAN-Modell bildet somit einen gemeinsamen konzeptuellen und terminologischen Rahmen für die Darstellung der emotionstheoretischen Grundlagen und ihre Integration in informatische Emotionsmodelle.

Abstract

Since some decades the research on emotional agents is a continuously developing and a very active branch within the research field of Affective Science. This fast growth is probably the factor that have led to the fragmentation of the research: Today the field is divided into numerous theories, concepts and models and it still lacks an integrated computational theory of emotion in view of their large-scale causes and effects, both for the individual agent and the multi-agent society. Hence computational models of emotion feature crucial differences with respect to its basic assumptions, terminology, validity, theoretical and empirical underpinning etc. The use of intuitive concepts and the lack of theoretical underpinnings have resulted in conceptual confusion surrounding the term “emotion“ in AI and DAI research. This has often been criticized by leading researchers of the field and the need for a consolidation has been proclaimed more than once.

On the most general level, the purpose of this theses is to contribute to a more elaborated theoretical underpinning for emotion modelling in (distributed) artificial systems. This purpose is being addressed by means of analyzing of the psychological, neurobiological and sociological theories of emotion and by integrating these different perspectives on emotion using a general modelling technique. This modelling technique is based on the High-level Petri nets formalism of Reference nets together with the Socionical modelling approach. Due to its operational semantics and the broad range of available analytical methods, modelling with Petri nets allows an evaluation and formal checking of semantics of the models. In addition, the formal representation of emotion models facilitates their integration into the computational domain. Reference nets introduce recursivity to the Petri net theory and the possibility of dynamic creation of net instances. Together with these extensions, Petri nets provide a powerful instrumentarium for modelling dynamic, hierarchical and recursive structures, as described by psychological, neurological and social theories of emotion.

It has often been argued that for an appropriate modelling of emotion in (distributed) artificial systems a proper theoretical foundation is required that should allow integration of psychological, sociological and neurological theories of emotion. In this matter, a conceptual and terminological framework is proposed here, which provides the possibility to combine different theoretical approaches and models of emotion in a single multi-agent system.

This approach is concerned with extending a multi-agent framework MULAN, which is a technical implementation of the agent concept on the basis of the reference net formalism, to a general reference architecture for modelling emotions in multi-agent systems (E-MULAN). This architecture provides a possibility to model and analyze emotional theories, to integrate and evaluate a number of (interdisciplinary) approaches within a single modelling framework and to study their possible interdependencies and reciprocities. Furthermore, it allows for analytical separation of cognitive and emotional components, which is also an important requirement for a general reference architecture for modelling emotions.

With E-MULAN a conceptual and terminological reference architecture is provided, which facilitates the integration of different theoretical approaches to emotion in a unique model, describing multiple interdisciplinary aspects of emotional phenomena and processes simultaneously and allowing for their representation and analysis, thus providing a more elaborated theoretical underpinning for (D)AI models of emotion.

Danksagung

Die vorliegende Arbeit ist während meiner Zeit als Stipendiatin der Abteilung Forschung und Wissenschaftsförderung der Universität Hamburg am Arbeitsbereich TGI (‘Formale Grundlagen der Informatik’) entstanden.

Meinem Betreuer Dr. Daniel Moldt möchte ich für seine Begeisterung für die Emotionsthematik danken, sowie für die Vielzahl an Diskussionen, Denkanstößen und neuen Ideen. Ohne seinen unermüdlichen Einsatz, seine ständige Diskussionsbereitschaft und Ermutigung wäre diese Arbeit gar nicht möglich gewesen.

Bei meinem Betreuer Prof. Dr. Rüdiger Valk möchte ich mich für seine stets konstruktiven und hilfreichen Anmerkungen bedanken, aber auch für die mir zugestandenen Freiräume, sowie für seine Bereitschaft, sich mit dem Thema zu beschäftigen.

Prof. Dr. Bernd Page bin ich sehr dankbar für seine Bereitschaft, die Begutachtung der Arbeit recht kurzfristig zu übernehmen. Die von ihm organisierten Doktorandenseminare der Arbeitsgruppe MBS (‘Modellbildung und Simulation’) fand ich in der Endphase der Promotion sehr hilfreich, insbesondere haben mir aber auch seine ausführlichen und aufschlussreichen Anmerkungen und Kommentare zu der ersten schriftlichen ‘Endfassung’ der Arbeit weitergeholfen.

Prof. Dr. Rolf von Lüde danke ich ebenso recht herzlich für seine Teilnahme am Promotionsverfahren. Seine Arbeiten im Kontext der Sozionik-Projekte lieferten zahlreiche Anregungen und Motivationen für diese Arbeit.

Ich bedanke mich bei Mitarbeitern und Promovenden der Arbeitsgruppe TGI, in erster Linie Dr. Lawrence Cabac, Dr. Matthias Wester-Ebbinghaus, Dr. Michael Köhler-Bußmeier und Dr. Heiko Rölke, die mir bei der Einarbeitung in die MULAN-Umgebung und bei der Umsetzung der Anwendungsszenarien, aber auch bei allen möglichen technischen und konzeptionellen Fragen immer wieder weitergeholfen haben. Dr. Michael Duvigneau danke ich ganz besonders für seinen maßgeblichen Beitrag zur Umsetzung der Synchronisationsmechanismus für MULAN-Agenten und sein stets offenes Ohr bei allen Fragen zu CAPA, MULAN, RENEW und Latex. Bei Dr. Christian von Scheve möchte ich mich für die Unterstützung bei den soziologischen Fragestellungen bedanken. Die Mitarbeit an den gemeinsamen Veröffentlichungen hat mir immer viel Spaß gemacht und ich konnte dabei viel lernen.

Außerdem bedanke ich mich bei meinen MBS-Kollegen Arne Koors, Philipp Joschko und Johannes Göbel für die Anregungen aus der simulationstechnischen Perspektive, sowie bei Sven Kruse, Milena Andonova und Prof. Dr. Fleischer für die vielen interessanten Diskussionen und konstruktiven Fragen.

Am allermeisten möchte ich aber meiner Familie - in erster Linie meinen Eltern - für ihre Geduld und Ausdauer danken! Ohne ihre fortwährende Unterstützung und ihre ständige Bereitschaft, spontan bei der Kinderbetreuung einzuspringen, Mittagessen für David zu kochen oder mit Benjamin Fußball zu spielen, wäre diese Arbeit gar nicht denkbar gewesen.

Inhaltsverzeichnis

I. Grundlagen	13
1. Einleitung und Übersicht	15
1.1. Motivation	15
1.2. Offene Fragen der informatischen Emotionsmodellierung	17
1.3. Ziele der Arbeit	22
1.4. Aufbau	25
1.5. Terminologie	27
2. Forschungs- und Anwendungsfelder für emotionale Agenten	31
2.1. Emotion und Künstliche Intelligenz	36
2.2. Emotion und Verteilte Künstliche Intelligenz	37
2.3. Emotion zur Optimierung der Mensch-Computer-Interaction	42
3. Agenten- und Multiagentensysteme	47
3.1. Begriffsbestimmung	47
3.1.1. Agentendefinition	47
3.1.2. Agenteneigenschaften	49
3.1.3. Umgebungsdimensionen	50
3.2. Agentenarchitekturen	51
3.2.1. Abstraktes Agentenmodel	51
3.2.2. Rein reaktiver Agent	53
3.2.3. Agent mit innerem Zustand	55
3.2.4. Agent mit Utility-Funktion	56
3.2.5. Deliberativer Agent	57
3.2.6. Hybride Architekturen	60
3.3. Emotionale Agenten	61
3.3.1. Grundlegenden Agenteneigenschaften und Emotion	62
3.3.2. Emotionale BDI-Agenten	65
3.4. Multiagentensysteme	67
3.4.1. Koordination und Kommunikation	68
3.4.2. Agentenkommunikationssprachen	69
3.4.3. FIPA-Referenzarchitektur	70
3.4.4. Modellierungstechniken für MAS	72

4. Modellierungstechnik	73
4.1. Benutzte Formalismen	73
4.1.1. Petrinetze	74
4.1.2. Objektnetze, Referenznetze	74
4.1.3. Agentennetze	76
4.1.4. CAPA	77
4.2. Eine petrinetzbasierte Multiagentenarchitektur: Mulan	78
4.3. PAOSE-Ansatz	81
4.4. Petrinetze als sozionische Darstellungstechnik	81
4.4.1. Sozionische Theorienmodellierung mit Petrinetzen	81
4.4.2. Vermittelnde Netze: Grundlagen und allgemeine Modellierungsprinzipien	84
4.5. Anwendung des sozionischen Ansatzes bei der Modellierung von Emotionstheorien	89
4.5.1. Motivation	89
4.5.2. Integration auf Basis der Modellierungstechnik	90
4.5.3. Zusammenfassung und weiteres Vorgehen	91
 II. Theorien und Modelle	 93
5. Emotionsdefinition	95
5.1. Abgrenzung von Emotionen zu verwandten mentalen Phänomenen	98
5.2. Emotionsklassifikationen	103
5.2.1. Kategoriale Klassifikation	104
5.2.2. Dimensionale Klassifikation	109
5.2.3. Emotion: Zustand vs. Prozess	110
5.3. Emotionspsychologische Arbeitsdefinition	110
5.4. Emotion als Anpassungsmechanismus: Evolutionspsychologische Definitionen	115
5.5. Eine neurobiologische Emotionsdefinition	120
5.6. Zusammenfassung und (Neu-)Definition des Emotionsbegriff für technische Systeme	124
 6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien	 133
6.1. Psychologische Theorien der Emotion	135
6.1.1. Klassische Emotionstheorien der Psychologie	136
6.1.2. Einschätzungstheorien	138
6.1.3. Zusammenfassung und Gesamtbewertung der psychologischen Emotionstheorien im Hinblick auf informatische Modellierung	176
6.2. Neurobiologische Grundlagen von Emotion	179
6.2.1. Neurowissenschaftlicher Forschungsstand zur Emotion	179
6.2.2. Emotion und soziale Norm: Soziale Emotionen bei der Regulation sozialen Verhaltens	185

6.3.	Soziologische Erklärungsansätze zur Emotion	191
6.3.1.	Emotionen bei soziologischen Klassikern: Simmel, Durkheim, Weber, Parsons	193
6.3.2.	Emotionen in der modernen soziologischen Forschung	195
6.3.3.	Die Rolle der sozialen Emotionen bei der Etablierung und Aufrecht- erhaltung sozialer Normen: Modellierung und Integration der Ansätze	199
6.3.4.	Auswertung der soziologischen Emotionstheorien und offene For- schungsfragen	214
7.	Architekturen emotionaler Agenten	217
7.1.	Klassifizierung der Emotionsarchitekturen	218
7.2.	KI	220
7.2.1.	Staller und Petta: Die TABASCO-Architektur	221
7.2.2.	Velásquez: Emotion bei der Entscheidungsfindung	228
7.2.3.	Dörners PSI und Bachs MicroPSI	239
7.2.4.	Sloman: CogAff als konzeptuelles Framework für Architekturen emo- tionaler intelligenter Systeme	254
7.3.	VKI	268
7.3.1.	Elliott: Affective Reasoner	268
7.3.2.	Das Oz-Projekt von Bates und Reilly	274
7.3.3.	Der Ansatz von Aubé und Senteni: Emotion als Metapher für Kon- trollstrukturen in MAS	279
7.4.	Auswertung der informatischen Emotionsmodelle	284
7.4.1.	Überblick über weitere informatische Emotionsarchitekturen	285
7.4.2.	Auswertung der Emotionsmodelle	288
III.	Anwendungen	295
8.	Referenzmodell für Emotionsmodellierung	297
8.1.	Grundsätzliche Annahmen und Ziele	298
8.1.1.	Definition der Modellebenen	298
8.1.2.	Interpretation von E-MULAN als Systemarchitektur und als Referenz- modell	299
8.1.3.	Analytische Trennung in rationale und emotionale Modellkomponenten	300
8.1.4.	Ein generisches Referenzmodell auf Basis der MULAN-Architektur .	301
8.2.	Vorstellung des allgemeinen Referenzmodells	302
8.2.1.	Gesellschaftsebene	306
8.2.2.	Soziale Struktur	307
8.2.3.	Akteurs- / Agentenebene	308
8.2.4.	Emotionsebene	310
8.3.	Modellierung mit E-MULAN	313
8.3.1.	Das Modell der Emotionseinschätzung	314
8.3.2.	Modell der emotionsbasierten Steuerung	323

Inhaltsverzeichnis

8.4. Verknüpfung der rationalen und der emotionalen Modellkomponenten	330
8.4.1. Lose vs. enge Kopplung	330
8.4.2. E-MULAN als Steuerungsmodell	331
8.4.3. Synchronisationsmechanismus für emotionale und rationale Modellkomponenten	334
8.4.4. Realisierung der emotionsbasierten Steuerung unter Verwendung der synchronen Kommunikation	337
8.5. Diskussion des Referenzmodells	338
9. Anwendungsbeispiele	351
9.1. Grundsätzliche Annahmen und Modellierungsansatz	351
9.2. Modell der reaktiven Emotionsgenerierung in E-MULAN	352
9.3. Modell der proaktiven Emotionsgenerierung in E-MULAN	362
9.4. Modell der emotionsbasierten Steuerung in E-MULAN	368
9.5. Modell der emotionsbasierten Normetablierung	376
10. Zusammenfassung und Ausblick	383
10.1. Diskussion der Ergebnisse	384
10.2. Offene Fragen und Ausblick	394
Literaturverzeichnis	399

Teil I.

Grundlagen

1. Einleitung und Übersicht

1.1. Motivation

Agentenorientierte Softwareentwicklung ist ein Schwerpunkt aktueller Forschung. Sie hat das Ziel, Softwaresysteme aus autonomen und intelligenten Einheiten (Agenten) zu komponieren. Unter Agenten sind intelligente, autonome, mobile, reaktive und / oder proaktive sowie z.T. sozial kompetente (Software-)Komponenten zu verstehen, die in der Lage sind, in unbekannten Umgebungen und unter nicht vorhersehbaren Bedingungen eigenständig und zielgerichtet Entscheidungen zu treffen, Verhaltensweisen auszubilden und ihre Handlungen mit denen anderer Agenten zu koordinieren (vgl. dazu exemplarisch [Bra97]). Neben Autonomie sind daher explizites Wissen über sich, andere Agenten und die Umgebung, zielgerichtetes Handeln sowie Anpassungsfähigkeit weitere wesentliche Eigenschaften eines Agenten.

Um Agentenarchitekturen zu entwickeln, die über diese Fähigkeiten verfügen, wird in der Informatik auf den Gebieten der Künstlichen Intelligenz (KI) und der Verteilten Künstlichen Intelligenz (VKI) versucht, biologische und / oder soziale Mechanismen nachzubilden, die in natürlichen Systemen Handlungswahl und Entscheiden leiten und diesen Systemen zu Leistungsfähigkeit, Fehlertoleranz, Adaptivität, Robustheit und Effektivität verhelfen. So werden beispielsweise Erkenntnisse aus der Psychologie und den Neuro- und Kognitionswissenschaften herangezogen, um grundlegende Prozesse der menschlichen Intelligenz, des Wahrnehmens, der Informationsverarbeitung, des Entscheidens, der Wissensorganisation und des Verhaltens bei der Konstruktion von Agentenarchitekturen nutzbar zu machen (vgl. [Bac08], [BD07], [BC06], [MGP12], [CP09]).

Bislang haben sich diese Anstrengungen auf die Übertragung der Eigenschaften und des Potenzials der *kognitiven* Systeme von Menschen und Tieren auf technische Systeme konzentriert. Seitdem von den Kognitionswissenschaften zunehmend erkannt wurde, dass für diese grundlegenden Prozesse intelligenten Verhaltens, die auf technische Systeme übertragen werden sollen, auch Emotionen und nicht nur kognitives Vermögen eine maßgebliche Rolle spielen, wurde auf informatischer Seite zunehmend versucht, die Wirkungsweisen der natürlichen emotionalen Mechanismen für technische Systeme nutzbar zu machen. Emotion wurde als integraler Bestandteil vieler kognitiven Prozesse gesehen, der insbesondere bei der Wahrnehmung, dem Entscheiden, der Verhaltenssteuerung usw. beteiligt ist: “AI, interested in cognitive processes modeling and simulation, clearly see that emotion is a crucial element to model perception, learning, decision processes, memory, behavior and others functions.” ([dFGQ05], S. 1).

Seitdem der Neurowissenschaftler Antonio Damasio nachweisen konnte, dass Emotion bei Menschen vor allem in Entscheidungssituationen, eine ausschlaggebende Rolle spielt

1. Einleitung und Übersicht

[Dam94] und auch weitere Forscher Emotion als eines der Schlüsselemente der menschlichen Intelligenz und Adaptivität identifiziert haben, das für Planen und Entscheidungsfindung essenziell notwendig ist (z.B. [Gol95], [LeD96a]), hat die Anzahl der informatischen Modellierungsansätze zu Emotion stetig zugenommen (vgl. [SMBB08], [MGP12]).

Die bekannte Aussage aus Minsky's Buch "Societies of Mind" fasst die vorherrschende Sichtweise der KI-Emotionsforscher zusammen: "[T]he question is not whether intelligent machines can have any emotion, but whether machines can be intelligent without any emotions." ([Min86], S. 163).

Mittlerweile hat sich die informatische Emotionsforschung so weit entwickelt, dass man nicht mehr über die mögliche Übertragbarkeit oder Modellierbarkeit der Emotion spricht, sondern immer stärker das "natürliche" Phänomen Emotion in den Fokus informatischer Forschung rückt.

Emotionale Agenten Angesichts der in der Informatik in letzter Zeit deutlich gewordenen Tatsache, dass die Anwendung der klassischen deliberativen KI-Methoden in dynamischen, unsicheren, ressourcenbeschränkten Umgebungen kritisch wird, die kognitive Grundlage also nicht ausreicht, um das Potenzial biologischer und sozialer Systeme in Agentensystemen nachzubilden (vgl. [Ell92], [Can03], [Vel99]), hat das Interesse an emotionalen Mechanismen in den letzten 15 Jahren deutlich zugenommen. Diese werden heute neben Kognition als zentrale Elemente des menschlichen "Verhaltensapparats" angesehen: "[E]motions are now seen as a key element in successful coping with a non-deterministic, dynamic, and social environment." [SP00].

Aktuelle Arbeiten basieren konzeptionell in überwiegendem Maße auf Agentensystemen, die auf Techniken der Künstlichen Intelligenz zurückgreifen. Die Konzeption und Entwicklung von Agentenarchitekturen, die Emotionen erkennen, darstellen oder generieren können, ist in den letzten Jahren zu einem bedeutenden und schnell wachsenden Gebiet innerhalb der Agentenforschung geworden. Unter dem Begriff *emotionale Agenten* [Bat94] werden in dieser Arbeit Agentenarchitekturen verstanden, die durch ihren Aufbau zur Generierung, Verarbeitung und / oder Kommunikation von Zuständen / Prozessen fähig sind, die biologisch / psychologisch definierten Emotionen funktional gleichwertig erscheinen.

Besonders im Hinblick auf die Modellierung von Handlungs- und Entscheidungsprozessen, bei der bislang überwiegend auf intentionale BDI-Architekturen zurückgegriffen wurde, ist deutlich geworden, dass rationale Gesichtspunkte alleine nicht genügen, um die Handlungswahl und Entscheidungsfindung von Akteuren bzw. Agenten realistisch zu modellieren. Hier könnte die Emotionsmodellierung insofern Abhilfe schaffen, als die zentrale Bedeutung der Emotion von den heutigen Emotionsforschern gerade in der Organisation und Motivation des Verhaltens gesehen wird:

"Der Handlungsfluß bedarf eines Auswahlsystems, das die Entscheidung zu und zwischen Handlungszielen steuert (*Handlungswahl*), und er bedarf eines Unterbrechungssystems, das bei veränderten Prioritäten schnelles Umschalten ermöglicht. Emotionen beeinflussen die Richtung des Verhaltens und ermöglichen dadurch eine Auswahl von verschiedenen Verhaltensmöglichkeiten, abhängig

1.2. Offene Fragen der informatischen Emotionsmodellierung

von Situation, Zustand und Erfahrungen - sei es, dass bestimmte Verhaltensweisen durch positive Emotionen verstärkt und negative Emotionen gehemmt werden, sei es, dass zuvor erfahrene Emotionen vorweggenommen werden. Reize erhalten dadurch eine Bedeutung: Eigene Verhaltensweisen, Handlungsfolgen oder Sachverhalte in der Umwelt werden emotional markiert und damit ein Bestandteil der inneren Repräsentation.“ [Vaa03].

Emotionen sind also “mächtige Motivatoren künftigen Verhaltens. Sie bestimmen ebenso den Kurs des Handelns von einem Moment zum nächsten, wie sie die Segel für langfristige Ziele setzen.“ [LeD98]. In Bezug auf allgemeine informatische Kernprobleme wie Performanz, Zuverlässigkeit, Kontrollierbarkeit, Adaptivität oder Effizienz verspricht man sich von der Analyse und Modellierung der Emotionen neue, mächtige Heuristiken zum Problemlösen: “We find close parallels between our resource-adapting mechanisms and emotional models in the life sciences. Indeed, most of the virtues that are cited for emotional intelligence in people are precisely the advantages we seek to obtain with the resource-adapting mechanisms in our software.“ ([Bur98], S. 1).

Neben interindividuellen Emotionseigenschaften lassen sich intraindividuelle Einflüsse von Emotionen auf der Ebene der sozialen Interaktion und Kommunikation identifizieren. Durch Einbeziehung neuer Erkenntnisse aus den psychologischen und (vor allem) soziologischen Emotionstheorien möchte man die Benutzerakzeptanz und Usability der Userinterfaces steigern und die Interaktion mit dem Benutzer auf einer erweiterten wissenschaftlich fundierten Grundlage gestalten. Benutzungsmodellierung unter Berücksichtigung des emotionalen Aspektes könnte zur Entwicklung von anpassungsfähigen und psychologisch adäquaten Lern- und Tutoriensystemen führen (vgl. [Ell97], [PPB⁺04] bzw. [PPP07] für eine Übersicht). Daneben spielen die sozialen Aspekte der Emotionalität möglicherweise für die Ausgestaltung der Kontrollstrukturen in einem natürlichen oder künstlichen Agentensystem eine grundlegende Rolle ([SP00],[vSM04], [MG01]).

In der aktuellen Forschung gibt es sehr unterschiedliche Anwendungsfelder für den Einsatz von emotionalen Mechanismen innerhalb eines Agentensystems. Sie werden z.B. als Steuerungsmechanismen eingesetzt ([Sim67], [DM85], [Slo93]), als Heuristiken bei der Aktionswahl und beim Entscheiden ([Can03], [MG09], [GM05]), um Suchkomplexität beim Planen zu reduzieren ([DP05], [Bac02], [Bac08], [DSD01]), die Effizienz der Entscheidungsfindung zu steigern ([Vel97], [Vel99]) oder das Ressourcenmanagement in Multiagentensystemen zu steuern ([Aub05], [Aub09]).

Die überwiegende Anzahl der aktuellen Arbeiten konzentriert sich auf die Nachbildung von Emotionen oder emotionalem Ausdruck (Gestik, Mimik usw.) in künstlichen Charakteren zur Verbesserung der Mensch-Computer-Interaktion. Diese und weitere Anwendungsfelder für emotionalen Agenten werden in Kapitel 2 näher beschrieben.

1.2. Offene Fragen der informatischen Emotionsmodellierung

In den letzten Jahrzehnten wurden in der Informatik eine Vielzahl von Anwendungen entwickelt, die mit einem Emotionskonzept operieren und es zu modellieren bzw. zu im-

1. Einleitung und Übersicht

plementieren versuchen. Hinsichtlich ihrer grundlegenden Annahmen, ihrer Terminologie, ihrer Gültigkeitsgrenzen, ihres Erklärungspotenzials und ihrer empirischen Überprüfung divergieren diese *informatischen Emotionsmodelle* sehr stark (siehe z.B. [TPP03] für eine Übersicht). Dabei wird ein sehr breites Spektrum von Konzepten als “Emotion“ beziehungsweise als “Emotionsmodell“ bezeichnet. Die intuitive Vorgehensweise und die mangelnde theoretische Fundierung solcher Modellierungsansätze ist vor allem auf fehlende Grundlagenforschung und mangelnde theoretische Verankerung der Emotionsproblematik in der Informatik zurückzuführen.

In diesem Zusammenhang ist eine Konsolidierung wünschenswert, für die in dieser Arbeit eine erste Grundlage geschaffen werden soll. Durch die Auswertung von existierenden Emotionstheorien im Hinblick auf ihre Verwendbarkeit als Lösungsangebot für spezifische informatische Fragestellungen bzw. Anwendungen sollen in dieser Arbeit Theorien, Definitionen und Konzepten strukturiert zusammengestellt und in einen Bezug zu entsprechenden Anwendungsgebieten für Emotionen in der Informatik gesetzt werden. Auf der allgemeinen Ebene besteht das Ziel dieser Arbeit somit darin, die gemeinsame Schnittmenge von Informatik und Emotionsforschung zu durchleuchten und zu systematisieren.

Die Vorarbeit (siehe [Fix04]) hat diverse offene Fragen und Probleme der zeitgenössischen Emotionsmodellierung erkennen lassen, die im Folgenden zusammengefasst werden.

- **Terminologieunschärfe:** Eines der grundlegendsten Probleme der modernen Emotionsforschung stellt die Definition des Forschungsgegenstandes dar, da die Emotionsdefinitionen verschiedener Emotionstheorien teilweise sehr weit auseinanderliegen. Die Schwierigkeiten der emotionstheoretischen Diskussion um den Emotionsbegriff spiegeln sich in der informatischen Emotionsmodellierung wieder. Da selbst das Phänomen natürlicher Emotionen keineswegs eindeutig definiert werden kann (vgl. Kapitel 5 bzw. [Fri09] für eine Übersicht), gibt es umso unterschiedlichere Auffassungen, wenn es um die Beschreibung bzw. Definition künstlicher Emotionen (“artificial emotions“) geht (vgl. [Bac02], [Pic03], [SS02]). So wird ein sehr breites Spektrum von Konzepten als eine Emotion beziehungsweise als ein Emotionsmodell bezeichnet, so dass nicht automatisch klar ist, welche mentalen Phänomene in dem jeweiligen Emotionsmodell als Emotionen verstanden werden. “Like for its ‘natural’ counterpart, there are many dissimilar notions, which focus on different aspects, typically depending on the function authors attribute to emotion, or on the application within the project.“ ([Bac02], S.4). Die unterschiedlichen Emotionsdefinitionen und das Fehlen einer allgemein anerkannten Terminologie erschweren den Architekturvergleich innerhalb der informatischen Emotionsforschung.
- **Fehlen einer integrativen theoretischen Grundlage:** Eine weitere Herausforderung für die Informatik stellt das Fehlen einer einheitlichen theoretischen Grundlage für informatische Emotionsmodellierung dar. Das interdisziplinäre Forschungsfeld “*Affective Sciences*“ [DSG03] ist trotz (oder gerade wegen) des kontinuierlichen Wachstums in den letzten Jahrzehnten sehr fragmentiert. Es gibt zahlreiche Emotionstheorien, -konzepte und -modelle, ohne dass die Integration der Grundlagen in einer allgemein akzeptierten Emotionstheorie erfolgte. Bislang existiert keine integrative Emotions-

1.2. Offene Fragen der informatischen Emotionsmodellierung

theorie, die eine umfassende Beschreibung, Konzeptualisierung und Klassifikation der verschiedenen Faktoren liefern könnte, die bei der Emotionsgenese und Emotionsexpression eine Rolle spielen. Vielmehr steht eine Vielzahl von Emotionstheorien aus verschiedenen Forschungsgebieten nebeneinander, die jeweils nur Teilaspekte der Emotionalität betrachten. Um die bestehenden informatischen Modelle um den Faktor Emotionalität zu ergänzen, ist jedoch eine integrative Emotionstheorie notwendig, die sowohl psychologische Erklärungsansätze als auch soziale Aspekte der Emotionen berücksichtigt und sie idealerweise in eine wechselseitige Beziehung setzt.

Die zeitgenössische Emotionsforschung hält weder eine einheitliche, interdisziplinäre Theorie der Emotion, noch eine allgemeine Emotionsdefinition als Grundlage für die informatische Modellbildung bereit. Bedingt wird dies u.a. durch die unterschiedlichen Wissenschaftskulturen in den verschiedenen Disziplinen, in denen Emotionen untersucht werden, wie z.B. Psychologie, Soziologie, Neurowissenschaften, Kognitionswissenschaften oder Medizin.¹ Wegen des Fehlens einer solchen integrativen Theorie sind die informatischen Emotionsforscher dazu gezwungen, entweder nur Teilaspekte der Emotion zu modellieren, oder aber auf eigene, meist intuitive, theoretisch weniger fundierte und empirisch nicht überprüfte Modelle zurückzugreifen.

- **Forschungsfragmentierung** Mit der Problematik der fehlenden Definition von Emotion und einer fehlenden theoretischen Fundierung hängen noch weitere Probleme, wie die zunehmende Forschungsfragmentierung und eine konzeptuelle Verwirrung auf Seiten der informatischen Emotionsmodellierung zusammen (vgl. [SS02]). Aufgrund dieses Mangels einer gemeinsamen emotionstheoretischen Grundlage, operiert die Forschung mit verschiedenen mentalen Konzepten von Emotion, gegensätzlichen Zielsetzungen, diametral unterschiedlichen Vorannahmen und Terminologien sowie differierenden methodologischen Vorgehensweisen.

“Some researchers, for instance, see emotions as special kinds of motivations (e.g., [BB05]), whereas others draw a clear distinction between motivations and emotions (e.g., [Can97]). Such subtle changes in terminology have the unhappy consequence that the research community as a whole can easily lose track of what its various members are talking about. This is not only undesirable for any discipline, but may eventually culminate in a terminology so conflated that claims involving any of its terms are nearly meaningless.“ ([Sch02], S. 3).

Die Notwendigkeit, durch konzeptuelle Klärung und theoretische Fundierung der Modelle die Vorherrschaft eines Ad-hoc-Prinzips bei der informatischen Emotions-

¹Im Folgenden werden die Emotionstheorien, die aus diesen Disziplinen stammen, vereinfachend als “sozialwissenschaftlichen Emotionstheorien“ bezeichnet. Diese Bezeichnung ist sicherlich formal nicht korrekt, da z.B. die Psychologie eher als naturwissenschaftliche Disziplin zu betrachten ist, ganz zu schweigen von der Kognitionswissenschaft oder der Medizin. Nichtsdestotrotz wird hier das Adjektiv “sozialwissenschaftlich“ verwendet, wenn zusammenfassend von emotionstheoretischen Grundlagen gesprochen wird, die aus anderen Wissenschaften als der Informatik stammen, um diesen Unterschied zu betonen.

1. Einleitung und Übersicht

modellierung zu reduzieren, wurde von renommierten Emotionsforschern, wie Aaron Sloman, wiederholt betont: “Because of this conceptual mess, anyone can produce a program, label some component of it the emotion module and proudly announce that they have developed a robot or software agent that has emotions. It will be hard to argue against such claims when there is no agreement on what emotions are.” ([Slo03], S. 38; vgl. auch [SS02]).

- **Fehlende Formalisierung der sozialwissenschaftlichen Emotionstheorien.**

Um eine informatische Umsetzbarkeit der sozialwissenschaftlichen Emotionstheorien zu ermöglichen, müssen die wesentlichen theoretischen Konzepte und Abläufe eindeutig formal definiert sein. Bei vielen Theorien fehlt jedoch eine solche Formalisierung: Die formulierten Sachverhalte werden vielerorts unpräzise und unvollständig beschrieben. Einige Theorien sind wiederum nur schwer operationalisierbar, weil eine eindeutige Beschreibung der mit Emotion zusammenhängenden Prozesse fehlt. Ferner sind sozialwissenschaftliche Emotionstheorien für Informatiker genauso schwierig zu erfassen, wie z.B. die Organisations-, Gesellschafts- und Spieltheorien der Soziologie, die im Rahmen der Sozionik-Forschung für eine informatische Modellierung nach und nach appliziert werden mussten. Ebenso muss für den hier behandelten Problemzusammenhang das Anwendungswissen im Bereich der Emotionstheorien erst erworben werden. Der Mangel dieses Anwendungswissens in Bezug auf die sozialwissenschaftlichen Emotionstheorien auf Seiten der Informatik führt einerseits zum Missverständnis bestimmter Konzepte, andererseits aber auch zur Entwicklung eigener intuitiver Emotionsmodelle.

- **Fehlen einer einheitlichen Beschreibungstechnik für emotionale Konzepte:** Aus den beschriebenen Problemen ergibt sich unmittelbar der Bedarf an einer Modellierungstechnik für emotionale Konzepte. Das Fehlen einer gemeinsamen Modellierungssprache zur Beschreibung der Emotionstheorien gehört zu den wesentlichen Problemen der zeitgenössischen Emotionsmodellierung (vgl. [SS02], [Slo04], [Sch02], [GM09a]). Um einerseits eine konzeptuelle Klärung und gegenseitige Anschlussfähigkeit der Emotionstheorien zu gewährleisten, andererseits existierende informatische Emotionsmodelle auf einer abstrakten, architekturunabhängigen Ebene beschreiben zu können, braucht man eine universelle Modellierungstechnik. “We need a framework that explains how there can be all the diverse phenomena that different people refer to when they talk about emotions and other affective states and processes.” ([Slo03], S. 35).

Besonders erstrebenswert wäre eine Repräsentationstechnik, die sozialwissenschaftliche Emotionstheorien und informatische Emotionsarchitekturen integriert. Mit der selben Modellierungstechnik sollen sich ferner sowohl emotionale Abläufe auf der Makroebene der soziologischen Strukturen, als auch psychologische und neurobiologische Prozesse auf der Mikroebene darstellen lassen. Gleichzeitig muss diese Repräsentationssprache eine formale Spezifikation der emotionalen Abläufe und ihre Abbildung in einem informatischen System ermöglichen. Ferner soll sie intuitiv und leicht verständlich sein, um zur Vereinfachung der interdisziplinären Forschungsdiskussion, der Überwindung der

1.2. Offene Fragen der informatischen Emotionsmodellierung

Terminologieunschärfe und der Integration verschiedener Emotionstheorien beitragen zu können (vgl. [Sch02], [Slo04], [GM09a], [MFvLvS06]).

- **Schwierigkeit des Architekturvergleichs und der Architekturklassifizierung:** Aaron Sloman sieht die Notwendigkeit für ein Übereinkommen der KI- und VKI-Forscher hinsichtlich weitgehend akzeptierter Termini für die Architekturbeschreibung, um eine gemeinsame Kommunikations- und Diskussionsgrundlage zwischen verschiedenen Forschungsfeldern innerhalb von KI und VKI zu schaffen. Mit einem gemeinsamen terminologischen Rahmenwerk zur Klassifizierung und Gegenüberstellung verschiedener Architekturen könnte man eventuell der starken Fragmentierung der Forschung, der “Erfindung“ bereits existierender Architekturen und Mechanismen und dem terminologischen Durcheinander entgegenwirken. In diesem Zusammenhang argumentieren Michael Scheutz und Aaron Sloman konsequent für eine gemeinsame Beschreibungstechnik zur Repräsentation und Klassifizierung emotionaler Agentenarchitekturen: “If people at least try to work with a common framework for talking about architectures and the states they support, it may be possible for more collaboration and more critical comparative analysis to accelerate progress.“ ([Slo04], S. 6; vgl. auch [Sch02], [SS02], [Slo03]), [GM09a], [GM09b]).
- **Vernachlässigung soziologischer Grundlagen:** Von besonderer Bedeutung im Kontext der Implementierung von Emotionen in Multiagentensystemen (MAS) ist die Auseinandersetzung mit soziologischen Grundlagen. Die Vielzahl soziologischer Arbeiten, die inzwischen international zum Thema Emotionen erschienen ist, bedarf jedoch einer Systematisierung und Konzeptualisierung ([vSvL05]). Dies ist einerseits notwendig, um zu einem konsistenten und umfassenden Bild derjenigen sozialen Strukturen zu gelangen, von den Emotionen maßgeblich beeinflusst und ggf. durch die sie bestimmt werden, andererseits, um den Einfluss des Sozialen an die Ergebnisse anderer Disziplinen anschlussfähig zu machen. Bislang werden sowohl in Bezug auf eine umfassende Konzeptualisierung von Emotionen als auch im Hinblick auf das Verhältnis von Sozialität und spezifischen Emotionen, kaum Querbezüge zu wegweisenden Ergebnissen anderer Disziplinen (Psychologie, Neuro- und Kognitionswissenschaften) hergestellt.

Weil die entsprechende theoretische Grundlage fehlt, weisen viele Emotionsmodelle der Informatik eine einseitige oder nur unzureichende sozialwissenschaftliche Fundierung auf. Insbesondere wird die soziale Komponente der Emotionalität oftmals vernachlässigt: Zwar ist die Auseinandersetzung mit soziologischen Grundlagen bei der Behandlung “klassischer“ sozialer Phänomene wie Interaktion, Kooperation, Koordination, Konflikt, Vertrauen, Verhandlung, Emergenz, etc. weitgehend anerkannt (vgl. [Gas93], [Mal98]). Dass jedoch auch Emotionen maßgeblich von Sozialität beeinflusst werden und selbst wiederum Sozialität beeinflussen, ist bislang nicht berücksichtigt worden. Nach wie vor dominieren bei der Entwicklung emotionaler Systemmodelle kognitionspsychologische Emotionstheorien - auch dann, wenn es sich um künstliche Sozialsysteme, beispielsweise auf Basis von Multiagentensystemen handelt (z.B. [Ell92], [Bat94]).

1. Einleitung und Übersicht

Um Emotionen in (Multi-)Agentensystemen sozialwissenschaftlich fundiert modellieren zu können, benötigt die Informatik eine interdisziplinär verortete Emotionstheorie, die psychologische, soziologische und neurobiologische Grundlagen der Emotion aufgreift und zueinander in Beziehung setzt. Viele Probleme der informatischen Emotionsmodellierung lassen sich auf das Fehlen einer solchen integrativen theoretischen Grundlage sowie gemeinsamer Repräsentationstechniken für emotionale Prozesse zurückführen. Diese grundlegenden Probleme bei der Emotionsmodellierung fassen wir als Motivation für die vorliegende Arbeit auf.

Agentenorientierte Petrinetze In den Projekten der Arbeitsgruppe Theoretische Grundlagen der Informatik (TGI) an der Universität Hamburg werden seit längerer Zeit Petrinetze bzw. Referenznetze zur Darstellung von Agentenarchitekturen erfolgreich eingesetzt ([MW97], [Kum02]). Zudem wurde die petrinetzbasierte Multiagentenarchitektur MULAN (Multi-Agenten-Netze) [Röl04] entwickelt, die eine einheitliche Modellierung und den Vergleich verschiedener Architekturen ermöglicht. Nicht nur die Repräsentation von internen Prozessen innerhalb der Agentenarchitektur, sondern auch die Repräsentation der sozialen Prozesse in Agentengesellschaften kann mit MULAN-Konzepten erfolgen. In [vLMV03] wurden Vorschläge zur Modellierung soziologischer Theorien vorgestellt. Dabei wurden Petrinetze als Modellierungstechnik in der Sozionik erfolgreich eingesetzt. Gesellschaftstheorien der Soziologie wurden im Hinblick auf ihre Kernaussagen analysiert und mit Petrinetzen bzw. Referenznetzen modelliert. Des Weiteren wurde ein erster Vorschlag für eine Agentenarchitektur vorgestellt, die unmittelbar soziale Aspekte mit Hilfe der konzeptionellen Plattform MULAN reflektiert.

Neben der Berücksichtigung der rationalen Modelle wurde auch an der Integration emotionaler Konzepte gearbeitet ([vS04], [MvS00], [MvS01b], [MvS01a], [MvS02], [vSvL05], [vSM04], [vSMFvL06], [Fix04]). Aufgrund der Hypothese der wechselseitigen Abhängigkeitsbeziehung zwischen Agenten-internen (psychologischen) Emotionsphänomenen und externen (soziologischen) Emotionseinflüssen auf der Ebene der Gesellschaft ([vSM04], [vSMFvL06], [SP00]) erscheint eine gemeinsame Repräsentationstechnik zur Formalisierung dieser verschiedenartigen Prozesse besonders erfolgsversprechend, zumal sie auch die Verifizierung der entstehenden Modelle ermöglicht.

Das Fehlen der Repräsentationsmöglichkeit für Emotion macht die Simulation emotionaler Prozesse bei der Validierung sozialwissenschaftlicher Theorien sowohl akteurintern als auch auf der Ebene der Gruppe unmöglich. Sobald die MULAN-Architektur die Möglichkeit der Emotionsrepräsentation auf beiden Ebenen bietet, können nicht nur soziologische (bzw. sozionische) Fragestellungen, sondern auch emotionspsychologische und insbesondere emotionssoziologische Theorieaspekte mit MULAN nachgebildet und simuliert werden.

1.3. Ziele der Arbeit

Ausgehend von den beschriebenen Problemen werden die Ziele der vorliegenden Arbeit gesetzt. In dieser Untersuchung wird der Versuch unternommen, Konzepte aus der Welt der Agenten- und Multiagentensysteme mit sozialwissenschaftlichen - und dabei insbeson-

dere emotionstheoretischen Kategorien - zusammenzubringen. Dieser Zusammenschluss erfolgt auf der Ebene der Modellierung, indem sowohl die Agentenkonzepte als auch die sozialwissenschaftlichen Theorieaussagen mit dem grafischen Formalismus der Petrinetze dargestellt werden. Hierzu werden zwei bereits etablierte Modellierungsansätze benutzt: Modellierung von Agenten und Agentensystemen mit Agentennetzen ([Röl04]) und Modellierung von sozialwissenschaftlichen Theorien nach einem sozionischen Modellierungsansatz ([vLMV03]).

Das Ziel der vorliegenden Arbeit besteht darin, durch eingehende Untersuchung der psychologischen, soziologischen und neurowissenschaftlichen Emotionstheorien eine theoretische Grundlage für die Agentensystemoptimierung durch Emotionen zu erarbeiten, um auf dieser Basis einen gemeinsamen konzeptuellen und terminologischen Rahmen für existierende informatische Emotionsarchitekturen zu entwickeln.

Um herauszufinden, welche prinzipiellen Aspekte der Emotion für die Informatik im Allgemeinen und speziell für die Agententechnologien von Interesse sein könnten, müssen zuerst die entsprechenden Grundlagen erforscht werden. In dieser Arbeit sollen die grundlegenden Motivationen und Probleme, die mit der Emotionsmodellierung in Agentensystemen zusammenhängen, gesammelt und diskutiert werden.

Deshalb ist das Forschungsinteresse bei der Aufbereitung der theoretischen Grundlagen möglichst breit angelegt. Es wird der aktueller Forschungsstand der Emotionsmodellierung ausführlich dargestellt, indem theoretische Überlegungen und praktische Modellierungsansätze aus allen Gebieten der Informatik und den Sozialwissenschaften gesammelt und diskutiert werden. Die Ergebnisse dieser breit angelegten Betrachtung werden im anwendungsorientierten Teil der Arbeit als Leitlinien für einen eigenen Modellierungsansatz im Kontext der Agentensysteme benutzt.

Zuerst sollen die theoretischen Grundlagen der Verwendung des Emotionsbegriffs in verschiedenen Gebieten der Informatik in Form von Definitionen, Theorien und Modellen gesammelt werden. Die herangezogenen Emotionstheorien aus den Gebieten Psychologie, Soziologie und Neurowissenschaften werden im Hinblick auf ihre Anwendbarkeit für die Emotionsmodellierung in verschiedenen informatischen Anwendungsfeldern beurteilt. Insbesondere soll dadurch die theoretische Fundierung für den Einsatz emotionaler Konzepte in informatischen Anwendungen (wie emotionale Agenten, Emotionsmodellierung in einem MAS, Affective Interaction usw.) und damit für das gesamte Gebiet des *Emotional Computing* geschaffen werden. Die wichtigen emotionstheoretischen Konzepte werden mit Hilfe von Petrinetzen modelliert und reformuliert. Diese Vorgehensweise folgt dem sozionischen Modellierungsansatz (siehe [vLMV03]) und liefert die Möglichkeit der Präsentation von Teilaspekten soziologischer, neurowissenschaftlicher und psychologischer Theorien der Emotion in formaler informatischer Darstellungstechnik.

Die Vorteile der Petrinetze gegenüber anderen Modellierungstechniken wurden speziell für die Modellierung soziologischer Theorien in [vLMV03] analysiert und diskutiert. Dort wird die schwache Ausdrucksmächtigkeit bisheriger Modellierungsansätze, die unpräzise Semantik, die nicht vorhandene Werkzeugunterstützung und die fehlende Berücksichtigung der Änderung von Strukturen, die üblicherweise als statisch und damit unveränderbar angesehen werden, kritisiert ([vLMV03], S. 4).

1. Einleitung und Übersicht

Mit Petrinetzen steht eine exakte formale Modellierungstechnik zur Verfügung, die eine adäquate Modellierung von soziologischen Konzepten und Strukturen ermöglicht [vLMV03]. Im Unterschied zu den anderen Automaten-Formalismen erlauben Petrinetze eine direkte Modellierung der Nebenläufigkeit (z.B. für unabhängige Ereignisse oder Prozesse). Durch die operationale Semantik und das weite Spektrum der Analysemethoden erlauben Petrinetze die Evaluation und formale semantische Überprüfung der erstellten Modelle. Petrinetze höherer Ordnung (Referenznetze, Objektnetze) ermöglichen die Repräsentation rekursiver Strukturen und emergenter Prozesse und sind in der Lage, sowohl strukturelle als auch prozessuale Modellperspektiven gleichzeitig darzustellen. Ferner liegt für die zahlreichen Erweiterungen des Basisformalismus ein breites Spektrum an Analysemethoden und aufwendigen Tools zur formalen Evaluierung vor. Wegen der intuitiven und schnell nachvollziehbaren graphischen Notation hat sich der Einsatz der Petrinetze insbesondere im interdisziplinären Kontext bewährt, wo sie eine gemeinsame Kommunikationssprache bildeten (siehe z.B. [KMRV05], [KLR03]). Schließlich wird durch die formale Repräsentation der Emotionsmodelle ihre Integrierbarkeit in die informatischen Modelle und Anwendungen unterstützt.

Zusammen mit ihren Erweiterungen und Analysemethoden stellen Petrinetze somit ein mächtiges Instrumentarium für die Darstellung der dynamischen, hierarchischen und rekursiven Strukturen dar, wie sie in psychologischen, soziologischen und neurowissenschaftlichen Emotionstheorien beschrieben werden. Auf der Erfahrung aus der sozionischen Modellierung von soziologischen Theorien sowie auf den Analyseergebnissen aus [vLMV03] aufbauend, schlagen wir hier daher die Anwendung der Petrinetze / Referenznetze für die Modellierung der Emotionstheorien aus Psychologie, Soziologie und Neurowissenschaft vor. Die Verwendung der Petrinetze als Beschreibungstechnik für die Emotionstheorien und -modelle soll im Laufe der Arbeit evaluiert und praktisch erprobt werden.

Somit werden zweierlei Forschungsziele - konzeptioneller und technischer Natur - verfolgt, die sich jedoch eng aufeinander beziehen. Die allgemeine Betrachtung der vorliegenden Emotionstheorien und ihre Modellierung mit einer informatischen Repräsentationstechnik soll eine emotionstheoretische Basis für die Abbildung der Emotion in informatischen Systemen bereitstellen. Ein Schwerpunkt wird hierbei auf die Erfassung der Funktion der Emotion innerhalb eines Systems sowie auf die Integration und den Vergleich verschiedener theoretischer Ansätze gelegt. Auf der technischen Ebene besteht die Zielsetzung in der Entwicklung eines petrinetzbasierten Rahmenwerks, das als Beschreibungstechnik für die Emotionsmodelle dienen kann. Auf diese Weise wird die Repräsentation sowohl der emotionstheoretischen Grundlagen als auch informatischer Modelle mit der gleichen Beschreibungstechnik ermöglicht.

Die Schwerpunkte der Arbeit liegen insbesondere auf den folgenden Aufgaben:

1. Zusammenstellung und Auswertung der emotionstheoretischen Grundlagen mit dem Ziel, eine Menge an gemeinsamen Definitionen und Konzepten sowie eine begrenzte Anzahl von Modellen für spezifische informatische Fragestellungen bzw. Anwendungen zu erarbeiten. Die Bewertung der Ansätze im Hinblick auf die informatische Modellierung - insbesondere die Ausarbeitung von operationalen Definitionen für verschiedene emotionsrelevante Kategorien - wie Affekte, Emotionen, Stimmungen,

Gefühle, emotionale / affektive Zustände usw. - soll den Kontext für das Studium von affektiven Phänomenen und eine Basis für die Erarbeitung einer angemessenen theoretischen Herangehensweise für die konkreten Anwendungen bilden.

2. Die Modellierung ausgewählter emotionstheoretischen Emotionskonzepte mit der informatischen Darstellungstechnik auf Basis des sozionischen Modellierungsansatzes soll die Transparenz und die Zugänglichkeit der emotionstheoretischen Konzepte für die informatische Modellierung erhöhen und die Analyse der Modelle auf Basis einer formalen Repräsentationstechnik ermöglichen.
3. Der besondere Schwerpunkt liegt auf der Aufbereitung der emotionstheoretischen Grundlagen für die Emotionsmodellierung in Agentensystemen. Deshalb werden insbesondere solche sozialwissenschaftlichen Theorien untersucht, die die Erfassung funktionaler und sozialer Emotionskomponenten ermöglichen, die Emotion also im Kontext ihrer Auswirkungen auf das Funktionieren eines isolierten kognitiven Systems (emotionsbasierte Verhaltenssteuerung) oder mehrerer Systeme (emotionsbasierte Normetablierung) betrachten. Damit wird die Betrachtung soziologischer und neurowissenschaftlicher Ansätze begründet, die den Zusammenhang von Emotion und gesellschaftlichen Normen theoretisch fundiert belegen. Die Umsetzbarkeit dieser theoretischen Ansätze in einem Agentenmodell soll anschließend in den Anwendungsbeispielen demonstriert werden.
4. Unter Berücksichtigung der emotionstheoretischen Grundlagen wird schließlich auf Basis der Agentenarchitektur MULAN eine Referenzarchitektur zur Modellierung von emotionalen Agenten (E-MULAN) konzipiert. Mit der Referenzarchitektur werden ausgewählte Anwendungsszenarios beispielhaft modelliert und mit Referenznetzen simuliert.

Die vorliegende Arbeit zielt nicht vordergründig darauf ab, die zentrale Hypothese der informatischen Emotionsforschung zu überprüfen bzw. in Frage zu stellen, dass die natürlichen emotionalen Mechanismen, wie sie in sozialwissenschaftlichen Theorien formuliert werden, bei entsprechender Modellierung / Umsetzung in informatischen Systemen zur Lösung der informatischen Problemen beitragen können. Vielmehr soll versucht werden, durch ein konzeptuelles und terminologisches Rahmenwerk für die Klassifikation und den Vergleich von sozialwissenschaftlichen Emotionstheorien bzw. informatischen Emotionsarchitekturen zur Transparenz und dadurch zur Effizienz der informatischen Emotionsforschung beizutragen.

1.4. Aufbau

Die Arbeit gliedert sich auf der größten Ebene in drei Teile:

- (1)Einführung in die Grundlagen und den aktuellen Forschungsstand
- (2)Analyse der sozialwissenschaftlichen Emotionstheorien und informatischen Emotionsmodelle
- (3)Vorstellung des MULAN-basierten Referenzmodells zur Emotionsmodellierung E-MULAN und der Anwendungsbeispiele.

1. Einleitung und Übersicht

Im ersten Teil erfolgt eine Übersicht über die Anwendungsgebiete für “Emotionen” in der Informatik sowie eine Darstellung der aktuellen Forschungsschwerpunkte der informatischen Emotionsforschung (Kapitel 2). Ferner gibt es in diesem Teil eine kurze Einführung in Agenten- und Multiagentensysteme (Kapitel 3), die benutzten Petrinetz-Formalismen und insbesondere in das Multiagentensystem MULAN (Kapitel 4), die zusammen die Grundlage für die im Teil III diskutierten Konzepte bilden.

Der zweite Teil der Arbeit widmet sich der Aufbereitung der theoretischen Grundlagen der Emotionsproblematik für die Informatik. In diesem Teil wird eine Zweiteilung vorgenommen, bei der sozialwissenschaftliche Emotionstheorien und informatische Emotionsmodelle separat betrachtet werden. Ein wesentliches Ziel dieser Betrachtung besteht in der Bestimmung der Angemessenheit der jeweiligen Emotionstheorien und -modelle für die konkreten informatischen Anwendungen. Zuerst werden in Kapitel 4.5 die Grundzüge des sozionischen Modellierungsansatzes nach [vLMV03]) dargestellt und die Anwendung dieses Ansatzes für die Modellierung der Emotionstheorien diskutiert. Ferner wird eine Menge von Standardkriterien für den Vergleich der Emotionstheorien bezüglich der zugrunde liegenden Definitionen, Komponenten und Prozesse definiert. Anhand dieser Kriterien werden die Emotionstheorien in Bezug auf bestimmte Charakteristiken, wie Formalisierbarkeit, Operationalisierbarkeit, Erklärungspotenzial, Exaktheit der Emotionsdefinition etc. untersucht und verglichen. In Kapitel 5 werden die sozialwissenschaftlichen Emotionsdefinitionen zusammengetragen und diskutiert, mit dem Ziel, Emotionen zu anderen emotionsrelevanten Kategorien – wie Reflexen, Trieben, Affekten, Stimmungen und Gefühlen abzugrenzen bzw. in Beziehung zu setzen und eine speziell im Kontext eines technischen Systems tragfähige operationale Emotionsdefinition zu formulieren. In Kapitel 6 wird durch die Zusammenstellung der existierenden Emotionstheorien der Psychologie, der Soziologie und der Neurowissenschaft ein möglichst umfassendes Bild des emotionstheoretischen Forschungsstandes erarbeitet. Das Kapitel schließt mit einer zusammenfassenden Bewertung der Emotionstheorien im Hinblick auf ihre Verwendung in informatischen Modellen. In Kapitel 7 werden die zurzeit in der Informatik angewendeten Emotionsmodelle in Bezug auf ihre emotionstheoretische Fundierung, die Gültigkeitsgrenzen der zugrunde liegenden Emotionskonzepte, die Modellarchitektur, die praktische Verwendbarkeit, die Implementierbarkeit etc. untersucht.

Im dritten Teil der Arbeit wird unter Berücksichtigung der Ergebnisse der Theorien- und Modellbetrachtung ein Referenzmodell für Emotionsmodellierung auf Basis des referenznetz-basierten Multiagentenarchitektur MULAN entwickelt, das neben der Modellierung der kognitiven auch eine explizite Modellierung der emotionalen Prozesse erlaubt. Das vorgestellte Referenzmodell soll deshalb einerseits als Modellierungstechnik für den Entwurf von emotionalen (Multi-)Agentensystem-Architekturen eingesetzt werden können, andererseits ist seine Verwendung als konzeptuelles und terminologisches Rahmenwerk für die Klassifikation und den Vergleich von sozialwissenschaftlichen Emotionstheorien und informatischen Emotionsarchitekturen denkbar. Die Anwendungsmöglichkeiten und Grenzen des Modells werden im Anschluss an die Modellbeschreibung in Kapitel 8 an ausgewählten Anwendungsbeispielen aufgezeigt und diskutiert (Kapitel 9).

Zum Abschluss werden die Ergebnisse der Arbeit übergreifend zusammengefasst und in einem Ausblick weitere mögliche Anschlussarbeiten aufgezeigt und diskutiert.

1.5. Terminologie

In diesem Abschnitt erfolgt eine konzeptuelle und terminologische Klärung der grundlegenden Termini, die in dieser Arbeit benutzt werden. So werden die Begriffe “Agent“, “Architektur“, “Emotion“, “Emotionsmodell“ usw. erläutert und ihre Bedeutung im Kontext der nachfolgenden Betrachtungen definiert.

Agent Unter Agenten werden in Rahmen dieser Arbeit unabhängige Ausführungseinheiten verstanden, die weitestgehend autonom in einer Umgebung agieren. Sie können bestimmte Eigenschaften dieser Umgebung wahrnehmen und in ihr handeln, indem sie ihre internen Ziele verfolgen. Die Definition des Agentenbegriffs wird absichtlich möglichst allgemein gehalten, damit sie sowohl die Softwareagenten aus dem (V)KI-Kontext, als auch die autonomen Systeme aus dem Kontext der Robotik umfasst.

Emotionale Agenten Eine allgemeine Definition eines *emotionalen Agenten* ist in der aktuellen Literatur noch nicht vorhanden. Unter emotionalen Agenten werden im Folgenden Agentenarchitekturen verstanden, die zur Generierung, Verarbeitung und / oder Kommunikation von solchen Zuständen bzw. Prozessen fähig sind, die zu biologisch bzw. psychologisch definierten Emotionen funktional äquivalent sind.

Bei diesen *emotionalen Agentenarchitekturen* kann es sich sowohl um simulierte Softwareagenten in virtuellen Umgebungen als auch um autonome Roboter handeln. Wir werden keine Unterscheidung bei der Betrachtung solcher Agenten machen, da die grundlegenden Eigenschaften und Funktionsweisen dieser beiden Systemtypen prinzipiell gleichartig und für Agentensysteme im Allgemeinen charakteristisch sind. Eine ausführliche Diskussion der charakteristischen Eigenschaften eines Agenten, sowie eine genauere Begriffsbestimmung findet sich in Kapitel 3.1.

Multiagentensystem Bei einem Multiagentensystem (MAS) handelt es sich um ein System aus mehreren gleichartigen oder unterschiedlich spezialisierten handelnden Einheiten, die kollektiv zu effektivem Problemlösen fähig sind (siehe Kapitel 3.4 für eine genauere Definition).

Der Begriff des Multiagentensystems wird auch in der Biologie (natürliche Multiagentensysteme) verwendet, wo er sich auf das kollektive Problemlösen innerhalb einer Gruppe von mehreren Organismen bezieht. Ein Beispiel für biologische Multiagentensysteme sind Ameisenstaaten. Einige der in Ameisenstaaten ablaufenden Algorithmen (Ameisenalgorithmen) stellen heuristische Lösungsverfahren für komplexe Optimierungsaufgaben dar und sind neben ihrer grundsätzlichen Bedeutung innerhalb der theoretischen Biologie auch ein Vorbild bei der Optimierung informatischer Problemlösungsprozesse. Man spricht in der Informatik auch von verteilter künstlicher Intelligenz oder DAI (distributed artificial intelligence).

1. Einleitung und Übersicht

Architektur Im Kontext der (Multi-)Agentensysteme muss zwischen der Architektur eines isolierten Agentensystems im Sinne der KI und eines aus mehreren Agenten zusammengefassten Multiagentensystems im Sinne der VKI unterschieden werden.

Im Rahmen der KI wird durch die Architektur eines Agenten ein kognitives Modell beschrieben, welches seinem Verhalten zugrunde liegt. Verschiedene Architekturen unterscheiden sich darin, auf welche Weise die Agentenziele und die Wissensbasis repräsentiert werden und wie die Aktionen und die Reaktionen des Agenten zustande kommen. Eine Agentenarchitektur beschreibt also das Zusammenspiel der internen Komponenten eines Agenten, das ihn zu seinen Entscheidungen führt.

Unter der *Architektur eines Agentensystems* bzw. der *Agentenarchitektur* werden wir folglich in dieser Arbeit das interne Organisationsprinzip eines isolierten Agenten verstehen, das den Entscheidungsprozess dieses Agenten determiniert. Durch eine solche Definition können so unterschiedliche KI-Konzepte wie die BDI-Architektur oder die Subsumptionsarchitektur unter dem Begriff der Agentenarchitektur gefasst werden.

Emotion Die genaue Definition des Forschungsgegenstandes Emotion ist aus mehreren Gründen besonders schwierig. Wegen der Interdisziplinarität des Forschungsfeldes herrscht in der sozialwissenschaftlichen Emotionsforschung weitgehend Uneinigkeit über eine allgemein akzeptierte Terminologie (vgl. Kapitel 6). Das definitorische Problem tritt insbesondere auch deshalb verschärft auf, weil Emotion in den verschiedenen Disziplinen, die sie zu ihrem Forschungsgegenstand machen, je nach Erkenntnisinteresse aus unterschiedlichen Blickwinkeln betrachtet wird und sie deshalb jeweils anders definiert wird. Aus diesem Grunde gibt es in den Sozialwissenschaften zurzeit keine interdisziplinär akzeptierte Definition der Emotion. Stattdessen findet man in der einschlägigen Literatur eine große Zahl unterschiedlicher expliziter und impliziter Definitionen.

In der Informatik wird Emotion, die in künstlichen intelligenten Systemen verwendet wird, oft als “artificial emotion“ (z.B. [Bac02], S. 2) bezeichnet. Ebenso wie in den Sozialwissenschaften wird dieser Begriff je nach Betrachtungsperspektive sehr unterschiedlich definiert und bezeichnet verschiedenste Phänomene.

In Kapitel 5 werden die sozialwissenschaftlichen Emotionsdefinitionen zusammengetragen und diskutiert, mit dem Ziel, Emotionen gegenüber anderen emotionsrelevanten Kategorien wie Reflexen, Trieben, Affekten, Stimmungen und Gefühlen abzugrenzen. Aus der Überblendung einiger sich nicht gegenseitig ausschließender Definitionen wird eine gemeinsame Grundlage abgeleitet, um eine speziell im Kontext eines technischen Systems tragfähige operationale Definition des Emotionsbegriffs zu formulieren.

Im Folgenden wird dem Ergebnis dieser Betrachtungen vorgegriffen, um bereits an dieser Stelle den Emotionsbegriff in dem Sinne zu definieren, wie er in den nachfolgenden Kapiteln angewendet wird. Wir geben also hier die Emotionsdefinition, die bei der Diskussion in Kapitel 5 schrittweise abgeleitet und begründet wird, ohne Herleitung an.

Speziell für die Zwecke dieser Arbeit werden wir Emotionen als komplexe Mechanismen verstehen, die in einem natürlichen intelligenten System als Resultat der Einschätzung seiner Umweltsituation entstehen und das System in einen

Zustand versetzen, in dem es auf seine Umwelt optimal reagieren kann (*emotionaler Zustand*). In einem emotionalen Zustand verändert sich der Ablauf vieler Systemprozesse wie Wahrnehmung, Aufmerksamkeitszuteilung, Entscheidungsprozesse, Planung und Handlungswahl. Die Funktionalität der Emotionen bei der Veränderung des Systemverhaltens lässt sich nach der hier vertretenen Hypothese in ihren einzelnen Aspekten in technischen Artefakten nachbilden und trägt zu ihrer Optimierung bei.

Emotionsmodell In dieser Arbeit wird konzeptuell zwischen Emotionstheorien und Emotionsmodellen unterschieden. Unter *Emotionstheorien* werden die Erklärungsansätze zum Emotionalitätsphänomen verstanden, die nicht aus der Informatik, sondern aus der sozialwissenschaftlichen bzw. neurowissenschaftlichen Emotionsforschung kommen und theoretisch fundierte Hypothesen mit empirisch bzw. klinisch überprüfbaren Aussagen über die Entstehung, Ablauf und Auswirkungen der Emotionsprozesse in Lebewesen formulieren. Dazu gehören nicht nur die *integrativen* Theorien der Emotion, die die Komplexität und die Multi-Komponentialität des Emotionsphänomens in seiner Gesamtheit zu beschreiben versuchen, sondern auch Erklärungsansätze, die Teilaspekte der Emotionalität beleuchten, z.B. Entstehung von Emotionen [Laz91], [Sch93] bzw. die Rolle der Emotion beim Entscheiden [Dam94], bei der Handlungswahl [Fri07], für die Normbefolgung im sozialen Kontext [Dam03] usw.

Mit *Emotionsmodellen* werden im Folgenden die informatischen Systementwürfe bezeichnet, die, teilweise auf den sozialwissenschaftlichen Theorien gründend, versuchen, Emotionen - je nachdem, wie dieser Begriff in dem jeweiligen Ansatz definiert wird - in einem System zu implementieren. Einem Emotionsmodell kann eine spezifische Emotionstheorie zugrunde liegen (z.B. [DLD⁺01]), die durch die Modellierung überprüft werden soll. Es kann aber auch auf intuitiven, sozialwissenschaftlich nicht fundierten Grundlagen basieren oder eine stark vereinfachende Ad-hoc-Theorie als Grundlage haben, welche aus Fragmenten der sozialwissenschaftlichen Originaltheorien zusammengesetzt ist.

Bei vielen informatischen Emotionsmodellen wird der Anspruch erhoben, aufgrund der Modellierung ausgewählter Aspekte der Emotion neue, wenn auch sozialwissenschaftlich nicht belegte Erklärungsansätze für das modellierte Phänomen darbringen zu können (z.B. [Slo01]). Daher ist die Unterscheidung in Emotionstheorien und Emotionsmodelle nicht immer unproblematisch.

Dörner (2001) sieht den Unterschied zwischen einer Emotionstheorie und einem Emotionsmodell vor allem in der Art der Beschreibung des zu erklärenden Phänomens. Diese ist bei der Emotionstheorie eher abstrakt, während in einem Modell die repräsentierten Phänomene tendenziell analog zu den realen Urbildern abgebildet werden: "Ein wichtiger Unterschied zwischen Modell und Theorie ist, dass eine Theorie eine symbolische Beschreibung eines bestimmten Realitätsausschnittes darstellt, ein Modell hingegen eine (teilweise) analoge Beschreibung ist." ([DLD⁺01], S. 2). "Ein Modell "verdinglicht" bestimmte Ausschnitte einer Theorie. Dazu wird die formale Symbolbeschreibung partiell in ein Objekt umgesetzt, das als Gegenstand in Raum und Zeit realisiert ist (z.B. ein Flugzeugmodell) oder als Prozess Verhalten in der Zeit realisiert (z.B. eine Computersimulation)" (ebd.). Im Falle von

1. Einleitung und Übersicht

Dörners Emotionsmodell, das die Implementierung und Evaluierung einer speziell auf die Umsetzung in einem Computersystem ausgerichteten integrativen Emotionstheorie darstellt, ist dieses Unterscheidungskriterium sicherlich ausreichend. Die Mehrzahl der informatischen Emotionsmodelle implementiert jedoch nur Fragmente von sozialwissenschaftlichen Originaltheorien oder nimmt eine spezifische Kombination von verschiedenen Emotionstheorien als Modellierungsgrundlage. Informatische Emotionsmodelle werden oft keiner empirischen Überprüfung unterzogen und sind nicht automatisch als sozialwissenschaftlich fundiert zu betrachten. Daher soll bei der Unterscheidung zwischen Emotionstheorien und (informatischen) Emotionsmodellen ein spezielles Augenmerk auf die Validität dieser Modelle gerichtet werden.

Um den Unterschied zwischen sozialwissenschaftlichen Emotionstheorien und oft nicht theoretisch fundierten informatischen Erklärungsansätzen bereits bei der Begriffsbildung festzuhalten, werden wir im Folgenden den Begriff *informatische Emotionstheorien* benutzen, um solche Emotionsmodelle zu bezeichnen, die, aus der Informatik kommend, einen eigenen theoretischen Erklärungsanspruch für das modellierte Phänomen erheben.

2. Forschungs- und Anwendungsfelder für emotionale Agenten

Die Idee, dass Emotionen einen integralen Bestandteil natürlicher intelligenter Systeme bilden und deshalb auch Eingang in die Modellierung von KI-Systemen finden sollten, ist nicht neu in der Informatik. Sie geht zurück auf die Arbeiten von [Sim67] und [SC81] sowie [Min86] und [New80] und wurde seit den 1980er Jahren kontinuierlich von unterschiedlichen Forschungsgruppen weiterverfolgt. In [Sim67] wurde Emotion zum ersten Mal zum Gegenstand eines systematischen Modellierungsansatzes kognitiver Prozesse. Einen ähnlichen Ansatz verfolgt Aaron Sloman in seinen Arbeiten. In die gleiche Kategorie sind auch die Arbeiten zur Simulation sozialen menschlichen Verhaltens einzuordnen [Wei66].

Heute lassen sich mehrere unterschiedliche Zweige innerhalb der Informatik identifizieren, die sich mit dem Entwurf und der Implementierung von Systemen beschäftigen, in denen Emotionen nachgebildet oder diese als Teil der Systemarchitektur integriert werden. In diesem Kapitel geben wir eine grobe Übersicht über verschiedene Forschungsrichtungen und Motivationen innerhalb der informatischen Emotionsmodellierung. Ferner werden die wichtigsten Anwendungsgebiete für emotional ausgestattete informatische Systeme vorgestellt.

Grundsätzliche Forschungsfragen Wir werden in dieser Arbeit drei grundsätzliche Zielsetzungen bei der Beschäftigung mit der Emotionsproblematik innerhalb der Informatik unterscheiden (vgl. auch [Weh98]):

1. Emotionen als Mittel der technischen Systemoptimierung und Systemleistungssteigerung,
2. Emotionen zur Verbesserung der Benutzbarkeit von Systemen,
3. Emotionen zur Modellierung und Validierung von soziologischen, psychologischen, kognitions- und neurowissenschaftlichen Theorien .

Die korrespondierenden Forschungsrichtungen setzen sich aus unterschiedlicher Motivation und aus verschiedenen Perspektiven mit der Modellierung und Implementierung der Emotionen in informatischen Systemen auseinander.

Entsprechend des breit angelegten Forschungsinteresses der Arbeit werden wir versuchen, Emotion ganzheitlich zu betrachten und alle drei Forschungsrichtungen zu beleuchten. Jedoch wird die Motivation der funktionalen Systemoptimierung durch Emotionen im Mittelpunkt stehen, weil es sich hierbei um eine informatische Kernmotivation handelt. Von dieser Motivation wird auch die Wahl der Emotionstheorien und Modellierungsansätze in den Kapiteln 6 und 7 geleitet.

2. Forschungs- und Anwendungsfelder für emotionale Agenten

1. **Emotionen als Mittel der technischen Systemoptimierung und Systemleistungssteigerung:** Die informatische Kernmotivation bei der Beschäftigung mit Emotionalität besteht darin, verbesserte technische Systeme zum Nutzen von Menschen zu entwickeln. Um entscheiden zu können, inwiefern das System durch Emotionen verbessert werden kann, muss jedoch zuerst die Problemklassifizierung im technischen Bereich erfolgen, mit dem Zweck, Probleme zu identifizieren, die durch Emotionsmodellierung beseitigt werden können. Dabei lassen sich zwei Klassen von Problemen benennen, die eine Emotionsimplementierung im System notwendig machen:

- a) Dort, wo Emotionen im Anwendungsbereich notwendig sind, geht es bei der Systemoptimierung in Bezug auf Emotionen hauptsächlich um die technische Verbesserung der Emotionswahrnehmung und den Emotionsaustausch mit der Umgebung: Das System nimmt eine Emotion in der Umgebung oder die Umgebung nimmt eine Emotion im System wahr.
- b) Die zweite Klasse der Optimierungsprobleme kann umrissen werden durch die Hypothese, dass soziale menschliche Konstellationen, wie sie in Psychologie, Soziologie oder Neurowissenschaft beschrieben werden, technische Systeme bei entsprechender Umsetzung wesentlich verbessern können.

Bei diesen beiden Problemklassen werden Emotionen zur Koordinierung von Technik eingesetzt, die in ihrem Aufbau verschiedene Phänomene aufweisen, die man mit Emotionen unterstützen möchte. Dabei stellen Emotionen unterschiedliche soziale Metaphern dar, die benutzt werden, um z.B. psychologische, soziologische oder neurowissenschaftliche Phänomene, die aus dem menschlichen Kontext bekannt sind, im technischen System abzubilden.

Klassische, deliberative KI-Methoden werden in dynamischen, unsicheren Umgebungen kritisch. Zeit, Speicher oder Rechenkapazität sind in bestimmten Domänen knappe Ressourcen und die Anwendung deliberativer Methoden führt diesbezüglich oftmals zu inakzeptablen Ergebnissen. Hier können vor allem nicht repräsentationsbasierte, sondern reaktive Mechanismen der Agentensteuerung, wie beispielsweise emotionale Zustände und Prozesse, zur Steigerung der Verarbeitungseffektivität beitragen und angemessenere Ergebnisse liefern (vgl. [Can03], [Sch02]).

Aus den Neurowissenschaften stammende neue Erkenntnisse schreiben Emotionen eine bedeutenden Rolle beim Entscheiden und der Aktionswahl zu: “In nature emotions provide decision state which serves as a context for limiting the scope of search for action selection (LeDoux 1996)” ([TWB07] S. 1537). In Bezug auf informatische Kernprobleme wie Performanz, Zuverlässigkeit oder Effizienz verspricht man sich deshalb von der Analyse und Modellierung von Emotionen neue, mächtige Heuristiken und Ansätze zum Problemlösen. Es wird versucht die für Agenten wesentlichen Mechanismen Handlungswahl, Planen und Entscheiden, Motivgenerierung und Interaktionsfähigkeit mit Hilfe emotionaler Komponenten zu optimieren ([BLR92a], [Vel97], [Vel99], [MG02], [Can03], [GM04], [ADP06], [BA08], [ML09] u.v.a.m.). So werden emotionale Mechanismen eingesetzt, um Steuerungsfunktionen in einem intelligenten System

zu übernehmen (vgl. [Sim67], [DM85]; [Slo01]), um die Aktionswahl zu unterstützen ([MG09], [GM05], [Can03]), um Suchkomplexität beim Planen zu reduzieren ([DSD01], [Bac02], [DP05], [Bac08]), die Effizienz bei der Entscheidungsfindung zu steigern ([Vel97], [Vel98], [GL00]) oder das Ressourcenmanagement in einem MAS zu steuern ([Aub05], [Aub09]).

In der VKI-Forschung rückt Emotion als potenzieller sozialer Steuerungsmechanismus im Hinblick auf Kontrollierbarkeits- und Strukturationsprobleme künstlicher Agentengesellschaften in das Feld des Forschungsinteresses.

Es stellen sich im Bereich der technischen Systemoptimierung und Systemleistungssteigerung durch Emotionen prinzipielle Forschungsfragen, einerseits in Bezug auf die Abgrenzung und Identifikation der zu modellierenden Phänomene, andererseits in Bezug auf ihre adäquate Umsetzung im technischen System. In erster Linie muss gefragt werden, welche psychologischen, soziologischen und neurowissenschaftlichen Theorien vorliegen, die die funktionale Rolle der Emotionen in einem natürlichen System aufdecken. Des Weiteren muss man nach der prinzipiellen Operationalisierbarkeit dieser Theorien und ihrer Umsetzbarkeit in einem technischen System fragen. Diese beiden Fragestellungen werden in der vorliegenden Arbeit schwerpunktmäßig behandelt.

2. Emotionen zur Verbesserung der Benutzbarkeit von Systemen.

Die Motivation, mit Hilfe von Emotionen verbesserte Modelle und Techniken für die Mensch-Computer-Interaktion (MCI) zu entwickeln, bildet den zweiten großen Schwerpunkt innerhalb der informatischen Emotionsforschung. Betrachtet man Informatik als Wissenschaft, die das Themenfeld der Emotion bearbeitet, so bilden Emotionen, die bei dem Benutzer erzeugt bzw. erkannt werden, ebenfalls einen Forschungsgegenstand der Informatik. Die zentrale Hypothese diesbezüglich lautet, dass Benutzerfreundlichkeit und Benutzerakzeptanz des Systems durch die Berücksichtigung des emotionalen Zustandes des Benutzers wesentlich gesteigert werden kann.

Im Bereich der Mensch-Computer-Interaktion werden deshalb neben klassischen ergonomischen Aspekten auch zunehmend die Emotionen der Benutzer betrachtet. Gegenstand dieses als "Affective Computing" [Pic97] bekannt gewordenen Forschungsfelds sind weniger Performanz- und Leistungskriterien im technischen Sinn, sondern vielmehr die sozial-emotionalen Aspekte einer verbesserten Mensch-Computer-Interaktion. Die MCI ist zurzeit das Forschungsfeld mit dem größten Anwendungspotenzial. Emotionale Agenten werden vor allem in der Unterhaltungsindustrie (z.B. als "believable agents" [Bat94]) bei der Schnittstellengestaltung ([AKGR00], [MG03], [PPP07]), in Computerspielen ([MAS02], [ZS11], [EB98], [Bat94], [Rei96]), im Bereich pädagogischer Anwendungen ([ARMK00], [LA08] [KMP01]) sowie zur (web-basierten) Präsentationsunterstützung [ARM98] eingesetzt.

Emotionen werden diesbezüglich als integratives Moment der interpersonellen Interaktion angesehen, an die die Agent-Benutzer-Interaktion angeglichen werden soll. Personifizierte (anthropomorphe) emotionale Agenten in Form von Körpern, Gesicht-

2. Forschungs- und Anwendungsfelder für emotionale Agenten

ter, Tieren, Karikaturen o.Ä. repräsentieren das technische System gegenüber dem Benutzer. Insbesondere ist hierbei die nonverbale Kommunikation von Interesse, d.h. die Emotionsexpression in Form von Mimik oder Gestik.

Im Gegensatz zu den beiden letzten Punkten geht es im MCI-Bereich vor allem um die sozial-interaktiven und kommunikativen Aspekte von Emotionen, wobei emotionale Interface-Agenten nicht notwendigerweise über eine elaborierte, emotionstheoretisch fundierte Repräsentation interner emotionaler Zustände verfügen müssen - von Bedeutung ist vielmehr die konsistente und für den Benutzer nachvollziehbare Manifestation von Emotionen. In diesem Zusammenhang stellt sich die Frage nach den relevanten soziologischen und psychologischen Theorien bzw. Hypothesen, die es erlauben, die Systeme im Hinblick auf die Benutzeremotionen optimal zu gestalten.

Der Modellgegenstand dieses Forschungsbereiches soll dahingehend abgegrenzt werden, dass Emotionsaspekte im Kontext der Informatik, die nicht unmittelbar mit der Programmgestaltung zu tun haben, nicht berücksichtigt werden.

So sollen etwa die Emotionen der Software-Entwickler in einem Entwicklungsprojekt nicht zum Forschungsgegenstand werden. Vielmehr soll in diesem Fall die Einbettung des Software-Entwicklungsprozesses in den organisatorischen Kontext erfolgen und die Abgrenzung zur Arbeitswissenschaft gezogen werden. Die Emotionen, die in einem Software-Entwicklungsprojekt auftreten, sollen dann zum Gegenstand der informatischen Emotionsmodellierung werden, wenn eine Software-Entwicklungsumgebung implementiert wird, die im Sinne der Benutzbarkeitssteigerung die Emotionen der Benutzer (in diesem Fall der Entwickler) miterfassen soll.

3. **Emotionen zur Modellierung und Validierung von soziologischen, psychologischen, kognitions- und neurowissenschaftlichen Theorien.** Aus Wissenschaften wie Psychologie, Neurowissenschaften, Kognitionswissenschaften, Soziologie, Biologie etc., die sich mit der Erforschung des "natürlichen" und sozialen Phänomens Emotion befassen, stammt die Forderung, Theorien mit Hilfe von computerunterstützten Simulationen zu validieren und ihre Visualisierung, Konsistenzprüfung, Formalisierung oder Operationalisierung zu ermöglichen (vgl. [Sch93], [ASSCL95], [Ell93], [FS87], [FM94], [Pfe88]).

Im Gegensatz zur technischen Systemoptimierung durch Simulation des Emotionsprozesses wird bei dieser Art der Modellierung nicht der Emotionsprozess selbst, sondern die Emotionstheorie zum Modellgegenstand. Dabei liegt das Ziel der Modellierung darin, die Möglichkeiten der simulationstechnischen Validierung, Formalisierung, Konsistenzprüfung oder Visualisierung wahrzunehmen bzw. so zu einem besseren Theorieverständnis zu gelangen. Diese Vorgehensweise erscheint insbesondere für bestimmte soziologische Fragestellungen von großem Interesse zu sein, da makrotheoretische Aussagen tendenziell nicht oder nur mit kaum zu rechtfertigendem Aufwand empirisch oder experimentell validierbar sind ([CG95], S.2). Die Resultate der simulationstechnischen Theorienmodellierung sind vor allem außerhalb der Informatik relevant.

Im Zusammenhang mit der Modellierung sozialwissenschaftlicher Theorien stellen sich allgemein folgende Fragen: Welche Anforderungen stellen sich aus der Sicht der zu simulierenden psychologischen, soziologischen oder kognitionswissenschaftlichen Theorie an das System? Welche soziologischen bzw. psychologischen Phänomene bzw. Gegenstände müssen im System beschreibbar sein bzw. was ist der zu simulierende Modellgegenstand? Welche Softwarearchitektur wird benötigt, um die Phänomene beschreiben zu können?

Innerhalb der oben dargestellten grundlegenden Motivation für die Implementierung der emotionalen Mechanismen in Agentensystemen lässt sich die informatische Emotionsforschung in drei wesentliche Forschungsgebiete aufteilen: Künstliche Intelligenz (KI), Verteilte Künstliche Intelligenz (VKI) und Mensch-Computer-Interaction (MCI). Diese Gebiete werden jeweils wie folgt eingegrenzt:

- Das Forschungsgebiet Künstliche Intelligenz (KI) betrachtet Emotion als Mittel zur Performanz- und Effizienzsteigerung isoliert agierender Entitäten (z.B. [Vel97], [Slo01], [Pet03], [PT06], [Can03]).
- Die Verteilte Künstliche Intelligenz (VKI) erforscht die Bedeutung und Funktionalität der Emotion für sozial interagierende Agenten in Multiagentensystemen, verteilten Systemen und künstlichen Gesellschaften (z.B. [Ell92], [BLR92a], [SP00], [GL00], [MG01], [Aub05], [Aub09], [MG09]).
- Schließlich bildet die Erforschung der Rolle der Emotion in der Mensch-Computer-Interaktion und in hybriden Systemen das Hauptinteresse des Forschungsgebietes Mensch-Computer-Interaction (MCI) (z.B. [Bat94], [Pic97], [AKGR00], [Dau00], [PPP07]).

Wir werden die Darstellung der aktuellen Richtungen in der informatischen Emotionsforschung im weiteren Verlauf dieses Kapitels entsprechend dieser Forschungsbereiche strukturieren und die wesentlichen Fragestellungen und Lösungsansätze innerhalb der benannten Bereiche darstellen.

Zuerst werden die wichtigen Forschungsfragen und Herangehensweisen an die Emotionsmodellierung innerhalb der KI dargestellt. Hierbei werden einige wichtige Modellierungsansätze beschrieben und ausgewählte Agentenarchitekturen vorgestellt. Entsprechend erläutern wir im nächsten Abschnitt Fragen und wichtige Agentenarchitekturen aus dem VKI-Bereich. Hier wird auch das Forschungsfeld Sozionik vorgestellt sowie die Einbettung des in dieser Arbeit verfolgten Ansatzes in den allgemeinen Forschungskontext der Sozionik vorgenommen. Im dritten Abschnitt werden Ansätze aus dem MCI-Gebiet vorgestellt, die Emotion zur Optimierung der Mensch-Computer-Interaktion einsetzen. Schließlich wird auf offene Fragen und Probleme der aktuellen informatischen Forschung zur Emotion eingegangen, um daraus Motivationen und Bewertungskriterien für die nachfolgende Betrachtung der emotionstheoretischen Grundlagen in Kapitel 6 abzuleiten.

2.1. Emotion und Künstliche Intelligenz

Das Forschungsgebiet Künstliche Intelligenz (KI) betrachtet Emotion als Mittel zur Performanz- und Effizienzsteigerung isoliert agierender Entitäten. Dazu werden sowohl eingebettete autonome Agenten in Robotern als auch Software-Agenten in simulierten Umgebungen gezählt. Dieser Abschnitt zeigt anhand einiger Motivationen und theoretischer Überlegungen aus der KI-Forschung, warum und wie autonome Agenten mit Emotionen ausgestattet werden sollen und welche Einsatzmöglichkeiten für solche Systeme denkbar sind.

Bei der Realisierung des zielgerichteten selbstständigen Agentenverhaltens greift man auf Mechanismen natürlicher biologischer und sozialer Systeme zurück, weil diese Mechanismen evolutionär erprobt und dafür optimiert sind, die Anforderungen von komplexen Umgebungen durch autonomes und adaptives Handeln des Organismus effektiv zu meistern [Laz91].

In der KI-Emotionsforschung zählt man Emotionen zu solchen biologischen Mechanismen und hält die Inspiration durch natürliche Systeme für ein sehr nützliches Design-Prinzip bei der Entwicklung des emotionalen Apparats von künstlichen Agenten, die mit den gleichen Umgebungen und Aufgaben wie natürliche Organismen konfrontiert werden (z.B. [Can03], S. 115f.): “The existence of affective systems in living creatures under the pressure of evolution suggests that it plays a number of functional roles“ (zitiert nach [ADP06]). Emotion als erprobter natürlicher Mechanismus der Komplexitätsbewältigung in nicht deterministischen, hoch dynamischen Umgebungen ist ein optimales Vorbild für die Verhaltenssteuerung in Softwarearchitekturen: “One aspect of emotions that is of particular interest to us as agent engineers is the fact that emotions and affect provide a glimpse of what could be termed as being really at work in the control of natural beings“ ([Pet03], S. 251).

Emotionsmodelle können somit zur Optimierung der allgemeinen Modelle des rationalen Verhaltens beitragen. Mehrere Autoren haben argumentiert, dass emotionale Einflüsse, die auf den ersten Blick irrational erscheinen, wichtige soziale und kognitive Funktionen haben, die für jedes intelligente System von Nutzen sind ([Min86], [Sim67], [OJL87], [SC81], [Dam94], [GL00]). Z.B. können soziale Emotionen wie Ärger und Schuld auf Mechanismen zurückgeführt werden, die die Gruppeneffizienz durch Minimierung sozialer Konflikte verbessern. Dadurch können auch “irrational choices“ in sozialen Spielen, wie beim Gefangenendilemma, erklärt werden [Fra88].

Emotionen werden bei der Aktionswahl in dynamischen und sozialen Umgebungen und zur Effizienzsteigerung beim Planen eingesetzt [ADP06]. Es wird versucht, die für Agenten wesentlichen Mechanismen der Entscheidungsfindung [Vel99], Motivgenerierung [Can03], Interaktionsfähigkeit und Kommunikation [BLR92a] mit Hilfe emotionaler Komponenten zu optimieren.

Im Zuge der Emotionsmodellierung können, manchmal als Nebeneffekt, neue Mechanismen gefunden werden, die im Allgemeinen für Agentensysteme nützlich sind: So führte Maos Versuch, Ärger zu modellieren, zur Entwicklung eines allgemeinen “social credit assignment“-Mechanismus und dem Modell des sozialen Zwangs [MG04].

Für neuere Ansätze der KI-Forschung ist die *funktionale* Rolle der Emotion innerhalb der Agentenarchitektur charakteristisch. Die funktionale Perspektive impliziert, dass Emotion als eine Lösung für bestimmte Probleme in Bezug auf das optimale Funktionieren einzelner

(oder mehrerer, siehe Abschnitt 2.2) Agenten angesehen werden kann. Sie bildet damit einen Ausgangspunkt für die Erforschung der Rolle der Emotion für die Leistungssteigerung der KI-Systeme.

Offene Forschungsfragen Emotionen werden von der gegenwärtigen informatischen Forschung einerseits als Mittel zur technischen Systemoptimierung und Systemleistungssteigerung gesehen, andererseits zur Simulation und Validierung von sozialwissenschaftlichen Theorien in Informatiksystemen benötigt und schließlich zur Verbesserung der Benutzbarkeit von Systemen eingesetzt.

Bei der Motivation der funktionalen Systemoptimierung durch Emotionen handelt es sich um die informatische Kernmotivation, Performanz, Zuverlässigkeit, Kontrollierbarkeit und Effizienz von Informatiksystemen durch die Nachbildung bzw. Einbeziehung der natürlichen Emotionsprozesse in die bisher weitgehend rational geprägte Agentenmetapher zu steigern.

Die bisher offenen Fragestellungen in diesem Bereich beziehen sich auf die Mechanismen der emotionsbasierten Verhaltenssteuerung in natürlichen Systemen, speziell den Mechanismus des Emotionseinflusses auf Handlungswahl (z.B. [Can03]) und Entscheidungsfindung (z.B. [Vel97]). In Kapitel 6 wird deshalb insbesondere nach den theoretischen Ansätzen gesucht, die zu einer Lösung für diese Fragestellung beitragen können.

2.2. Emotion und Verteilte Künstliche Intelligenz

Multiagentensysteme stellen in gewisser Weise eine Erweiterung des lange vorherrschenden kognitivistischen Paradigmas der KI dar: Mehrere intelligente Systeme (Agenten) werden in ein Multiagentensystem (MAS) eingebunden, das entsprechend leistungsfähiger, aber auch schwieriger zu beherrschen ist.¹ Das Agentenkonzept in der VKI stimmt grundsätzlich mit der KI-Definition überein, jedoch verfügen die Agenten in einem MAS zusätzlich über die Fähigkeit, ihre Aktionen mit denen anderer Agenten zu koordinieren, um übergreifende Probleme zu lösen

Charakteristische Merkmale von Multiagentensystemen sind:

- **Verteiltheit:** Die Lösung eines Problems wird auf mehrere Problemlösungskomponenten verteilt, die jeweils Teillösungen erarbeiten.
- **Handlungsfähigkeit:** Die Problemlösungskomponenten sind relativ eigenständige Akteure, also Agenten, die sich in ihrem Verhalten nach eigenen Handlungsprogrammen richten.
- **Koordination:** Die Agenten müssen ihre jeweiligen Beiträge zu der Problemlösung in geeigneter Weise miteinander koordinieren.

Die besonderen Herausforderungen bei der Konstruktion von MAS liegen im Bereich des Emergenten, also bei solchen Phänomenen, die sich nicht explizit modellieren lassen

¹Eine genauere Begriffsbestimmung sowie Darstellung des aktuellen Forschungsstandes zu Multiagentensystemen erfolgt in Kapitel 3.4.

2. Forschungs- und Anwendungsfelder für emotionale Agenten

(bzw. die nicht explizit modelliert werden sollen), sondern sich aus der Interaktion einzelner Agenten untereinander ergeben (vgl. [Fer95], [Dor97]). Aus diesem Grund gehören Gesellschaftsmetaphern wie Normen, Regeln, Institutionen, Kooperation, Struktur oder Kohäsion zu den interessanten und vielversprechenden Forschungsthemen der VKI.

Das Vorhandensein mehrerer Agenten macht die Simulation sozialer Zusammenhänge möglich. Mit Hilfe von MAS können soziale und soziologische Phänomene modelliert und erforscht werden, wie z.B. Koordination und Kooperation, Steuermechanismen und Regeln der Gesellschaftsstrukturierung, das Mikro-Makro-Problem usw. Dies macht die MAS zu einem wichtigen Werkzeug soziologischer Forschung, vor allem bei der Validierung sozialtheoretischer Modelle.

Das VKI-Forschungsfeld *Sozionik* nutzt Anregungen und Analogien aus der sozialen Welt dazu, "intelligente Technologien zu entwickeln, um damit technische Probleme nach dem Vorbild sozialer Problembewältigung zu lösen und gleichzeitig die soziologische Theoriebildung durch Modellierung und experimentelle Simulation voranzutreiben ([vLMV03], S. 2)". Das primäre Ziel der Sozionik ist die Modellierung künstlicher Agentengesellschaften, und zwar sowohl als Modell menschlicher Gesellschaften als auch als Problemlösungsmethode für komplexe intelligente Prozesse.

Aus der sozionischen Perspektive muss zur besseren Konstruktion und Analyse agentenbasierter sozialer Systeme eine Verschränkung von soziologischen und informatischen Forschungsgrundlagen stattfinden. Dabei werden Theorien und Verfahren der Soziologie und Informatik so verknüpft, dass die Grundlage für ein soziologisch fundiertes Agentenmodell geschaffen wird.

Die ersten Arbeiten zur Verbindung der Konzepte von Emotion mit grundsätzlichen Fragen der VKI-Forschung geben Hinweise darauf, dass dieser Ansatz möglicherweise zur Lösung einiger grundlegender sozionischer Probleme führen kann, z.B. des Problems sozial intelligenter Koordination autonomer Agenten, des Mikro-Makro-Problems oder des Problems der Kontrollierbarkeit von Multiagentengesellschaften ([SP00], [vSM04], [MG01], [MG09]).

Kontrollierbarkeit und Handlungskoordination in MAS Das Problem der Kontrollierbarkeit in Multiagentensystemen stellt eine wichtige Herausforderung für die sozionische Forschung dar. Das Bestreben, auf Seiten der Agenten maximale Autonomie, Flexibilität sowie Handlungs- und Lernfähigkeit zu erreichen, steht der festen Verankerung der Kontrollstrukturen im System entgegen. Dieser Widerspruch ist auch als sogenanntes *Emergenzproblem* bekannt [vSM04].

Ein weiteres sozionisches Problem, die Handlungskoordination autonomer Agenten in einem MAS, kann einerseits auf der Mikroebene aus der Perspektive des einzelnen Agenten angegangen werden, indem die Koordination als in den internen Eigenschaften des Agenten begründet gesehen wird (z.B. kann die Kooperationsbereitschaft durch interne Agenteneigenschaften wie Gutwilligkeit oder Egoismus festgelegt werden). Andererseits können die gesellschaftlichen Strukturen der Makroebene modelliert werden. Diese kann man als allgemeine gesellschaftliche Normen oder Regeln realisieren, nach denen sich alle Agenten richten (*social laws*), oder als Rollen, auf die Agenten mehr oder minder festgelegt sind, bzw. als sprachliche Festlegungen, derer sich alle Agenten in Austauschprozessen

2.2. Emotion und Verteilte Künstliche Intelligenz

bedienen müssen (z.B. Kommunikationsprotokolle). Schließlich gibt es den Weg über Abhängigkeitsbeziehungen, die aus arbeitsteiliger Spezialisierung resultieren.

Man hofft, durch die Anwendung der sozialen Mechanismen, die die Kontrollierbarkeit und Koordination in menschlichen Gesellschaften sicherstellen, die Kontrolle über das Gesamtsystem zu erlangen, ohne die Vorteile der VKI-Systeme (Autonomie, Adaptivität, Flexibilität, Proaktivität) aufgeben zu müssen. Es wird deshalb nach sozialen Mechanismen gesucht, die die Gesellschaft mehrerer autonomer Agenten koordinieren und kontrollierbar machen, ohne die Autonomie eines einzelnen Agenten zu stark einzuschränken. Insbesondere sind die Mechanismen der dynamischen Entwicklung von Kontrollstrukturen in natürlichen Gesellschaften als Vorbild für MAS von Interesse.

Eine mögliche Lösung in diesem Zusammenhang ist die Implementierung eines Systems der geteilten Normen, die nicht durch die Systemarchitektur fest vorgegeben sind, sondern sich dynamisch aus der Interaktion der Agenten (und Benutzer) ergeben (siehe [ST97]). Um diesen Ansatz zu verfolgen, ist es zwingend erforderlich, sich mit den Mechanismen der Normetablierung, -aufrechterhaltung und -befolgung in menschlichen sozialen Systemen auseinanderzusetzen, sie tiefgründig zu erforschen und an den VKI-Kontext anzupassen.

Sowohl theoretische Analysen als auch empirische Untersuchungen zeigen, dass Emotionen eine essenzielle Rolle bei der Etablierung sozialer Normen und Sanktionierung des nichtkonformen Verhaltens spielen ([Els96], [Eli76], [FU04]). Die strukturierende Funktion der Emotionen im sozialen Kontext wird auch durch neueste neurowissenschaftliche Untersuchungsergebnisse gestützt ([Dam96],[Dam03],[EL04]), [R.10]. Laut diesen Ergebnissen treten Emotionen als Mechanismen der Unterstützung von sozial akzeptierten normkonformen Verhaltensweisen und der Bestrafung von abweichendem Verhalten auf([Dam03]). In die gleiche Richtung weisen Untersuchungen, die darauf hindeuten, dass Emotionen eine Schlüsselfunktion bei der Gesellschaftsregulation bzw. bei der Verlinkung der Aktionen eines Akteurs mit dem globalen Systemverhalten zukommt ([vSvL05]).

Soziale Normen und Emotion Die soziologischen und neurowissenschaftlichen Forschungsergebnisse zeigen deutlich, dass die Modellierung bzw. Implementierung der Wechselwirkungen zwischen Emotionen und sozialen Normen zur Etablierung dynamischer und flexibler Kontrollstrukturen in einem Multiagentensystem beitragen kann und deshalb von der VKI-Forschung gründlich untersucht werden sollte. Jedoch hat die Untersuchung der Emotion aus dieser Perspektive bislang nicht den entsprechenden Eingang in die VKI-Forschung gefunden ([vSM04], S. 2f.). Es existieren zwar Ansätze, Emotionen in sozial interagierenden Agenten in Multiagentensystemen zu realisieren ([Ell92], [BLR92a], [Bat94]), jedoch gehen diese Arbeiten auf die strukturierende Bedeutung und grundlegende Funktionalität der Emotionen für die MAS in der Regel nicht ein. So werden in [Ell92] vor allem die kurzfristigen Koordinationseffekte der Emotionen in diadischen Interaktionen untersucht. In [AS96] bzw. [Aub09] wird die Rolle der Emotionen als Operatoren bei der Ressourcenzuordnung thematisiert. Langfristige Strukturierungseffekte der Emotionen auf der Makroebene wurden bei den meisten Ansätzen jedoch nicht betrachtet.

Conte und Castelfranchi ([CC95b]) führten eine Studie durch, in der sie die makrosoziale Funktion der Normen bei der Aggressionskontrolle in einer Agentengesellschaft erforschten.

2. Forschungs- und Anwendungsfelder für emotionale Agenten

In ihrer Untersuchung konnten sie zeigen, dass bei einer “normativen Strategie” (normative strategy), bei der alle Agenten einer Multiagentengesellschaft eine Norm befolgten, Aggression (gemessen an der Anzahl aggressiver Vorfälle zwischen Agenten) wesentlich seltener vorkam als bei einer “funktionalen Strategie” (utilitarian strategy), bei der keine solche Verhaltensnorm existierte ([CC95b], vgl. auch [SP00], S. 4). Diese Untersuchung hat die VKI-Forschung nachhaltig beeinflusst und eine Reihe von weiteren Arbeiten nach sich gezogen ([WW95], [CCP98], [SH99], [MG01], [MG09]). Staller und Petta ([SP00]) kritisieren jedoch, dass weder [CC95b] noch die Autoren der nachfolgenden Forschungsarbeiten das aggressive Verhalten explizit als emotionales Verhalten identifiziert bzw. in den Kontext der Emotionsforschung gestellt haben.

Auf [SP00] geht der Vorschlag zurück, den Zusammenhang von Emotionen und sozialen Normen im VKI-Kontext zu untersuchen. In [SP00] diskutieren die Autoren erstmalig in der VKI-Forschung die theoretischen und praktischen Möglichkeiten und Perspektiven der Implementierung sozialer Normen in Form von agenteninternen Emotionen und kommen zu dem Schluss, dass “the computational study of social norms can profit by modeling emotions among agents in artificial societies” ([SP00], S. 9). Auf diese Hypothese zurückgreifend, wird in Kapitel 6.3.3 untersucht, wie Emotionen in größeren sozialen Einheiten einen robusten Strukturierungs- und Kontrollmechanismus erzeugen und wie sie beim Design von Agenten- und Multiagentensystemen berücksichtigt werden können.

Das Mikro-Makro-Problem In komplexeren und dynamischeren Multiagentensystemen stellt sich die Frage nach dem Verhältnis und der möglichen Verschränkung agentenorientierter und strukturorientierter Ansätze: das sogenannte Mikro-Makro-Problem. Dabei stehen folgende Fragen im Vordergrund: Sind die gesellschaftlichen Strukturen das Resultat individueller Handlungsstrategien? Ist individuelles Handeln umgekehrt determiniert durch gesellschaftliche Strukturen? Oder gelten die beiden Aussagen gleichzeitig?

Um diesen Frage nachzugehen, versucht man, die Mikro-Makro-Zusammenhänge auf den unterschiedlichen Ebenen sozialer Koordination zu rekonstruieren:

- Auf der Ebene der Interaktion definiert man emergente Problemlösungen als nicht direkt auf die einzelnen Agenten zurückführbar, sondern als Ergebnis von Interaktionsprozessen. Um flexible und dynamische adaptive Systeme zu bilden, müssen die soziologischen Modelle ein Höchstmaß an übergreifender Problemlösungskompetenz mit einem Mindestmaß an prädefinierter Intelligenz des Einzelagenten verbinden.
- Auf der Organisationsebene unterscheidet man zwischen hierarchischen und heterarchischen Organisationsformen: Standardisierte Handlungsabstimmung ist effizient, dafür aber unflexibel; hierarchische Arbeitsteilung reduziert Komplexität, kann aber nicht gut mit unvollständiger Information umgehen.
- Auf der Ebene gesellschaftlicher Strukturen stellt sich dann wiederum die Frage, welche soziologischen Mechanismen eine soziale Ordnung innerhalb von Multiagentensystemen implementieren können und wie diese ohne fest programmierte Strukturationsregeln in MAS umgesetzt werden können.

2.2. Emotion und Verteilte Künstliche Intelligenz

Staller und Petta [SP00] diskutieren die Möglichkeit der Implementierung von Emotionen in Multiagenten-Settings zur Konstituierung von sozialen Normen sowie das Potenzial eines solchen Ansatzes für die Lösung des Mikro-Makro-Problems in natürlichen und künstlichen Gesellschaften. Die Autoren schlagen die Anwendung der Einschätzungstheorie der Emotion vor, um die Mikro- und Makroebene zu verlinken: “Macro-level functions of emotions such as sustaining cooperation and reciprocity depend on the micro-level processes of appraisal and the generation of action tendencies“ ([SP00], S. 9).

In [vSM04] wird die These aufgestellt, dass Emotionen als soziale Steuermechanismen innerhalb einer (Agenten-)Gesellschaft fungieren, die Mikro- und Makroebene miteinander verknüpfen. Die Autoren nehmen an, dass die Akteure in einer (homogenen) Gesellschaft aufgrund der Sozialisation innerhalb eines bestimmten sozialen Aggregats über ähnliche Ausprägungen mentaler Objekte, wie soziale Normen und Konventionen, verfügen. Diese im Akteur verankerten sozialen Strukturen sind dann Instrumente, die die Normkonformität des Akteurs und u.a. seinen Umgang mit den eigenen Emotionen und den Emotionen anderer Akteure regulieren. “Auf diese Weise wirkt einerseits die durch ein soziales Umfeld geprägte emotionale Handlungs- und Verhaltenssteuerung direkt in die unmittelbare soziale Umgebung zurück, andererseits stellt die Rezeption einer Emotionsexpression durch andere Akteure in Aussicht, in eben diesen Akteuren “reaktive“ Emotionen hervorzurufen, die eine entsprechend konforme oder konträre Handlungsregulation zur Konsequenz haben“ ([vSM04], S. 8).

Wenn man diesen sozialen Mechanismus auf die Agentengesellschaften überträgt, so bedeutet dies, dass man einerseits soziale Strukturen, wie Normen und Regeln, im Agenten fest verankern muss, so dass sie die Agentenemotionen direkt regeln. Andererseits sollen die Agenten aufgrund der Emotionsexpressionen der anderen Agenten Rückschlüsse auf die soziale Adäquatheit (Normkonformität) ihres eigenen Verhaltens ziehen können und das Verhalten anderer wiederum mittels emotionaler Reaktionen sanktionieren oder unterstützen. Emotionen fungieren auf diese Weise als Instrumente sozialer Kontrolle auf der Mikroebene, indem sie das Handeln des Agenten mitbestimmen, und dienen zugleich der Aufrechterhaltung der Gesamtstruktur der Gesellschaft auf der Makroebene, indem sie normkonformes Verhalten unterstützen [vSM04].

Ein weiteres Anwendungsfeld für Emotionsmodelle stellen Modellierungs- und Simulationsanwendungen dar. So wird für die Modellierung der Gruppenreaktionen in Extremsituationen [SJO⁺02], für die Modellierung von Gruppenkonflikten [MPS04] bzw. für groß angelegte soziale Simulationen eine zusätzliche Fundierung durch Emotionsmodelle benötigt ([GM01], [GM03], [GM05], [MG01], [MG09]).

Offene Forschungsfragen Neuere Ansätze im Bereich der VKI untersuchen die Möglichkeit des Einsatzes von Emotion in Multiagentensystemen zur Lösung grundlegender sozionaler Probleme wie das Mikro-Makro-Problem [vSM04] oder das Problem der Kontrollierbarkeit von Multiagentengesellschaften bzw. die Etablierung sozialer Normen [SP00]. In der vorliegenden Arbeit soll schwerpunktmäßig die Frage nach der Rolle der Emotion bei der Etablierung sozialer Kontrollstrukturen der Makro-Ebene, insbesondere sozialer Normen, betrachtet werden. Im theoretischen Teil der Arbeit (Kapitel 6) wird nach theoretischen

2. Forschungs- und Anwendungsfelder für emotionale Agenten

Ansätzen gesucht, die die Rolle der Emotionen als sozialen Steuermechanismen belegen und den Prozess der Gesellschaftsstrukturierung und -regulation durch emotionale Mechanismen näher spezifizieren, sodass eine emotionstheoretische Grundlage für die Implementation der beschriebenen Mechanismen geschaffen werden kann.

Im Abschnitt 6.3 wird aus den betrachteten Ansätzen ein genereller Mechanismus der Etablierung und Aufrechterhaltung sozialer Normen durch Emotion abgeleitet und in Form von Petrinetz-Modellen dargestellt.

2.3. Emotion zur Optimierung der Mensch-Computer-Interaction

Die MCI ist zurzeit ein schnell wachsendes Forschungsfeld mit dem größten Potenzial an Anwendungen, die durch Emotionsmodellierung grundlegend optimiert werden könnten. Vor allem werden emotionale Agenten in der Unterhaltungsindustrie als *believable agents* oder *lifelike characters* eingesetzt, beispielsweise in Computeranimation und Computerspielen, zur benutzerfreundlicheren Schnittstellengestaltung sowie zur Präsentationsunterstützung (z.B. im Frontend von Webauftritten, Onlineshops usw.). Man erhofft sich, durch den Einsatz der Emotionen in der MCI die Interaktionen mit dem Benutzer einfacher und intuitiver gestalten zu können und in langfristiger Perspektive die Adaption des Systems an den emotionalen Zustand des Benutzers zu erreichen.

Die Untersuchung in [RN96] hat gezeigt, dass die Mensch-Computer-Interaktion im Wesentlichen nach natürlichen sozialen Interaktionsmustern abläuft und von den gleichen Faktoren bestimmt wird wie die natürliche zwischenmenschliche Kommunikation. Es konnte festgestellt werden, dass die sog. emotionale Intelligenz bei der menschlichen Interaktion einen wichtigeren Platz einnimmt als kognitive oder verbale Fähigkeiten. Zur emotionalen Intelligenz gehört in erster Linie die Fähigkeit, die Emotionen der anderen zu erkennen und auf sie angemessen zu reagieren (vgl. auch [Pic03]). Zum Beispiel wäre es im Zusammenhang mit der wachsenden Anzahl von Anwendungen, die den Benutzer mit Hilfsinformationen und Tipps, Empfehlungen und Erinnerungen zu unterstützen versuchen, wünschenswert, dass diese Programme die (emotionale) Reaktion der Benutzer auf plötzliche Unterbrechungen erkunden und ihr Verhalten dieser Reaktion anpassen. “Suppose that a computer starts to give you help at a bad time. You try ignoring, then frowning at, and then maybe glaring at it. The savvy computer infers you do not like what just happened, ceases the interruption, notes the context, and learns from the feedback“ ([PVH01]). Die emotional kompetente Schnittstellen- und Interaktionsgestaltung eröffnet neue Perspektiven auf die Reduzierung von Frustration und Verwirrtheit bei der Interaktion mit dem Computer (z.B. [PMMI03]). Es konnte ferner gezeigt werden, dass die Qualität der Interaktion im Hinblick auf Benutzerfreundlichkeit und Steigerung der Motivation des Benutzers durch den Einsatz von emotionalen Interface-Agenten verbessert werden kann [CK02], [KTB03].

Emotion als Instrument der sozialen Kontrolle wird in Anwendungen eingesetzt, die beanspruchen, menschliches Verhalten zu lenken und zu steuern, z.B. im psychotherapeutischen Zusammenhang ([GM03], [RHA⁺99], [PG08]), in Tutoriensystemen ([LBS99], [RVC03], [Ell97], [MMSL01], [PPB⁺04]) oder Marketing-Systemen ([AKGR00], [LA08], [CBC⁺00]).

2.3. Emotion zur Optimierung der Mensch-Computer-Interaction

Therapeutische Systeme als eine Spezialform von Trainingssystemen versprechen eine der wichtigsten Einsatzdomänen für emotionale Interface-Agenten zu werden, da die Qualität dieser Systeme stark davon abhängt, inwieweit die Nutzerschnittstelle sich dem Erregungszustand des Nutzers anpassen kann. Insbesondere lernbehinderte Nutzer profitieren beim Lernen von der emotionalen Komponente eines solchen Systems ([MVDT04], [Moh05]).

Im Folgenden wird ein Überblick über diese und weitere wichtigen Anwendungsdomänen für emotionale Agenten aus dem MCI-Bereich gegeben.

Affective Computing Gegenstand des als *Affective Computing* ([Pic97]) bekannten Forschungsfelds innerhalb von MCI sind die sozial-emotionalen Aspekte einer verbesserten Mensch-Computer-Interaktion. Die Modellierung der emotionalen Zustände wird in diesem Zusammenhang als eine viel wichtigere Voraussetzung betrachtet als rationales Verhalten [Rei96].

Auf dem Weg zur Modellierung der emotionalen Zustände der Benutzer wird einerseits an der Entwicklung integrierter Sensoren zum Messen der physiologischen Werte, die vermutlich im Zusammenhang mit den Emotionen der Benutzer stehen, gearbeitet. Andererseits wird nach Lösungen gesucht, wie aus den gesammelten Daten auf die Benutzeremotionen geschlossen werden kann. Schließlich werden Modelle der emotionalen Zustände der Benutzer entwickelt, die in vielen Anwendungen eingesetzt werden könnten: “Perhaps the most fundamental application of affective computing will be to inform next-generation human interfaces that are able to recognize, and respond to, the emotional states of their users“ [Pic97].

Believable Agents und Lifelike Characters Aus dem Alltag ist die Tendenz bekannt, dass die Benutzer dazu neigen, die technischen Systeme, mit denen sie interagieren, zu personifizieren. Diesen Anthropomorphismus möchte man bewusst ausnutzen und verstärken. Emotionale Interface-Agenten, auch bekannt als Avatare oder anthropomorphe Interaktionsagenten, repräsentieren in Form von animierten Figuren (Gesichter, Tiere, Cartoon-Figuren usw.) das technische System gegenüber dem Benutzer. Dabei wird versucht, das natürliche Verhalten und dem Benutzer vertraute Interaktionsmuster so realistisch wie möglich nachzubilden. In diesem Zusammenhang sind die emotionalen Verhaltensweisen besonders wichtig, speziell auch Emotionsexpressionen in Form von Mimik oder Gesten [Dau02].

Bei diesem Forschungsgebiet geht es vor allem um die sozial-interaktiven und kommunikativen Aspekte von Emotion, wobei emotionale Interface-Agenten nicht notwendigerweise über elaborierte, Theorie-basierte interne emotionale Zustände verfügen müssen. Von Interesse ist vielmehr die glaubhafte, konsistente und für den Benutzer nachvollziehbare Manifestation von Emotionen ([BLR92a], [Rei96], [ARMK00], [AKGR00], [LA08], [CdSCS11], [PCLC07], [CZ02], [Dau98], [ENIY98], [LBS99], [ZS09], [KTB03]; s. [PPP07] für eine Übersicht). Um eine solche überzeugende emotionale Kommunikation zu ermöglichen, muss man die theoretischen Grundlagen der Emotionalität, insbesondere die sozialen Aspekte der Emotionserzeugung und des -austauschs, genau erforschen.

2. Forschungs- und Anwendungsfelder für emotionale Agenten

Benutzer-Modellierung Ein Spezialgebiet innerhalb des Forschungsfelds Affective Computing befasst sich mit der Erfassung und Modellierung emotionaler Benutzerzustände, um unter Berücksichtigung dieser Modelle möglichst angemessen - u.a. auch in emotionaler Hinsicht - auf den Benutzer reagieren zu können ([VP99], [Pic03], [PVH01], [Con02] u.v.a.m.). Wenn das System nicht nur Ausdrücke von Frustration, Konfusion, Interesse, Langeweile, Angst oder Freude bei dem Benutzer identifiziert, sondern diese Emotionen auch mit dem Systemverhalten und der Benutzertätigkeit assoziiert, könnte es auch besser lernen, welche Interaktionen erfolgreich waren und welche verbessert werden sollten, etwa welche Teile des Interface oder des Systems die größte Frustration für den User verursachten. Diese direkte Usability-Auswertung könnte als Feedback für die Systemdesigner oder auch für das System selbst genutzt werden, falls es über die Möglichkeit der dynamischen Anpassung seines Verhaltens an den emotionalen Zustand der Benutzer verfügt.

Die Erfassung der Benutzeremotionen wird aufgrund der Selbstdarstellungen der Benutzer, physiologischer Daten, von Verhaltensmerkmalen (Körperstellung, Bewegungsart, Gesichtsausdruck; vgl. [VP99], [PVH01]) und teilweise von verbaler Expression (Wortwahl, Stimmchwankungen [Pic03]) vorgenommen. Zur Aufnahme der physiologischen Daten werden Eingabegeräte entworfen, die mit Hilfe von integrierten Sensoren Hautwiderstand, Herzschlag, Muskelanspannung, Atmung, Blutdruck, Temperatur und andere physiologische Parameter messen und an das Computersystem weiterleiten. Kameras, Mikrofone, spezielle Eingabegeräte wie die am MIT entworfene "Sentic Mouse" [Kir97], "expression glasses" [JRJR01] sowie affektiver Schmuck und Accessoires ([PH97], [HP98]) und andere Instrumente aus dem Gebiet des *Wearable Computing* machen die Erfassung physiologischer Parameter möglich, die, so die Annahme, emotionale Zustände der Benutzer kennzeichnen und sie identifizieren lassen (vgl. [Pic97]). Es wird versucht, aus den Veränderungen dieser Parameter im Laufe der Interaktion mit dem Computersystem Rückschlüsse auf den zugrunde liegenden emotionalen Zustand des Benutzers zu ziehen.

Affektive Lernsysteme Die Benutzer-Modellierung findet verstärkt Anwendung in der Lernsoftware. Es wird versucht, Lernsysteme mit der Fähigkeit zur Emotionserkennung und -modellierung zu versehen, um das Systemverhalten besser an den Benutzer anpassen zu können. Emotionsmodelle werden in Tutor- und Dialogsystemen eingesetzt, um Benutzermodelle um emotionale Komponenten zu ergänzen ([LBS99], [RVC03], [Ell97], [MMSL01], [Con02], [CM04], [JR98], [KMP01], [PPB⁺04]).

Insbesondere Lernsysteme für Kinder könnten von adaptiven Fähigkeiten des Systems profitieren, wenn sie Zeichen der Frustration oder Langeweile beim Kind erkennen und z.B. die Lernstoffkomplexität oder die Präsentationsgeschwindigkeit entsprechend anpassen. Picard sieht die primären Ziele der affektiven Lernprogramme folglich darin, "to help keep the child's exploration going, by occasionally prompting with questions or feedback, and by watching and responding to the affective state of the child watching especially for signs of frustration and boredom that may precede quitting, for signs of curiosity or interest that tend to indicate active exploration, and for signs of enjoyment and mastery, which might indicate a successful learning experience" [Pic03]. In [Ell97] wurde mit Hilfe eines Multiagentensystems ein tutorielles Lernsystem implementiert, das verschiedene emotionale

2.3. Emotion zur Optimierung der Mensch-Computer-Interaction

“Lehrerpersönlichkeiten“ simuliert und so dem Benutzer erlaubt, einen für ihn optimalen Lernstil zu finden.

Spiele- und Unterhaltungsindustrie Ein weiteres Anwendungsfeld für “Believable Characters“ findet man in der Spiele- und Unterhaltungsindustrie. Mit Hilfe der emotionalen Modelle kann z. B. das Verhalten von Spielfiguren in Video-Spielen oder bei *Interactive Storytelling* gesteuert werden (vgl. z.B. [MJL00], [MAS02], [ZS09], [ZS11], [RVC03], [PCLC07], [SMBB08]). Ferner werden emotionale Spielzeuge produziert, z.B. “Entertainment Roboter“ von Sony (Aibo und Qrio), die sechs Emotionen durch Körperhaltung, Musik und Lichteffekte ausdrücken können.

Therapeutische Systeme Therapeutische Systeme werden bei Menschen eingesetzt, um bestimmte Defizite zu kompensieren bzw. eine Verbesserung bestimmter Fähigkeiten zu trainieren. Im Bereich des Autismus werden Experimente mit therapeutischen Computersystemen unter Einsatz von emotionalen Interface-Agenten durchgeführt ([TT02], [BP02], [PG08]). Dabei werden die Schwerpunkte auf die Kompensation fehlender Fähigkeiten von autistischen Kinder wie die Erkennung und Interpretation von emotionalem Ausdruck, Dialogführung und Nachahmungslernen gelegt.

Im Projekt Aurora [Dau02], dessen Ziel ebenfalls in der Entwicklung einer Therapie für autistische Kinder unter Einsatz von emotionalen Agenten bestand, konnten erhebliche Verbesserungen in allen oben genannten Problempunkten erzielt werden.

In [MVDT04] bzw. [Moh05] wurde ein Prototyp eines computerbasierten Trainingssystems für lernbehinderte Kinder auf Basis eines “emotionalen Computersystems“ erstellt, das sowohl Erregungszustände erkennen als auch über Interface-Agenten Emotionen simulieren kann.

Textual Processing Um die maschinelle Übersetzung von literarischen Texten zu unterstützen, versucht man, die mentalen Zustände der agierenden Akteure nachzubilden, um so die Suche nach dem äquivalenten Ausdruck in der anderen Sprache zu erleichtern [Ree91].

Offene Forschungsfragen Die Emotion wird in der MCI weniger aus der funktionalen Perspektive einer technischen Systemoptimierung betrachtet, sondern vielmehr aus der Motivation heraus, die Benutzbarkeit von Informatiksystemen zu steigern. Die vielen verschiedenen Anforderungen in der MCI-Forschung erfordern eine interdisziplinäre Betrachtung der Emotion. Die Erfassung physiologischer Aspekte für die Erkennung von Benutzeremotionen erfordert den Rückgriff auf Methoden und Theorien der Neurowissenschaften. Psychologische und soziologische Emotionstheorien sollen die Mechanismen erklären und Techniken entwerfen, mit deren Hilfe auf die Benutzeremotionen aufgrund der Situationsfaktoren und Einschätzungen geschlossen werden kann.

So muss unter Rückgriff auf emotionstheoretische Grundlagen entschieden werden, welche Form des Systemverhaltens bei der Interaktion mit dem Benutzer am wenigsten frustrierend ist, aufgrund von welchen expressiven Verhaltensweisen der Benutzer, die einen bestimmten emotionalen Zustand charakterisieren, das System diesen Zustand erkennen kann und

2. Forschungs- und Anwendungsfelder für emotionale Agenten

aufgrund von welchen soziologischen Gesetzen der Emotionsregulation durch soziale Interaktion das System sein Verhalten an den Benutzer anpassen soll. Man sieht, dass auch auf dem Gebiet der MCI die interdisziplinäre Betrachtung der Emotionstheorien verschiedener Wissenschaften in der Art des Forschungsgegenstandes selbst begründet ist. Ferner ist für MCI die Frage nach der Integrierbarkeit verschiedener Emotionstheorien - vor allem psychologischer und soziologischer Provenienz - in hohem Maße relevant.

Es stellen sich auch im MCI-Bereich die prinzipiellen Forschungsfragen nach den psychologischen, soziologischen und neurowissenschaftlichen Theorien, die die funktionale Rolle der Emotion in einem natürlichen System aufdecken. Des Weiteren muss man nach der prinzipiellen Operationalisierbarkeit und Integrierbarkeit dieser Theorien und ihrer Umsetzbarkeit in einem technischen System fragen. Diese Fragestellungen werden in Kapitel 6 schwerpunktmäßig behandelt.

3. Agenten- und Multiagentensysteme

Agentenorientierung bietet vor allem in dynamischen Umgebungen und unter nicht vorhersehbaren Konditionen die Möglichkeit, robuste Lösungen auszuarbeiten und die Systemkomplexität durch Aufteilung des Problembereichs in unabhängig voneinander agierende Software-Einheiten (sog. *Software-Agenten* [FG97]) zu bewältigen. Agenten haben die Fähigkeit, eigenständige und zielgerichtete Verhaltensweisen auszuprägen, und verfügen über die Eigenschaften Intelligenz, Autonomie, Reaktivität oder Proaktivität und z.T. über soziale Kompetenz, die in komplexen, dynamischen, nicht deterministischen Anwendungskontexten in besonderem Maße zum Tragen kommen.

Dieses Kapitel gibt einen Überblick über grundlegende Konzepte und Fragestellungen der Agentenforschung. Zuerst wird der Agentenbegriff eingegrenzt und durch ein abstraktes Agentenmodell formalisiert. Anschließend werden die grundlegenden Agentenarchitekturen vorgestellt und diskutiert. Die Übersicht und Klassifizierung der Agentenarchitekturen in Kapitel 3.2 bildet den theoretischen und terminologischen Rahmen für die Betrachtung der Architekturen emotionaler Agenten in Kapitel 7. In Kapitel 3.3 wird der Einsatz der Emotion bei der Implementierung von Agenteneigenschaften wie Autonomie, Adaption, Reaktivität bzw. Proaktivität und soziale Kompetenz diskutiert. Schließlich wird auf spezielle Fragen zu Multiagentensystemen eingegangen und insbesondere existierende Kommunikationssprachen, Modellierungstechniken und Referenzarchitekturen für (Multi-)Agentensystemen näher betrachtet.

3.1. Begriffsbestimmung

Zuerst wird in diesem Abschnitt der Agentenbegriff durch eine Zusammenstellung der gängigen Agentendefinitionen und die Beschreibung der wichtigen Agenteneigenschaften spezifiziert. Anschließend werden die Eigenschaften der Anwendungsumgebungen betrachtet, die den Einsatz von Agenten sinnvoll machen.

3.1.1. Agentendefinition

Die oft zitierte Definition von Wooldridge and Jennings (1998) beschreibt Agenten folgendermaßen:

“[Agent is] an encapsulated computer system, situated in some environment, and capable of flexible autonomous action in that environment in order to meet its design objectives“ ([JW98], S. 5).

Als grundlegende Agenteneigenschaften werden also Einbettung in eine Umgebung und Autonomie genannt. Dabei bedeutet die Einbettung in diesem Kontext die Fähigkeit des

3. Agenten- und Multiagentensysteme

Agenten, sensorischen Input aus der Umgebung wahrzunehmen und Aktionen auszuführen, die Veränderungen in dieser Umgebung verursachen können.

In diesem Zusammenhang wird ein wichtiger Unterschied zwischen Agenten- und Expertensystemen deutlich: Expertensysteme interagieren oft nicht unmittelbar mit einer Umgebung, sondern nur indirekt, indem sie Informationen von einem Benutzer erhalten (vgl. “disembodied intelligence“ [JSW99]). In diesem Fall vermittelt der Benutzer zwischen System und Umwelt und ersetzt den sensorischen Input, welcher bei einem Agentensystem zu den grundlegenden Systemeigenschaften gehört. Entsprechend wird die Systemantwort auf die Eingabe nicht in Form von unmittelbaren Aktivitäten in der Umgebung, sondern als Feedback bzw. Auskunft an den Benutzer gegeben.

In der Definition von Franklin und Graesser (1996) wird ein autonomer Agent nicht nur als eine in die Umgebung eingebettete Einheit, sondern sogar als Bestandteil der Umgebung beschrieben:

“An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future“ [FG97].

Kinny, Georgeff und Rao (1996) fügen zu den bereits genannten Merkmalen die Reaktivität als grundlegende Agenteneigenschaft hinzu und betonen in ihrer Definition die Eigenschaften der Agentenumgebung:

“The agent paradigm in AI is based upon the notion of reactive, autonomous, internally-motivated entities embedded in changing, uncertain worlds they perceive and in which they act“ [KGR96].

In [vST02] wird ein Agent als ein autonomer Prozess definiert, der mit anderen Agenten oder Benutzern interagieren kann:

“[W]e define a software agent as an autonomous process capable of reacting to, and initiating changes in, its environment possibly in collaboration with users and other agents. The feature that makes an agent more than just a process is its capability to act on its own, and, in particular, to take initiative where appropriate“ ([vST02], S. 173).

Eine sehr allgemeine Definition geben Russel und Norvig (1995):

“An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators“ [RN95].

Diese Definition bringt alle aufgeführten Agentendefinitionen auf den kleinsten gemeinsamen Nenner: Die essenzielle Eigenschaft eines Agenten wird demnach in seiner Fähigkeit gesehen, Informationen aus der Umgebung wahrzunehmen und Aktionen auszuführen, welche die Umgebung verändern können.

3.1.2. Agenteneigenschaften

Die meisten Agentendefinitionen beschreiben den Agentenbegriff extensional, indem sie wesentliche Fähigkeiten bzw. Merkmale eines Agenten aufzählen. Zusätzlich zu den oben genannten essenziellen Eigenschaften werden Reaktivität, Proaktivität, soziale Kompetenz, Mobilität, Rationalität, Kommunikation usw. als wichtige Agenteneigenschaften genannt. Verschiedene Autoren nehmen unterschiedliche Sätze von Eigenschaften als notwendige Charakteristika eines Agenten an. Die Auswahl der charakteristischen Agentenmerkmale hängt dabei erheblich von der jeweiligen Anwendungsdomäne ab.

Im Folgenden werden die wichtigsten Agenteneigenschaften aufgezählt und erläutert. Die Zusammenstellung erhebt keinen Anspruch auf Vollständigkeit, vielmehr bietet sie eine Übersicht über in der Literatur besonders oft erwähnte Eigenschaften.

In [Woo02] werden Autonomie, Reaktivität, Proaktivität und soziale Kompetenz als charakteristische Merkmale eines intelligenten Agenten hervorgehoben ([Woo02], S. 23; vgl. auch “weak notion of agency“ in [WJ95]):

- **Autonomie** bezeichnet die Fähigkeit eines Agenten, selbstständig und unabhängig Entscheidungen über sein weiteres Handeln treffen zu können. Ein autonomer Agent hat die volle Kontrolle über seinen internen Zustand und seine Handlungen. Er kann selbst, basierend auf seinem internen Zustand und anhand seiner Planungs- und Entscheidungsverfahren, die Art, den Zeitpunkt und die Dauer seiner Aktionen bestimmen. Der Auslöser einer Aktion kann sowohl eine Änderung in der Umgebung (*reaktive Aktionsauslösung*) als auch eine Änderung des internen Zustands des Agenten (*proaktive Aktionsauslösung*) sein. Autonomie ist somit eine notwendige Voraussetzung für die weiteren Agenteneigenschaften (Reaktivität, Proaktivität, soziale Kompetenz) und wird zu den grundlegenden Eigenschaften eines Agenten gezählt (s.o.).
- **Reaktivität** bezeichnet die Fähigkeit eines Agenten, seine Umgebung fortlaufend wahrzunehmen und auf Veränderungen in angemessener Zeit zu reagieren [Woo97]. Diese Reaktionsfähigkeit ist nicht unbedingt mit einem sofortigen Handeln verbunden, sondern eher mit einer Veränderung des inneren Zustands des Agenten, welcher wiederum eine Reaktion hervorrufen kann.
- **Proaktivität**: Intelligente Agenten zeigen zielgerichtetes Verhalten, indem sie nicht nur auf Umgebungsveränderungen reagieren, sondern auch selbst Initiative ergreifen, um ihre Ziele zu erreichen. Ein Agent verhält sich proaktiv, wenn seiner Aktivität kein unmittelbar auslösendes Ereignis in der Umwelt zugrunde liegt.
- **Soziale Kompetenz**: Agenten haben die Fähigkeit, mit anderen Agenten (oder Menschen) zu interagieren, zu kooperieren und zu verhandeln, um ihre internen Ziele zu erreichen.

3. Agenten- und Multiagentensysteme

Außerdem werden in der Literatur häufig noch folgende Agenteneigenschaften hervorgehoben (vgl. [Röl04]):

- **Mobilität** ist die Fähigkeit eines Agenten, seinen Standort im System zu wechseln. Agenten können sich in der realen Welt (z. B. Roboter) oder in einem virtuellen System (z. B. Software-Agenten auf verschiedenen Agentenplattformen) bewegen.
- **Rationalität** ist die Fähigkeit eines Agenten, sich in Bezug auf seine Ziele immer so zu verhalten, dass diese erfüllbar bleiben. Ob ein Verhalten als rational bewertet werden kann, ist abhängig von Performanzkriterien, dem Vorwissen des Agenten (in Bezug auf die Umgebung), Handlungsmöglichkeiten bzw. -fähigkeiten des Agenten sowie seinen Perzeptionsfähigkeiten.
- **Adaptivität** ist die Fähigkeit eines Agenten, sich an Veränderungen seiner Umgebung anzupassen.

Im Zusammenhang mit Multiagentensystemen sind ferner folgende Agenteneigenschaften von Bedeutung (vgl. auch Kapitel 3.4):

- **Kommunikation** bezeichnet einen Informationsaustausch zwischen den Agenten eines Multiagentensystems und ermöglicht den Agenten miteinander zu kooperieren und gemeinsam Ziele zu verfolgen.
- **Koordination** zwischen den Agenten eines Multiagentensystems ist nur unter der Voraussetzung vorhandener Kommunikation möglich. Unter diesem Begriff werden die *Kooperation* und *Konkurrenz* von Agenten bei der Verfolgung ihrer individuellen Ziele zusammengefasst.

Agenten sind in erster Linie für das zielgerichtete Handeln in komplexen Umgebungen konzipiert. Sie müssen in der Lage sein, in unbekannten Umgebungen und unter nicht vorhersehbaren Konditionen eigenständig Entscheidungen zu treffen, Verhaltensweisen auszuprägen und ihre Handlungen mit denen anderer Agenten zu koordinieren. Die in diesem Abschnitt aufgezählten Anforderungen an die Agentensysteme leiten sich im Wesentlichen aus den Eigenschaften der Umgebung ab, in der diese agieren müssen. Die Kriterien für die Klassifizierung von Agentenumgebungen werden im folgenden Abschnitt vorgestellt.

3.1.3. Umgebungsdimensionen

Nach [RN95] sind folgende Dimensionen für die Klassifizierung einer Agentenumgebung von Relevanz ([RN95], S. 46):

- **Beobachtbarkeit:** Eine Umgebung ist *beobachtbar* (*observable, accessible*), falls alle ihre Eigenschaften, die für die Lösung der aktuellen Aufgabe relevant sind, dem Agenten über seine Sensorik zugänglich sind. Der sensorische Input des Agenten muss daher zuverlässig, korrekt, präzise und vollständig sein. Die meisten real-weltlichen Umgebungen sind in diesem Sinne nicht bzw. nicht vollständig beobachtbar.

- **Vorhersagbarkeit:** Die Umgebung ist *vorhersagbar (deterministisch)*, falls der Folgezustand der Umgebung durch den gegenwärtigen Zustand bzw. durch die Aktionen des Agenten im gegenwärtigen Zustand vorbestimmt ist. In einer deterministischen Umgebung sind insbesondere die Auswirkungen der Agentenaktionen auf die Umgebung eindeutig definiert. In der Praxis sind fast alle Agentenumgebungen nicht deterministisch.
- **Dynamik:** Eine *statische* Umgebung wird nur durch Agentenaktionen verändert und kann ansonsten als gleichbleibend betrachtet werden. Im Gegensatz dazu kann eine *dynamische* Umgebung außer durch den Agenten selbst noch durch weitere Prozesse verändert werden. In einer *dynamischen bzw. semidynamischen* Umgebung kann der Agent nicht annehmen, dass zur Zeit t_1 die Umgebung so bleibt, wie sie zur Zeit t_0 war (mit t_0 vor t_1). Da sich der Umgebungszustand zum Zeitpunkt t_1 verändert haben kann, während der Agent seine weiteren Aktionen von dem zur Zeit t_0 beobachteten Umgebungszustand ausgehend plant, besteht für Agenten in dynamischen Umgebungen Handlungsdruck.
- **Interne Struktur der Umgebung, der Sensorik und der Aktorik:** Eine Umgebung ist *diskret*, falls es in ihr eine fixe, endliche Anzahl von Aktions- und Perzeptionsmöglichkeiten gibt. Dieser Aspekt bezieht sich auf den Zustandsraum. Ebenso können sensorische Wahrnehmung und Aktivitäten eines Agenten *diskret oder kontinuierlich* sein.

Eine besonders komplexe Klasse von Umgebungen sind die *nicht vollständig beobachtbaren, nicht deterministischen, dynamischen und kontinuierlichen Umgebungen*, die auch *offene Umgebungen* genannt werden [Hew86].

3.2. Agentenarchitekturen

Den Ausgangspunkt der weiteren Betrachtungen bildet eine formale Darstellung des Agentenbegriffs nach Wooldridge (2002). Von seiner Definition eines formalen Agentenmodells ausgehend werden die aus der Literatur bekannten Agentenarchitekturen vorgestellt. Der Begriff “Agentenarchitektur“ wurde in Kapitel 1.5 als “das Zusammenspiel der internen Komponenten eines Agenten, die ihn zu seinen Entscheidungen führt“, definiert und wird im Weiteren in diesem Sinne verwendet. In [Röl04] wurde der Einsatz der Petrinetze zur Darstellung der Agenteneigenschaften und Agentenarchitekturen illustriert. An diese Vorgehensweise anknüpfend, werden wir ausgewählte formale Definitionen ebenfalls mit Netzmodellen illustrieren.

3.2.1. Abstraktes Agentenmodell

Die oben aufgeführten Eigenschaften eines Agenten können in einem abstrakten Agentenmodell formalisiert werden ([Woo02], S. 31). Dazu wird eine Menge von Umgebungszuständen E , ein Aktionsrepertoire eines Agenten Ac sowie eine Abbildung Ag von E auf Ac definiert.

3. Agenten- und Multiagentensysteme

Sei also

$$E = (e_1, \dots, e_n)$$

eine abzählbare Menge von möglichen diskreten Umgebungszuständen. Zu jedem Zeitpunkt t befindet sich die Umgebung in genau einem dieser Zustände. Das Handlungsrepertoire eines Agenten ist durch eine abzählbare Menge Ac von Aktionen definiert, mit

$$Ac = (\alpha_1, \dots, \alpha_n)$$

Ein Zustandsverlauf eines Agenten in einer Umgebung ist eine Menge R aller möglichen endlichen Sequenzen r über E und Ac der Form

$$r : e_0 \xrightarrow{\alpha_0} e_1, \dots, e_{n-1} \xrightarrow{\alpha_{n-1}} e_n.$$

Jede Aktion eines Agenten verändert die Umgebung. Eine Übergangsfunktion (state transformer function) τ beschreibt die Auswirkung der Agentenaktionen auf die Umgebung. Sie bildet den Zustandsverlauf eines Agenten (der mit einer Aktion endet) auf eine Menge der möglichen Umgebungszustände ab, die aus den Agentenaktionen resultieren können.

$$\tau : R^{Ac} \rightarrow \wp(E).$$

Die Entscheidung des Agenten über seine weiteren Aktionen beruht auf den Zustandsverläufen, die er bis dato ausgeführt hat (history). Deshalb kann ein Agent (bzw. die *Entscheidungsfunktion* eines Agenten) durch die folgende Funktion beschrieben werden:

$$Ag : R^E \rightarrow Ac \tag{3.1}$$

Der Zustandsverlauf eines Agenten ist somit eine Folge von Umgebungszuständen und Agentenaktionen. Formal kann der Zustandsverlauf eines Agenten Ag in einer Umgebung $Env = \langle E, e_0, \tau \rangle$ mit einer Sequenz

$$(e_0, \alpha_0, e_1, \alpha_1, e_2, \alpha_2, \dots)$$

repräsentiert werden mit:

1. e_0 ist der initiale Zustand der Umgebung Env ;
2. $\alpha_0 = Ag(e_0)$ und
3. für $u > 0$,
 $e_u \in \tau((e_0, \alpha_0, \dots, \alpha_{u-1}))$,
 wo
 $\alpha_u = Ag((e_0, \alpha_0, \dots, \alpha_u))$.

3.2. Agentenarchitekturen

Dieses abstrakte Agentenmodell kann anhand der bisherigen Begriffsbildung dahingehend spezifiziert werden, dass nunmehr der Entscheidungsprozess des Agenten in *Perzeption* und *Aktion* aufgeteilt wird. Ein Agent beobachtet seine Umgebung und nimmt bestimmte Merkmale des aktuellen Umgebungszustandes wahr. Diese Umgebungsmerkmale werden intern abgebildet als sogenannte *Perzepte*. Durch die Perzeption wird festgelegt, welche Umgebungsmerkmale wahrgenommen werden, d. h. die Gesamtmenge der Umgebungszustände E wird in Bezug auf die relevanten Umgebungsmerkmale eingeschränkt. So können zwei verschiedene Umgebungszustände e_1 und e_2 in E zu einem gleichen Perzept p in Per führen (mit Per - als eine (nicht leere) Menge von Perzepten), d.h. die perzeptuellen Informationen, die ein Agent von diesen Umgebungszuständen bekommt, sind identisch; die assoziierten Umgebungszustände sind dann aus der Sicht des Agenten nicht unterscheidbar. Dadurch ist der Zustandsraum Per in der Regel viel kleiner als E .

Per sei eine (nicht leere) Menge von Perzepten. Mit der Funktion *see*

$$see : E \rightarrow Per \quad (3.2)$$

wird die Abbildung der Umgebungszustände auf Perzepte bezeichnet. Das Verhalten des Agenten (*Agentenfunktion*) wird dann durch die Abbildung *action* beschrieben, die Perzeptionsfolgen auf Aktionen abbildet:

$$action : Per^* \rightarrow Ac \quad (3.3)$$

Dann kann ein Agent durch ein Paar $Ag = \langle see, action \rangle$ beschrieben werden.

3.2.2. Rein reaktiver Agent

Eine spezielle Klasse der Agenten, die ihre Aktionswahl bzw. ihre Entscheidungen allein aufgrund der aktuellen Wahrnehmungen der Umgebung trifft, heißt *rein reaktive* ([Woo02] oder *tropistische* Agenten ([Gen87]).

Das Verhalten eines rein reaktiven Agenten kann repräsentiert werden durch die Funktion:

$$Ag : E \rightarrow Ac$$

bzw. in der verfeinerten Darstellung:

$$action : Per \rightarrow Ac$$

Die Leistung eines rein reaktiven Agenten ist abhängig von der vollständigen Beobachtbarkeit der Umgebung. Dies bedeutet, dass rein reaktive Agenten in solchen Umgebungen effektiv funktionieren, die sie über ihre Sensorik vollständig erfassen können.

In [Röl04] wurde ein Petrinetz-Modell eines reaktiven Agenten vorgestellt, das wir hier zur Veranschaulichung der reaktiven Architektur wiedergeben (Abb. 3.1):

Die übereinander angeordneten Transitionen **Verhalten 1**, **Verhalten 2** usw. stellen die verschiedenen Verhaltensalternativen dar, aus denen der Agent bei der Perzeption eines spezifischen Stimulus (Nachricht n) eine assoziierte Verhaltensweise auswählt. In der Abbildung ist diese bedingte Auswahl durch den Überwachungsausdruck *guard* an der oberen Verhaltenstransition angedeutet.

3. Agenten- und Multiagentensysteme

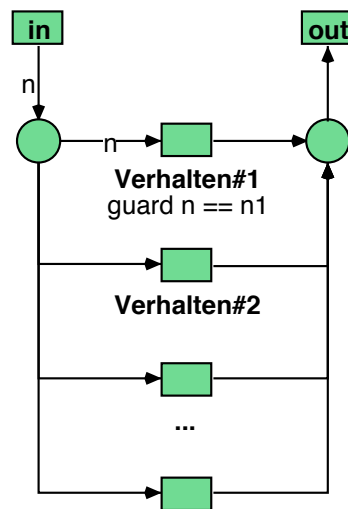


Abbildung 3.1.: Reaktive Agentenarchitektur (aus [Röl04], S. 30).

Subsumptionsarchitektur von Brooks Das wohl bekannteste Beispiel für die Umsetzung eines reaktiven Agentenmodells in einer konkreten Agentenarchitektur ist die *Subsumptionsarchitektur* von Brooks (1991). Eine charakteristische Eigenschaft dieser Architektur ist der Verzicht auf Modellierung der Entscheidungsmechanismen auf Basis der symbolischen (Umwelt-)Repräsentation (siehe Kapitel 3.2.5) zugunsten einer adaptiven und flexiblen Steuerung im Zuge der Agenteninteraktion mit der Umwelt.

Der Entscheidungsprozess eines Agenten wird durch eine Reihe von Verhaltensmodulen (*task accomplishing behaviours*) ([Woo02], S. 90) realisiert. Jedes Verhaltensmodul bildet den sensorischen Input aus den aktuell perzipierten Eigenschaften der Umweltsituation auf eine (vordefinierte) Aktion ab. Dabei wird weder komplexe symbolische Repräsentation (z. B. der Umwelt), noch Entscheidungsprozess auf Basis der Manipulation einer solchen symbolischen Repräsentation verwendet. In der Architekturimplementierung von Brooks wurden die Verhaltensmodule durch endliche Automaten realisiert. In vielen Implementationen werden die Verhaltensmodule durch einfache (Verhaltens-)Regeln der Form

$$situation \rightarrow action$$

implementiert

Die Verhaltensmodule arbeiten asynchron, so dass mehrere Verhaltensmodule aufgrund der aktuell perzipierten Eigenschaften der Umweltsituation gleichzeitig aktiviert werden können. In diesem Falle muss ein Priorisierungsmechanismus definiert werden, aufgrund dessen die als Nächstes auszuführende Aktion ausgewählt wird. Brooks schlug die Verwendung einer *Subsumptionshierarchie* vor, bei der die Verhaltensmodule auf mehrere Ebenen aufgeteilt werden. Die Module der niedrigeren Hierarchieebenen haben Priorität vor den Verhaltensweisen der höheren Ebenen.

Die *see* -Funktion kann, wie beim allgemeinen Modell, durch

$$see : E \rightarrow Per$$

definiert werden. Die Entscheidungsfunktion *action* wird realisiert durch eine Menge von Verhaltensregeln und eine Prioritäts-Relation auf dieser Menge. Eine Verhaltensregel wird durch ein Paar (c, a) mit $c \subseteq P$ und $a \in A$ notiert. Eine Verhaltensregel (c, a) wird in einem Umgebungszustand $s \in S$ aktiviert gdw. $see(s) \in c$.

Eine Prioritäts-Relation (*inhibition function*) ist eine strenge Totalordnung auf der Menge der Verhaltensregeln R mit $\prec \subseteq R \times R$. Bei der Aktionswahl werden im ersten Schritt alle Verhaltensregeln berechnet, die durch aktuelle Perzeptionsergebnisse aktiviert werden. Im zweiten Schritt wird für jede aktivierte Verhaltensregel (c, a) geprüft, ob sie durch eine Verhaltensregel mit höherer Priorität dominiert wird. Falls keine Verhaltensweise höherer Priorität gefunden wird, ist a die nächste auszuführende Aktion.

Bewertung Die Hauptvorteile der reaktiven Agentenarchitekturen allgemein und insbesondere der Subsumptionsarchitektur sind ihre Einfachheit, Leistungsfähigkeit und Robustheit. Jedoch gibt es eine Reihe von fundamentalen Problemen, die für rein reaktive Architekturen nicht gelöst sind.

- Weil ein reaktiver Agent kein Umgebungsmodell besitzt, müssen die in der Umgebung auftretenden Ereignisse, die von dem Agenten mit entsprechenden Aktionen beantwortet werden sollen, durch den Programmierer explizit vorgegeben werden.
- Da ein rein reaktiver Agent ausschließlich auf Basis der aktuellen Wahrnehmungen handelt, kann er bei seiner Aktionswahl keine Umgebungseigenschaften berücksichtigen, die gegenwärtig nicht wahrgenommen werden. Entsprechend kann er keinen Bezug auf die Umstände nehmen, die zu dem aktuellen Zustand der Umgebung geführt haben.
- Dadurch trifft der reaktive Agent lediglich (zustands-)lokale Entscheidungen; dies erschwert das Lernen (da z.B. eine Performanzsteigerung aufgrund der durchlaufenen Situationen - *history* - nicht möglich ist) erheblich und macht eine dauerhafte Adaption des Agenten in der Umgebung kaum möglich.

Damit das Agentenverhalten in Bezug auf diese Aspekte optimiert werden kann, bedarf ein Agent einer Art Gedächtnis, eines inneren Zustandes, in dem er das Modell der Umgebung und seine Veränderung aufbewahrt.

3.2.3. Agent mit innerem Zustand

Bei einem Agenten mit einem innerem Zustand werden die wahrgenommenen Informationen über die Umwelt und ihre Veränderungen auf einen inneren Zustand abgebildet, so dass der Agent ein internes Modell seiner Umgebung besitzt, das er ständig durch neue Perzepte aktualisieren kann. Aufgrund seines aktuellen inneren Zustandes entscheidet der Agent über seine nächsten Aktionen.

3. Agenten- und Multiagentensysteme

I sei eine Menge aller möglichen internen Zustände eines Agenten. Die Funktion *action* wird für einen Agenten mit innerem Zustand definiert als

$$action : I \rightarrow Ac.$$

Aufgrund der aktuellen Perzepte kann ein Agent seinen inneren Zustand aktualisieren bzw. verändern. Die Funktion *next* formalisiert diesen Sachverhalt folgendermaßen:

$$next : I \times Per \rightarrow I.$$

Das Verhalten eines Agenten mit internem Zustand kann wie folgt beschrieben werden: Der Agent startet in einem initialen internen Zustand i_0 mit einem Zustand e . Durch die Umgebungsbeobachtung generiert er ein Perzept $see(e)$. Der interne Zustand wird durch die *next*-Funktion aktualisiert zu $next(i_0, see(e))$. Aufgrund des aktualisierten internen Zustandes wird eine Aktion $action(next(i_0, see(e)))$ aus der Menge der Agentenaktionen ausgewählt und ausgeführt. Diese auszuführende Aktion wird gleichzeitig durch eine Veränderung des internen Zustandes dokumentiert.

Bewertung Der interne Zustand kann so modelliert werden, dass alle Perzepte und Zustandsänderungen gespeichert werden. Dadurch entsteht eine Abfolge aller vom Agenten durchgelaufenen Zustandsänderungen - *Historie (history)*. Da der Entscheidungsprozess eines Agenten mit internem Zustand nicht nur auf den aktuellen Perzepten, sondern auch auf dem internen Zustand und daher auf der Historie der Wahrnehmung beruht, werden beim Entscheiden auch Teile der Umgebung berücksichtigt, die gegenwärtig nicht wahrgenommen werden. Dies ist insbesondere in nicht statischen, vorhersagbaren Umgebungen von Vorteil, da der Agent auch die Auswirkungen der eigenen (vergangenen und aktuellen) Handlungen auf den Umgebungszustand berechnen und in den Entscheidungsprozess miteinbeziehen kann.

Der Agent mit innerem Zustand handelt somit auf der Basis eines internen Modells der Umgebung, wie sie "gegenwärtig" ist. In den dynamischen, nicht deterministischen Umgebungen wäre es jedoch wünschenswert, auch gewisse Aspekte der Umwelt berücksichtigen zu können, die in der Zukunft eintreten werden, insbesondere Veränderungen, die sich aufgrund der Handlungen des Agenten ergeben.

3.2.4. Agent mit Utility-Funktion

Eine weitere Optimierung der Entscheidungsprozesse eines Agenten in Bezug auf die Aktionswahl liefert das Konzept der *Utility*-Funktion. Eine Utility-Funktion u weist dem aktuellen Umgebungszustand E einen reellen Wert

$$u : E \rightarrow \mathbf{IR}$$

zu.

Bewertung Ein Agent mit Utility-Funktion wird stets bestrebt sein, Umgebungszustände mit einem maximalen Utility-Wert herbeizuführen. Aufgrund der Utility kann die Performanz eines Agenten bewertet werden. So ist ein optimaler Agent dadurch definiert, dass er seine Utility maximiert ([Woo02], S. 39).

Um nicht nur einzelne Umweltzustände, sondern aus einer langfristige Perspektive auch die Agentenperformanz zu bewerten, kann den Agentenläufen ein Utility-Wert zugewiesen werden

$$u : R \rightarrow \mathbf{IR}.$$

Dieser Ansatz ist insbesondere für autonome Agenten vorteilhaft, die über längere Zeiträume selbstständig zielgerichtet handeln müssen.

3.2.5. Deliberativer Agent

Bei einem deliberativen Agenten wird sein interner Zustand durch eine Wissensbasis dargestellt, die ein symbolisches Modell der Agentenumwelt enthält. In der Wissensbasis sind explizite Informationen abgelegt, die ein Agent über seine Umgebung besitzt. Neben dem internen Umweltmodell ist die Fähigkeit zum (logischen) Schlussfolgern bzw. zur Wissensverarbeitung ein wesentliches Charakteristikum für einen deliberativen Agenten: Aufgrund von vorgegebenen Beweis- / Ableitungsverfahren kann er, von dem in der Wissensbasis explizit kodierten Wissen ausgehend, implizit kodiertes Wissen ableiten / folgern. Ebenso können mögliche Aktionen und ihre Auswirkungen in der Wissensbasis symbolisch repräsentiert werden. Durch Manipulation dieser symbolischen Repräsentationen kann ein Agent nach einer Problemlösung suchen oder planen (Planung als Suchproblem). Im Rahmen dieser Schlussfolgerungsprozesse verwendet ein Agent das in seinem Umweltmodell vorhandene Wissen, um seinen internen Zustand zu modifizieren.

Dieser Ansatz gründet auf der klassischen KI-Forschungsstrategie “physical symbol hypothesis” [NS72] bzw. “symbolische KI“. Diese geht davon aus, dass in einem System intelligentes Verhalten erzielt werden kann, indem das System eine symbolische Repräsentation seiner Umgebung und seiner gewünschten Verhalten erhält und auf diesen Repräsentationen syntaktische Manipulationen durchführt.

Deduktiver Agent Die symbolische Repräsentation kann z.B. mit logischen Formeln erfolgen; in diesem Fall entspricht die syntaktische Manipulation dem Deduktionsprozess. In [Gen87] wird der Term “deliberativer Agent“ auf solche Agentenarchitekturen beschränkt, deren interner Zustand durch eine Wissensbasis aus Formeln der Prädikatenlogik 1. Ordnung dargestellt wird. Um die Abgrenzung zur BDI-Architektur (s.u.) zu betonen, werden wir hier zur Bezeichnung solcher Logik-basierten Agentenarchitekturen den Begriff “deduktiver Agent“ verwenden.

Bewertung Die Vorteile der Verwendung von Logiken, z.B. der Prädikatenlogik, liegen vor allem in der Eleganz und der sauberen (logischen) Semantik eines solchen Ansatzes. Jedoch ist ein wesentlicher Nachteil die hohe Berechnungskomplexität der deduktiven

3. Agenten- und Multiagentensysteme

Theorembeweise. Wegen des hohen Rechenaufwands, der für die Bestimmung der optimalen Aktion in einem zum Zeitpunkt t_0 gegebenen Umweltzustand (repräsentiert durch den aktuellen Wissensbasiszustand) erforderlich ist, kann nicht davon ausgegangen werden, dass die Umgebung zum Zeitpunkt t_1 so bleibt, wie sie zum Zeitpunkt t_0 war, als der Agent mit dem Deduktionsprozess zum Ableiten der optimalen Aktion begonnen hat. Falls sich der Umgebungszustand zum Zeitpunkt t_1 verändert, ist die aufgrund vom Umgebungszustand zum Zeitpunkt t_0 abgeleitete Aktion möglicherweise nicht mehr optimal. Man spricht in diesem Zusammenhang von *Berechnungsrationalität*: Ein Agent besitzt diese Eigenschaft genau dann, wenn sein Entscheidungs- bzw. Aktionswahlmechanismus eine Aktion liefert, die zu Beginn der Entscheidungsprozesses optimal war. Solche Agenten sind nur unter der Voraussetzung effizient, dass zum einen in der Umwelt keine signifikanten Veränderungen stattfinden, solange der Entscheidungsprozess des Agenten abläuft, und zum anderen die Aktion, die zu Beginn des Entscheidungsprozesses rational war, auch dann rational bleibt, wenn der Entscheidungsprozess abgeschlossen ist.

Weitere Schwierigkeiten des Logik-basierten Ansatzes hängen mit der *symbolischen* Abbildung der perzipierten Informationen zusammen. So sind die Probleme der symbolischen Repräsentation von (bewegten) Bildern oder Geräuschen bis heute nicht zufriedenstellend gelöst. Weitere offene Fragen betreffen die Repräsentation und Verarbeitung von Wissen in dynamischen Umgebungen, z.B. Repräsentation von temporären, unsicheren oder unscharfen Informationen.

Wegen dieser Einschränkungen ist der praktische Einsatz von rein deduktiven Agenten mit Logik-basierten Entscheidungs- bzw. Aktionswahlmechanismen in vielen Anwendungsfeldern aus Effizienz- bzw. Performanzgründen nicht möglich - vor allem nicht in dynamischen Umgebungen, in welchen für den Agenten wegen der schnellen Umweltveränderungen Handlungsdruck besteht.

BDI-Architektur. Eine der bekanntesten Architekturen für deliberative Agenten ist die sogenannte BDI-Architektur. Dieser Ansatz stellt eine Abwandlung des rein Logik-basierten Ansatzes hin zur Einschränkung des logischen Deduktionsmechanismus zugunsten der effizienteren anwendungsorientierten Lösungen dar. Die BDI-Logik gründet auf dem Konzept der "Alltags-Deduktion" (Practical Reasoning) nach [Bra90]. Danach ist *Practical Reasoning* in erster Linie auf *Aktionen* gerichtet und besteht in der Abwägung von gegensätzlichen Überlegungen für und gegen konkurrierende (Handlungs-)Optionen, wobei Wünsche, Bewertungen, Prioritäten und Überzeugungen eines Agenten die Basis für diese Überlegungen bilden ([Bra90], S. 17).

Wooldridge (2002) betrachtet *Practical Reasoning* als einen in zwei Komponenten aufgeteilten Prozess: Er unterscheidet den *Deliberationsprozess*, bei dem ein Agent darüber entscheidet, *welchen* Zielzustand er durch sein Verhalten erreichen möchte, vom *Planungsprozess*, bei dem entschieden werden muss, *wie* der gewünschte Zielzustand zu erreichen ist.

Rao und Georgeff (1995) entwickelten davon ausgehend ein Modell, in dem rationales Verhalten durch die Konzepte *Beliefs*, *Desires* und *Intentions* beschrieben wird [Rao95]:

- **Beliefs** stellen das durch Perzeption gewonnene Wissen eines Agenten über seine Umwelt dar.
- **Desires** leiten sich direkt aus den *Beliefs* ab. Sie repräsentieren “Wünsche“ eines Agenten bezüglich der zukünftigen Umweltzustände. Durch die Formulierung von *Desires* wird noch keine Aussage über ihre Realisierbarkeit durch den Agenten getroffen. Auch miteinander in Konflikt stehende oder widersprüchliche Wünsche sind möglich.
- **Intentions** sind eine Untermenge von Wünschen, die von dem Agenten aufgrund seiner aktuellen Überzeugungen (*Beliefs*) zur direkten Umsetzung ausgewählt wurden. Sie repräsentieren konkrete (Handlungs-)Absichten eines Agenten und sind unmittelbar mit den Aktionen verknüpft. Bei den Intentionen sind logische Widersprüche nicht mehr zulässig.

An dieser Stelle wird noch keine Aussage darüber getroffen, wie *Beliefs*, *Desires* und *Intentions* eines Agenten repräsentiert werden sollen. Bei der konkreten Implementierung der Architektur kann *symbolische* Repräsentation (s. o.) benutzt werden (vgl. [CM81]), dies ist jedoch keine notwendige Anforderung des BDI-Ansatzes.

Rao und Georgeff (1995) entwickelten eine Reihe von BDI-Logiken, die sie zur Formalisierung der Schlüsseigenschaften der BDI-Architektur benutzten ([Rao91], [Rao92], [Rao93]). Diese Modelle wurden anschließend erweitert, z.B. in Bezug auf die Agentenkommunikation [Had96].

Zur Umsetzung der Intentionen führen einige der auf BDI-Logik aufbauenden Architekturimplementierungen das Konzept der *Ziele* (*goals*) ein. Ziele sind eine in sich konsistente Teilmenge von Intentionen, durch die eine Intention in konkrete, direkt ausführbare Aktionen (Handlungsprimitiven) übersetzt wird. Eine Intention benötigt zu ihrer Umsetzung im Allgemeinen mehrere einzelne Ziele.

Die bekannteste Implementierung des BDI-Modells ist das *Procedural Reasoning System* (PRS), das in [GL87] und [GI89] vorgestellt und anschließend mehrmals re-implementiert (u. a. in C++ und Java [Hub99]) und in einer Reihe von Anwendungen erprobt wurde. Die Entwicklung von BDI-Agenten in Java wird durch die Programmiersprache Jack unterstützt, die Java-Bibliotheken zur Implementierung von BDI-Eigenschaften zur Verfügung stellt [BHRH00].

In [Röl04] wurde ein Petrinetz-Modell des BDI-Agenten erstellt, das wir hier zur Veranschaulichung der BDI-Architektur mit kleinen Änderungen wiedergeben (Abb. 3.2).

Ein Agent perzipiert Informationen aus seiner Umgebung (Transition *wahrnehmen*) und speichert diese in seiner Wissensbasis (an der Stelle *Wissen*) bzw. ersetzt unter Umständen einzelne, nicht mehr aktuelle *Beliefs* durch aktuellere Informationen *Belief Revision*. Aus diesem Wissen kann der Agent *Wünsche* erzeugen (Transition *erzeugen*) und sich aufgrund seines aktuellen Wissenstands auf bestimmte Wünsche *festlegen*. Diese werden dann auf die Stelle *Intentions* kopiert. Durch die Verbindung zur Wissensbasis wird sichergestellt, dass Intentionen sich nicht gegenseitig verhindern bzw. sich widersprechen. Wenn das Wissen eines Agenten durch *Belief Revision* aktualisiert worden ist, kann es vorkommen, dass Intentionen nicht mehr (sinnvoll) verfolgt werden können. Für solche Fälle wurde eine

3. Agenten- und Multiagentensysteme

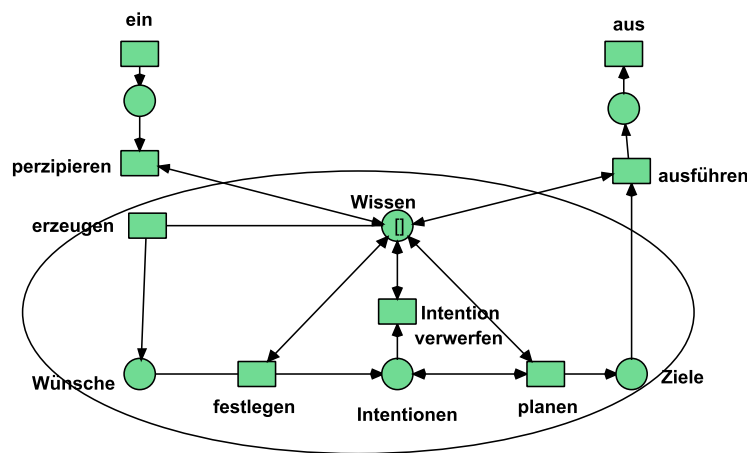


Abbildung 3.2.: BDI-Architektur (aus [Röl04], S. 39).

Transition (*Intention verwerfen*) für die Rücknahme der Intentionen ergänzt. Beim Schalten von *planen* erfolgt die Abbildung der Intentionen auf Handlungsprimitive des Agenten. Direkte Ziele des Agenten können unmittelbar ausgeführt werden. Transition *ausführen* übersetzt Ziele in Aktionen. Der Stand der Umsetzung wird in der Wissensbasis mitgeführt, weshalb den Transitionen jeweils der reservierende Zugriff auf die Stelle *Wissen* gewährt wird, der bei bestimmten Implementierungen u.U. durch Testzugriff ersetzt werden kann.

3.2.6. Hybride Architekturen

Zahlreiche Agentenarchitekturen jüngerer Datums verbinden reaktive und deliberative Prinzipien der Architekturgestaltung in einem Modell. Es entstehen sogenannte *Hybridarchitekturen*, die üblicherweise aus verschiedenen Schichten zusammengesetzt sind. Die Architekturschichten können sowohl horizontal, als auch vertikal angeordnet sein.

Bei der *horizontalen Schichtung* werden die Perzeptionsergebnisse zeitgleich auf alle Architekturschichten übertragen. Diese bearbeiten die Eingaben unabhängig voneinander, jeweils wie ein eigenständiger Agent. Anschließend wird aus den Ausgaben der Schichten ein tatsächlich auszuführender Plan ausgewählt. Eine schematische Darstellung einer horizontal geschichteten Architektur findet sich in Abb. 3.3. In der Abbildung wird die Unabhängigkeit der Verarbeitung auf verschiedenen Ebenen durch die Petrinetz-Semantik der nebenläufig zueinander ausführbaren Transitionen *layer 1* bis *layer n* angedeutet.

Bei der vertikalen Schichtung verläuft die Verarbeitung sequenziell, d.h. die Schichten werden nacheinander aktiv, indem Schicht *n* seine Ausgabe an Schicht *n+1* weiterleitet, die für Schicht *n+1* eine Eingabe darstellt. Eine mögliche Ausprägung der vertikalen Schichtenarchitektur wird in Abb. 3.4 gezeigt.

In Kapitel 7 werden mehrere Beispiele einer geschichteten Architektur vorgestellt, z.B. die TABASCO-Architektur von Staller und Petta [SP98] (siehe Kapitel 7.2.1) bzw. die CogAff-Architektur von Aaron Sloman [Slo01] (siehe Kapitel 7.2.4).

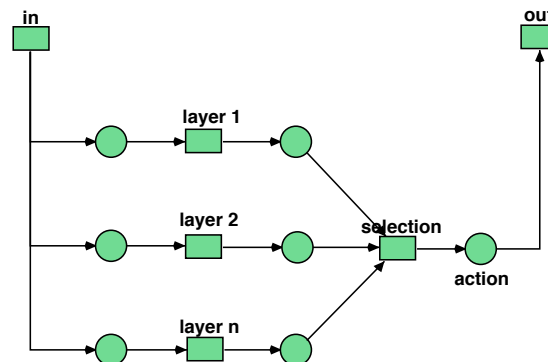


Abbildung 3.3.: Hybride Architektur mit horizontaler Schichtung (aus [Röl04], S. 41).

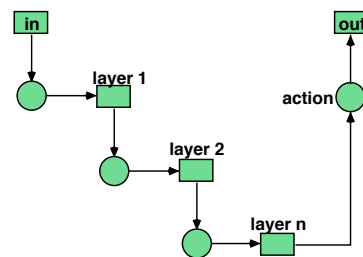


Abbildung 3.4.: Hybride Architektur mit vertikaler Schichtung (aus [Röl04], S. 42).

3.3. Emotionale Agenten

In der aktuellen informatischen Forschung zu emotionalen Agenten gibt es zwar eine Vielzahl von Architekturvorschlägen, eine Formalisierung bzw. Klassifizierung der verschiedenen Architekturen hat bisher jedoch in diesem Bereich kaum stattgefunden. Insbesondere findet man in der aktuellen Literatur keine allgemein anerkannte formale Definition des Begriffs “emotionaler Agent“. Einen der wenigen Versuche einer Definition von emotionalen (BDI-)Agenten findet man in [POM08]:

“Emotional-BDI agents are agents whose behaviour is guided not only by beliefs, desires and intentions, but also by the role of emotions in reasoning and decision-making.“

In der vorliegenden Arbeit wird der Begriff des emotionalen Agenten etwas breiter gefasst: Unter emotionalen Agenten werden solche Agentenarchitekturen verstanden, die durch ihren Aufbau zur Generierung, Verarbeitung und / oder Kommunikation von solchen Zuständen bzw. Prozessen fähig sind, die zu biologisch bzw. psychologisch definierten Emotionen funktional äquivalent sind (vgl. Kapitel 1.5).

In diesem Abschnitt geben wir eine Übersicht über die theoretischen Ansätze, die eine verbesserte bzw. erweiterte Realisierung der grundlegenden Agenteneigenschaften (vgl. die

3. Agenten- und Multiagentensysteme

Darstellung in Kapitel 3.1.2) mit Hilfe der Emotion diskutieren. Ferner werden um eine formalisierte Darstellung der Emotion erweiterte *allgemeine* Agentenmodelle betrachtet. Insbesondere wird auf die in [POM08] erstmalig vorgestellte Beschreibungslogik *EBDI* zur formalen Spezifikation emotionaler BDI-Agenten eingegangen. Eine ausführliche Betrachtung, Diskussion und Auswertung der *konkreten* Architekturen emotionaler Agenten findet in Kapitel 7 statt.

3.3.1. Grundlegenden Agenteneigenschaften und Emotion

In Folgenden wird skizziert, auf welche Weise Emotion in Bezug auf grundlegenden Agenteneigenschaften wie Autonomie, Adaption, Reaktivität / Proaktivität und soziale Kompetenz von Relevanz ist und wie sie für die Realisierung bzw. Optimierung dieser Eigenschaften eingesetzt werden kann (vgl. [TPP03], [OS03]).

Emotion und Autonomie Bei der Beschreibung der grundlegenden Agenteneigenschaften im Kapitel 3.1.2 wurde Autonomie definiert als die Fähigkeit eines Agenten, basierend auf seinen internen Zustand selbstständig und unabhängig Entscheidungen über die auszuführenden Aktionen treffen zu können.

Canamero [Can03] betrachtet das Konzept der Autonomie differenzierter und unterscheidet abhängig von der funktionalen und kognitiven Komplexität des Agenten, von den Eigenschaften seiner Umgebung und von der Art der Agenteninteraktion mit dieser Umgebung drei Autonomiegrade (vgl. [Can03], S. 116):

1. Ein Agent kann bereits als autonom bezeichnet werden, wenn er selbstständig handeln kann, ohne von außen gesteuert werden zu müssen. Eine solche Art der Autonomie korrespondiert mit einem sehr einfachen Verhaltensmuster, wie z.B. dem reflexiven Verhalten, das allein aufgrund von Reaktionen auf äußere Stimuli zustande kommt, wie im Falle von rein reaktiven Agenten. Bei solchen einfachen Stimulus-Reaktion-Verhaltensschemen sieht Canamero keinen Platz und keine Notwendigkeit für Emotion.
2. Ein höherer Autonomiegrad ist erreicht, wenn ein Agent selbst entscheidet, ob er einen aktuellen Umweltstimulus beachten und darauf reagieren soll, und zwar in funktionaler Abhängigkeit von der Relevanz dieses Stimulus für seine Ziele und geplanten Handlungen. Ein Agent mit einer solchen Konzeption von Autonomie - von Canamero als *motivationally autonomous* bezeichnet ([Can03], S. 116) - ist mit dem Problem der Handlungswahl konfrontiert, das darin besteht, zu jedem Zeitpunkt zu entscheiden, welches Verhalten im Hinblick auf das aktuelle Ziel besonders erfolgsversprechend ist. Bei der Entscheidung müssen die Chancen und Risiken der aktuellen Umweltsituation und die langfristigen globalen Ziele berücksichtigt werden.
3. Im Zusammenhang mit der Notwendigkeit der besonders schnellen Reaktion auf dringende interne oder externe Umgebungssignale sieht Canamero eine Einsatzmöglichkeiten von Emotionen bei der Implementierung des Interrupt-Mechanismus (vgl. [Sim67]) in deliberativen autonomen Agenten: "These situations are better dealt with

when a further level of autonomy exists that allows the agent to free itself from the demands of internal needs in order to interrupt ongoing behaviour and change goal priorities when the situation requires it - therefore to adapt quickly to rapidly changing circumstances“ ([Can03], S. 117).

Nach [MB93] besteht der grundsätzliche Unterschied zwischen einem endlichen Automaten mit zustandsabhängigem Verhalten und einem autonomen Agenten in der Fähigkeit des Letzteren zur Selbstkontrolle und Motivation. Motivation wiederum wird definiert als ein interner Prozess, der für die Lenkung der Agentenaktionen hin zur Erfüllung seiner Ziele und insbesondere für eine dynamische Änderung des Agentenverhaltens verantwortlich ist: “Motivation is a driving force (dependent both on the internal state of the agent and the state of the external world) that arouses and directs action towards the achievement of goals“ ([NL95], S. 281).

Neben Motivationen wird das Konzept der Anliegen (*concerns*) als grundlegender Mechanismus zur Steuerung der Agentenaktionen definiert [Fri86] (siehe Kapitel 6.1.2). Ein Anliegen ist die Disposition eines Systems, bestimmte Zustände der Umwelt und des eigenen Organismus der Abwesenheit solcher Zustände vorzuziehen: “A concern can be thought of as a representation of a desired state or event“ [MF00], S. 60).

Emotion passt in dieses Bild als situierter Motivator in Szenarios, in denen ein Agent mit mehreren Anliegen ausgestattet wird, und sich darum bemüht, bestimmte Umgebungszustände zu erreichen bzw. zu vermeiden. Das emotionale System signalisiert die Relevanz der aktuellen Ereignisse für die *concerns* des Agenten und greift auf dessen Reaktionsmöglichkeiten zurück, um den erkannten Chancen und Risiken optimal zu begegnen [PT06], [Pet03].

Unter Einbeziehung des Einflusses, den Emotionen auf interpersonelle Abhängigkeiten und Verpflichtungen in sozialen Szenarios ausüben, kann man eine flexible Administration des Autonomiegrades des Agenten realisieren. In dem Maße, wie die Emotionen sich zur Laufzeit entwickeln und ihrerseits zur Entwicklung eines persistenten Agentenverhaltens über seine gesamte Lebenszeit beitragen, formen sie auch die Basis für die Modellierung der Agentenpersönlichkeit und der Agentenidentität (vgl. [Dam94]).

Emotion und Adaption Unter Adaption wird die Fähigkeit eines Agenten verstanden, sich an neue Umstände und eine veränderte Umgebung anzupassen. Das bedeutet einerseits die Notwendigkeit, “auf unerwartete Eingaben adäquat zu reagieren (kurzfristige Adaption) und langfristigen Veränderungen einer eventuell als solche erkannten Umwelt durch interne Anpassung zu begegnen (langfristige Adaption bzw. Lernfähigkeit)“ ([Röl04], S. 16).

Ein Agent in einer dynamischen Umgebung muss in der Lage sein, die für seine Ziele relevanten Umgebungsveränderungen zu erkennen und sein Verhalten entsprechend anzupassen: “An effective autonomous agent must be capable of identifying changes that are relevant to its purposes and creating appropriate goals in response to these changes“ ([NL95], S. 277). Es wurden mehrere Mechanismen vorgeschlagen, um die Aufmerksamkeit des Agenten auf relevante Umgebungsveränderungen zu lenken, darunter Motivationen [NL95], Anliegen [Fri86] und in neueren Arbeiten zunehmend auch Emotionen (siehe z.B. [BB05] für eine Übersicht).

3. Agenten- und Multiagentensysteme

Da Emotion von der Mehrheit der Autoren als natürlicher Anpassungsmechanismus definiert wird, welches den natürlichen Organismen das Überleben und eine optimale Anpassung an eine dynamische Umwelt ermöglicht (vgl. [Laz91], [Dam94], [Fri86], [Can03]), liegt der Gedanke nahe, einen äquivalenten Mechanismus zur Optimierung der Adaptionsfähigkeit von autonomen Agenten anzuwenden. Nach dem Muster der natürlichen intelligenten Systeme könnten sich Agenten durch schnelle emotionale Reaktionen auf unerwartete Situationen besser an sich schnell verändernde Umweltbedingungen anpassen (*kurzfristige Adaption*). Durch die Generierung und Speicherung der emotionalen Reaktion, gekoppelt an die emotionsauslösende Situation, könnte eine *langfristige Adaption* modelliert werden [Pet03].

Emotion und Reaktivität bzw. Proaktivität Ein Agent muss auf jede mögliche Eingabe in geeigneter Weise *reagieren* können. Die Aktivitäten eines Agenten, denen kein unmittelbar auslösendes Ereignis zugrunde liegt, nennt man *proaktiv*. Die Notwendigkeit, flexibel auf externe Ereignisse reagieren zu können und die Reaktionen an neue Anforderungen anzupassen (vgl. Adaption), kann durch die statische und zur Übersetzungszeit festgelegte Zuordnung von Ereignis und Reaktion nicht geleistet werden ([Röl04], S. 16).

Eine der zentralen Probleme beim Design eines Agentensystems besteht darin, ein Gleichgewicht zwischen zielgerichtetem und reaktivem Verhalten zu finden ([Woo02], S. 24). Ein Agent soll systematisch versuchen, seine Ziele zu erreichen, z. B. indem er eine komplexe Abfolge von Aktionen ausführt. Jedoch soll er nicht blindlings mit der Ausführung dieser Aktionsfolge fortfahren, wenn z.B. klar wird, dass dies nicht zum gewünschten Ergebnis führt bzw. wenn das Ziel, das durch die Aktionsabfolge erreicht werden sollte, nicht mehr gültig ist. Unter diesen Umständen soll der Agent in angemessener Zeit auf den geänderten Umgebungszustand reagieren. Jedoch ist es wiederum nicht wünschenswert, dass der Agent sich ausschließlich reaktiv verhält und sich auf ein Ziel nicht lange genug konzentriert, um es erreichen zu können. Das Problem des optimalen Verhältnisses zwischen reaktivem und zielgerichtetem Verhalten ist noch weitgehend offen.

Hier könnte die Integration von Emotion eine Abhilfe schaffen. Emotion kann als schnelle adaptive Reaktion auf eine Situation bzw. ihre Konsequenzen charakterisiert werden. Das impliziert, dass Emotion eine operationalisierte Abstraktion von spezifischen Situationsgegebenheiten darstellt, sodass die gleichen objektiven Stimuli unterschiedliche Emotionen bei verschiedenen Individuen bzw. sogar bei demselben Individuum auslösen können (dann typischer-, aber nicht notwendigerweise zu verschiedenen Zeiten). Diese Eigenschaft der Emotion ist für die Realisierung einer sehr flexiblen Stimulus-Response-Abbildung von großer Bedeutung. Die Fähigkeit, adaptive Reaktionen in Akteuren einer spezifischen Domäne hervorzurufen, deutet auf eine effiziente Integration der statischen (zur Übersetzungszeit festgelegten) und dynamischen Laufzeitlösungen hin (vgl. [LeD96a], [Pet03], [PT06]).

Emotion und soziale Kompetenz Die Relevanz der Emotionen für soziale Fähigkeiten zeigt sich in der beachtlichen Anzahl soziologischer Theorien, die auf Emotionen als bestimmenden Faktor in den interpersonalen Beziehungen zurückgreifen. So definiert der *Weak Social Constructionism* Emotionen vor allem als soziale Konstrukte (siehe Kapitel 6.3).

Nach dieser Theorie zeigen die Individuen nur eine beschränkte Anzahl von natürlichen emotionalen Reaktionen außerhalb des sozialen Kontextes. Als soziale Konstrukte, beruhen Emotionen auf tieferen Modellen des eigenen Selbst, des Selbst der anderen usw. Ähnlicherweise erreichen die interpersonellen Emotionsmodelle mit nur einer kleinen Anzahl von erzeugenden Elementen ein signifikantes Erklärungspotential (vgl. z.B. [Kem78]).

Ein weiterer Themenkomplex beruht auf dem Zusammenspiel von Emotionen und sozialen Normen (siehe Kapitel 6.3.3). Emotionen werden als Mechanismen der Belohnung und Bestrafung bei instrumenteller Konditionierung des Individuums auf die sozialen Normen gesehen. Umgekehrt dienen soziale Normen der Regulation der Emotionen.

3.3.2. Emotionale BDI-Agenten

In [POM08] wird eine multi-modale Logik *EBDI* zur formalen Spezifikation emotionaler BDI-Agenten vorgestellt. Die vorgeschlagene Spezifikation ermöglicht zunächst nur die Modellierung von Furcht (*Fearful Agents*) und definiert die für diese Emotion spezifischen Auslösebedingungen und Aktionen. Jedoch ist eine Erweiterung des vorgestellten Modells auf weitere Emotionen denkbar. Im Folgenden werden die wesentlichen Modellelemente vorgestellt.

Zusätzlich zu den Standardkomponenten der herkömmlichen BDI-Architektur (siehe Kapitel 3.2.5) werden in [POM08] weitere Konzepte definiert:

- Die Menge *CAP* der „Fähigkeiten“ (*capabilities*) eines Agenten beschreibt sein abstraktes Vermögen, Veränderungen in seiner Umgebung herbeizuführen. Dieses Konzept erinnert stark an die abstrakten Pläne bzw. Aktionen eines herkömmlichen BDI-Agenten: “In fact, we can consider the set of capabilities as a dynamic set of plans which the agent has available to decide what to do in each of its execution states“ ([POM08], S. 71).
- Die Menge *RES* - *Ressourcen* - umfasst physikalische bzw. virtuelle Voraussetzungen für die Ausführung von Aktionen. Wenn eine Ressource für die Durchführung einer Aktion α nicht vorhanden ist, ist der Erfolg dieser Aktion nicht gesichert.
- *Furcht* (*fear*) wird definiert als ein modaler Operator $\text{FEAR}(\sigma)$, welcher als “Agent fürchtet, dass σ wahr ist“ interpretiert wird.
- Die Menge *FDES* der *grundlegenden Wünsche* (*Fundamental Desires*) sind spezielle *Desires*, die unter allen Umständen von Agenten verfolgt und als erfüllbar erhalten werden müssen: “Fundamental desires are special desires which are vital desires of the agent, or desires which cannot be failed to achieve, in any condition, since may put in cause the agent’s own existence. Fundamental desires must always be true and the agent must always do its best to maintain them valid“ [POM08], S. 72).

Zusätzlich werden verschiedene Arten von Auslösebedingungen für Furcht unterschieden: *Bedrohungen* (*Threats*) und *unangenehme Ereignisse* (*Unpleasant Events*) repräsentieren Fakten in der Umgebung, welche die Erfüllbarkeit eines oder mehrerer grundlegender Wünsche (Fundamental Desires) eines Agenten in Frage stellen und daher vom Agenten

3. Agenten- und Multiagentensysteme

als unmittelbar bedrohlich für seine Existenz bewertet werden. Die beiden Arten von Bedrohungen können unterschiedlich gewichtet werden (*dangerous, serious, possible, highly unpleasant, strongly unpleasant, possibly unpleasant*).

Ferner werden spezielle Aktionen bzw. Verhaltensweisen definiert, die ein Agent unter bestimmten Bedingungen ausführt. Diese spezifizierten Verhaltensweisen werden abstrakt durch atomare Aktionen repräsentiert, stellen aber eher komplexe Planungs- und Verhaltensstrategien dar, deren Auswahl durch Emotion gesteuert wird.

- **Selbsterhaltung** (*spresserv*¹): Dieses Verhalten wird aktiviert, wenn ein Agent fürchtet, einige seiner fundamentalen Wünsche nicht erfüllen zu können.
- **Motivated Processing Strategy** (*mpsdel*): Diese Verarbeitungsstrategie wird eingesetzt, wenn ein Wunsch des Agenten bedroht ist, das aktuell sein Verhalten bestimmt. Diese Strategie ist besonders berechnungsintensiv, da der Agent Möglichkeiten für die Aufrechterhaltung seines Wunsches ableiten muss.
- **Direct Access Strategy** (*dasdel*): Diese Strategie gründet auf der Existenz der vordefinierten Strukturen bzw. des Wissens. Es ist die einfachste Strategie, die am wenigsten Berechnungsressourcen erfordert und zu schnellen Lösungen führt.

Nun kann ein emotionaler BDI-Agent durch die Vereinigung der oben genannten Axiomengen beschrieben werden, die um folgende zusätzliche Axiome erweitert werden:

1. Jede Bedrohung ψ eines fundamentalen Wunsches φ löst beim Agenten Furcht vor $\neg\varphi$ aus:

$$Threat(\psi, \varphi) \rightarrow FEAR(\neg\varphi).$$

2. Jede Bedrohung ψ eines fundamentalen Wunsches φ löst beim Agenten ebenfalls Furcht aus:

$$Threat(\psi, \varphi) \rightarrow FEAR(\psi).$$

3. Wenn ein Agent nach einem Deliberationsprozess unter Benutzung der “Motive Processing Strategy“ zum Schluss kommt, dass entweder die Bedrohung ψ sich nicht bestätigt oder dass sein fundamentaler Wunsch φ mit ψ kompatibel ist, dann bewertet er ψ als einen “Unangenehmen Fakt“ in Bezug auf die Erfüllung von φ :

$$< mpsdel > ((BEL(\neg\psi) \vee BEL(\psi \wedge \varphi)) \wedge DES(\varphi)) \rightarrow Unpleasant(\psi, \varphi).$$

4. Wenn ein Agent bedroht wird und über keine freien Ressourcen für die Ausführung der “Direct Access Strategy“ verfügt, kann er eine selbsterhaltende Strategie einsetzen:

$$Threat(\psi, \varphi) \wedge \neg EffCap < dasdel > \rightarrow < spserv > \top.$$

¹In den Klammern sind die Bezeichnungen der entsprechenden atomaren Aktionen in *EBDI* angegeben.

5. Wenn ein Agent über freie Ressourcen verfügt, wird er die “Direct Access Strategy“ für seine Deliberationsprozesse einsetzen:

$$Threat(\psi, \varphi) \wedge EffCap < dasdel > \rightarrow < dasdel > \top.$$

Auf diese Weise lassen sich weitere Bedingungen für die emotionsbasierte Wahl der Verhaltens- bzw. Planungsstrategie formal darstellen.

Bewertung Grundsätzlich ermöglicht das Modell von [POM08] die formale Darstellung des funktionalen Einflusses der Emotion auf das Agentenverhalten und auf Deliberationsprozesse. Es bietet eine gute Basis für die Formalisierung der Emotion und ihre Einbindung in den BDI-Kontext, weist jedoch keine emotionstheoretische Fundierung für die spezifizierten Mechanismen der Emotions- und Aktionsauslösung auf, d.h. der Wahl einer Verhaltens- bzw. Planungsstrategie in Abhängigkeit ist mit keiner psychologischer bzw. sozialwissenschaftlicher Emotionstheorie begründet. Ferner ist die Spezifizierung bisher nur auf eine Emotion (Furcht) beschränkt.

Im Kapitel 6.1.2 wird ein weiteres BDI-basiertes Modell eines emotionalen Agenten vorgestellt, das auf Basis eines emotionstheoretischen Ansatzes von [Fri86] bzw. [FM93] abgeleitet wird.

3.4. Multiagentensysteme

Mehrere autonome Agenten können in ein Multiagentensystem eingebunden werden, das entsprechend leistungsfähiger, aber auch schwieriger zu beherrschen ist. Die meisten Autoren sehen ein MAS als ein aus mehreren Agenten zusammengesetztes System, in dem die Agenten kommunizieren und interagieren, um spezifische persönliche oder kollektive Aufgaben zu erfüllen.

Ferber (1995) definiert ein MAS als System, das aus den folgenden Elementen besteht [Fer95]:

- einer Umwelt;
- einer Menge von situierten Objekten (“situiert“ bedeutet hier, dass den Objekten zu jedem Zeitpunkt eine Position zugeordnet ist). Die Agenten können Objekte wahrnehmen, erzeugen, modifizieren und löschen. Die Kommunikation zwischen den Agenten erfolgt über den Austausch von Nachrichten (ebenfalls Objekte);
- einer Menge von Agenten, die als aktive Objekte angesehen werden;
- einer Menge von Beziehungen, die die Objekte miteinander verbindet;
- einer Menge von Operatoren, mit denen Agenten Objekte empfangen, erzeugen, konsumieren, verändern und löschen können. Zusätzlich gibt es Operatoren, die die Veränderungsversuche an Objekten und die Reaktionen der Umwelt auf diese darstellen.

3. Agenten- und Multiagentensysteme

In [Wei99] werden folgende allgemeine Eigenschaften von Multiagentensystemen definiert: “The characteristics of MAS are:

- each agent has incomplete information, or capabilities for solving the problem, thus each agent has a limited viewpoint;
- there is no global system control;
- data is decentralized; and
- computation is asynchronous“.

Ausgehend von diesen Charakteristiken können verschiedene Ausprägungen von MAS beschrieben werden:

Situierte Agentensysteme interagieren mit der Umwelt. Als *rein situierte* Agentensysteme werden Systeme beschrieben, in denen die beteiligten Agenten nicht unmittelbar über Nachrichten miteinander kommunizieren, sondern sich nur über die jeweils verursachten Veränderungen in der Umwelt wahrnehmen.

Werden Nachrichten zur Kommunikation verwendet, wird das System als *kommunizierendes* Multiagentensystem bezeichnet. Als *rein kommunizierendes* Multiagentensystem wird ein System bezeichnet, in dem die Umwelt nur aus Agenten besteht, und die Agenten (nur) über Nachrichten miteinander kommunizieren. Diese Art von Systemen wird häufig in der Verteilten Künstlichen Intelligenz verwendet. In [Fer95] wird der Begriff des Softwareagenten synonym zu rein kommunizierenden Agenten benutzt.

3.4.1. Koordination und Kommunikation

Agentenkoordination ist ein wichtiges Thema innerhalb der Multiagentenforschung. Bereits bei der Vorstellung des Forschungsgebiets VKI im Abschnitt 2.2 und anschließend bei der Diskussion der Agenteneigenschaften im Abschnitt 3.1.2 wurde deutlich, dass *Koordination*, d.h. Interaktion mit anderen Agenten, als eine grundlegende Eigenschaft von Multiagentensystemen gesehen wird. Die Zusammenfassung von mehreren Agenten reicht an sich noch nicht aus, um ein Agentensystem zu bilden. Dazu ist die Möglichkeit einer gegenseitigen Abstimmung zwischen den Agenten eines Multiagentensystems notwendig, um Mehrwert gegenüber einem einzelnen Agenten in Bezug auf die Leistungsfähigkeit des Systems zu erreichen.

Das Spektrum der verschiedenen Ansätze zur Realisierung der Koordination ist breit. In [HS99] wird Koordination in *Kooperation*, d.h. das gemeinsame Lösen eines Problems, und *Konkurrenz*, bei der Agenten gegeneinander arbeiten, unterteilt. Bei der Kooperation wird weiterhin *zentrales* und *verteiltes* Planen unterschieden. Beim zentralen Planen wird zentral ein verteilt auszuführender Plan erstellt, wogegen beim verteilten Planen auch die Planerstellung verteilt abläuft. Konkurrenzsituationen entstehen, wenn Agenten individuelle Ziele verfolgen, die sich widersprechen bzw. sich gegenseitig ausschließen. Solche Situationen werden durch *Verhandlungen* gelöst. Unter diesem Begriff werden im Bereich der Multiagentenforschung viele unterschiedliche Kommunikationsprozesse verstanden, zu denen Auktionen, Koalitionsbildungen, Verträge u.a. gehören.

Eine formale Behandlung von Verhandlung erlauben die Ansätze der Spieltheorie, die eine genaue Analyse verschiedener Situationen und präzise Voraussagen über die möglichen Ausgänge bei verschiedenen Handlungsalternativen ermöglichen. Spieltheoretische Ansätze gehen jedoch von einem vereinfachten Agentenmodell aus: Verhandlungspartner handeln aus dieser Sichtweise rein rational, haben beliebig viel Rechenzeit zur Verfügung und folgen strikten Verhandlungsprotokollen. Der emotionale Aspekt der Entscheidungsfindung wird bei spieltheoretischen Ansätzen komplett vernachlässigt.

3.4.2. Agentenkommunikationssprachen

Die Koordination von Agenten in einem Multiagentensystem ist nur unter der Voraussetzung realisierbar, dass in diesen Systemen die Möglichkeit der Kommunikation gegeben ist. Denn lediglich durch Kommunikation können die Agenten eines Systems gemeinsam an der Lösung eines Problems arbeiten bzw. in einer Konkurrenzsituation zu einem Kompromiss finden.

Die Kommunikation in Multiagentensystemen kann auf unterschiedliche Weise realisiert werden: Viele Systeme verwenden (asynchronen) Nachrichtenaustausch in einer eigens dafür entworfenen Agentenkommunikationssprache. Die Definition einer Agentenkommunikationssprache ist deshalb ein wichtiges Thema in der MAS-Forschung.

Um Interoperabilität von Agentenimplementationen verschiedener Hersteller zu erreichen, sind Standards auf dem Gebiet der Agentenkommunikation notwendig. Es wurden deshalb mehrere Versuche unternommen, eine standardisierte Kommunikationssprache zu entwickeln.

So existieren Standardisierungsinitiativen, die sich zum Ziel gesetzt haben, die Entwicklung von allgemeinen Spezifikationen aus dem Gebiet der Agententechnologien voranzutreiben und Standardisierungsvorschläge für den Einsatz von Agenten im industriellen Umfeld auszuarbeiten. Außer der Entwicklung einer standardisierten Kommunikationssprache wird an einem allgemeinen Referenzmodell für agentenbasierte Software-Architekturen gearbeitet. Die heute bedeutendste Standardisierungsinitiative im Bereich der Agententechnologien ist die FIPA (Foundation for Intelligent Physical Agents). Weitere Standardisierungsvorschläge stammen von der OMG (Object Management Group).

Den ersten Versuch einer Vereinheitlichung der Kommunikationssprache von Agenten stellt die KQML (*Knowledge Query and Manipulation Language*) dar, die ursprünglich zu Datenabfrage, -austausch und -bearbeitung dienen sollte (siehe [Lab96], [Lab01]).

Durch die Weiterentwicklung von KQML wurde die FIPA ACL (*FIPA Agent Communication Language*) geschaffen. Mit FIPA ACL steht eine Agentenkommunikationssprache mit vollständiger formaler Spezifikation zur Verfügung [fIPA02b]. Sie basiert auf der Sprechakttheorie [Aus72] und setzt eine Drei-Schichten-Architektur um, in der Inhalt, Bedeutung und Sprechakt einer Nachricht unterschieden werden (siehe Kapitel 3.4.3 für weitere Details). Die ACL ist inzwischen zum De-facto-Standard auf dem Gebiet der Agentenkommunikation geworden.

Mit OMG MASIF (*Mobile Agent System Interoperability Facility*) wurde ein Standardisierungsvorschlag vorgestellt, der explizit auf die Modellierung von mobilen Agenten ausgerichtet ist. MASIF verwendet die *Common Object Request Broker Architecture* (CORBA, ebenfalls von der OMG spezifiziert) zur Kommunikation und zum Transport der

3. Agenten- und Multiagentensysteme

Agenten. Es handelt sich bei CORBA um eine objektorientierte Middleware, die plattformübergreifende Protokolle und Dienste definiert. Ziel dieser Architektur sind verteilte Anwendungen in heterogenen Umgebungen. Der MASIF-Standard definiert eine generelle Architektur und eine eigene Schnittstelle in der von CORBA verwendeten Interface Definition Language (IDL), die kompatible Plattformen implementieren müssen.

3.4.3. FIPA-Referenzarchitektur

Neben der Spezifikation der Agentenkommunikationssprache FIPA-ACL beinhalten FIPA-Spezifikationen Standardisierungsvorschläge für die MAS-Architektur, insbesondere ein Referenzmodell für die Standardisierung von Plattformdiensten. Der FIPA-Standard für Agentenplattform wird durch die folgenden Elemente festgelegt (vgl. [FIPA02a]):

- Ein Agent wird in der FIPA-Spezifikation als eine autonome und kommunizierende Einheit definiert, die bestimmte Fähigkeiten kapselt, welche anderen Agenten als Dienste angeboten werden können. Ein Agent verpflichtet sich, die von ihm offerierten Dienste zur Verfügung zu stellen, deshalb muss über die Bereitstellung von Diensten nicht verhandelt werden. Dadurch wird zwar der Kommunikationsaufwand vermieden, der bei Verhandlungen über die Bereitstellung von Diensten anfällt, gleichzeitig wird jedoch auch die Autonomie des Agenten begrenzt.
- Über eine Schnittstelle benutzen Agenten die Funktionen der Plattform. Die Agentenplattform stellt die Umgebung der Agenten dar und bietet Dienste wie Nachrichtentransport, Kommunikationskanäle zu anderen Plattformen sowie Verzeichnis- und Verwaltungsdienste an. Die FIPA-Spezifikation macht keine Angaben über die Realisierung der Dienste einer Agentenplattform, sie werden nur über ihre Funktionalität und die Syntax ihres Aufrufs beschrieben. Die Plattform kann auch verteilt realisiert werden. Der Aufbau einer Agentenplattform nach FIPA wird in der Abb. 3.5 gezeigt.
- Das *Agenten-Management-System (AMS)* wird durch spezielle Agenten realisiert, die die Agentenverwaltung übernehmen und den kompletten Lebenszyklus jedes Agenten überwachen. Das AMS unterhält ein Verzeichnis der Identifikatoren aller auf der Plattform residierenden Agenten.
- Beim *Directory Facilitator (DF)* werden die Services registriert, die die jeweiligen Agenten anbieten. Der DF bildet eine Art gelber Seiten für die Agentenplattform. Die beim DF registrierten Dienste können auch von Agenten angesprochen und in Anspruch genommen werden, die auf anderen Plattformen residieren.

Agenten können die über den DF angebotenen Dienste identifizieren und ihre Ausführung anfragen. Diese Dienste werden nur angefragt und nicht aufgerufen, da der Agent die Kontrolle behält. Aufgrund seines inneren Zustandes kann er sich gegen eine Ausführung des Service entscheiden und diesen ablehnen.

- Das interne Nachrichtentransportprotokoll *Message Transport Service (MTS)* bietet einen Nachrichtenaustauschdienst für Agenten innerhalb des Systems. Alle Agenten

auf einer Plattform haben einmalige Identifikatoren, die die eindeutige Zuordnung und Zustellung von Nachrichten innerhalb der Plattform möglich machen.

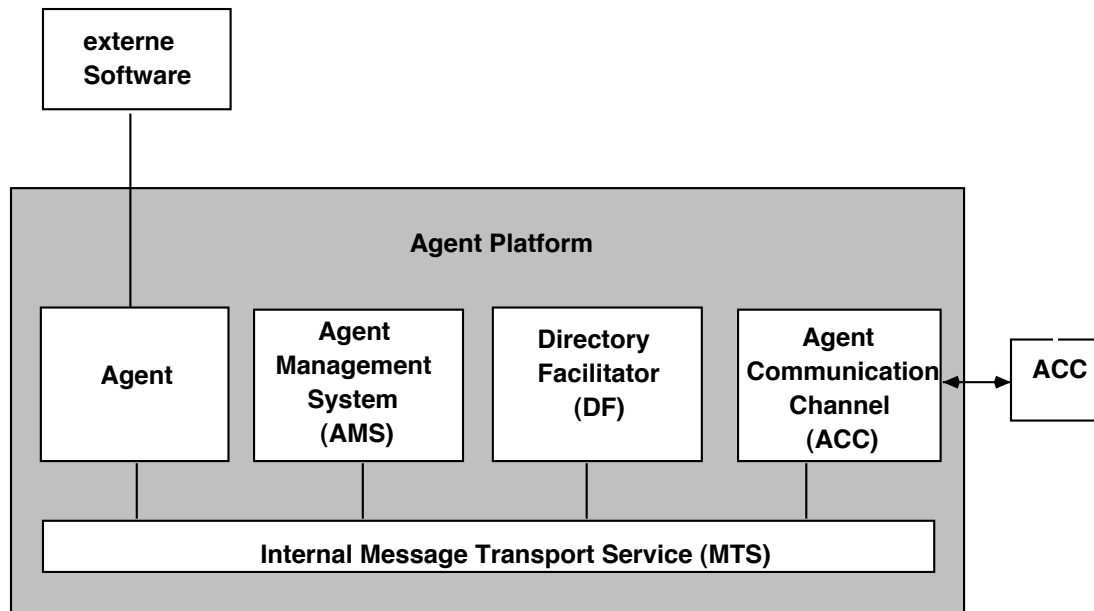


Abbildung 3.5.: Agentenplattform nach FIPA.

Die Kommunikation zwischen den Agenten erfolgt über den *Agent Communication Channel (ACC)*. Der ACC ermöglicht auch für Agenten auf verschiedenen Plattformen eine zuverlässige Kommunikation. Die Agenten kommunizieren über Nachrichten. Der Austausch von Nachrichten erfolgt auf der Anwendungsebene mittels Kommunikationsprotokollen im ACL-Format. Bei ACL wird zwischen der Bedeutung einer Nachricht und seinem Inhalt unterschieden (vgl. Kapitel 3.4.2). Der Format der Nachrichten und deren mögliche Bedeutung sind in weiteren FIPA-Spezifikationen definiert [fIPA02b]. Die Informationen werden in Prolog übertragen und es wird eine Ontologie zur Interpretation der Informationen angegeben.

Gegenwärtig existieren über 40 verschiedene Multiagentenplattformen (siehe z.B. [KNB⁺03] bzw. [PBL05] für eine Übersicht). Grundsätzlich lassen sich zwei Kategorien von Plattformen unterscheiden:

Die erste Kategorie fokussiert die Kommunikations- und Plattformstandards nach FIPA (z.B. JADE [BPR01], FIPA-OS [Lim01]). Grundsätzlich bieten Plattformen, die FIPA-Standards implementieren, die Möglichkeit, Agenten zu erzeugen und zu löschen, andere Agenten zu lokalisieren und Kommunikationsmöglichkeiten zwischen den Agenten zu schaffen.

Die zweite Plattformkategorie unterstützt agenteninterne Strukturen durch mentalistische Konzepte (z.B. JACK [BRHL99], JAM [Hub99], AgentBuilder [RS04]). Letztere beruhen auf

3. Agenten- und Multiagentensysteme

unterschiedlichen Agentensprachen, da kein Konsens darüber besteht, welche mentalistischen Begriffe am Besten für die Beschreibung dieser Strukturen geeignet sind.

3.4.4. Modellierungstechniken für MAS

Neben der technischen Unterstützung durch standardisierte Plattform- und Infrastrukturdienste sind etablierte Entwicklungsmethoden eine zwingende Voraussetzung für die Weiterentwicklung und Akzeptanz von Agententechnologien. In den vergangenen Jahren wurde eine Reihe von Vorschlägen für solche Entwicklungsmethoden (z.B. Gaia [WJK00], Tropos [GMP02], PASSI [CP02], MASSIVE [Lin01]) und Modellierungstechniken (z.B. AUML [OPB01]) veröffentlicht. Einen Überblick über (teilweise ältere) Methoden bieten [WD01], [IGG99] und [KNB⁺03].

Gaia [WJK00] ist einer der bekanntesten Ansätze für agentenorientierte Analyse und Entwurf. Es wird darin eine Entwicklungsmethodologie vorgeschlagen, bei der die Entwicklung eines Multiagentensystems auf Basis von Rollen ausgetragen wird. Das System wird in eine Reihe von Rollen unterteilt, die die Grundlage für die Agenteninteraktion bilden. Der Entwicklungsprozess beginnt mit der Anforderungsanalyse und liefert als Ergebnis einen über eine Reihe verschiedener Modelle verfeinerten Entwurf als Basis für die Implementierung.

Der Schwerpunkt der TROPOS-Methodologie [GMP02] liegt auf dem Requirement Engineering und der Spezifikation des zu entwickelnden Systems. TROPOS bietet eine Notation für die Definition von Agenten, sowie Agentenzielen, -plänen und -wissen. Das Prozessmodell von TROPOS umfasst alle Entwicklungsphasen von der Anforderungsspezifikation über die Architekturentwicklung bis hin zur Implementierung.

Der *Process for Agent Societies Specification and Implementation* (PASSI) [CP02] erstreckt sich von der frühen Anforderungsanalyse bis zur Implementierung. Dabei wird meist UML verwendet und es werden Entwurfstechniken aus der Softwaretechnik und der Künstlichen Intelligenz zusammengeführt und durch ein Zusatzmodul für das Modellierungswerkzeug Rational Rose unterstützt.

Das *Multi-Agent SystemS Iterative View Engineering* (MASSIVE) [Lin01] dagegen präferiert eine modellbasierte Vorgehensweise. In Form von Sichten werden verschiedene, iterativ zu verfeinernde Modelle für das MAS angelegt. Ein Schwerpunkt liegt dabei auf der Wiederverwendung von Entwurfsfragmenten durch den Aufbau einer Sammlung von Entwurfsmustern.

Es wurde mehrfach untersucht, wie UML an die Anforderung der agentenorientierten Modellierung angepasst werden kann. Dabei kamen mehrere Autoren zu dem Schluss, dass Multiagentensysteme mit den bestehenden UML-Konzepten nicht adäquat modelliert werden können ([BOvD00], [OPB01]) und eine Erweiterung der Modellierungssprache um zusätzliche (agentspezifische) Komponenten notwendig ist.

FIPA hat UML als Grundlage für die Entwicklung der speziell für die Modellierung von MAS konzipierten Sprache *Agent UML* (AUML) genutzt. Die Grundkonzepte von Gruppen und Rollen bilden die Grundlage der AUML-Spezifikation ([BMO01], [BOvD00], [OPB01]). Die Spezifikation ist zurzeit noch nicht vollständig ausgearbeitet. Im folgenden Kapitel wird AUML in Bezug auf die Modellierung von Agentensystemen diskutiert und mit anderen Modellierungstechniken, insbesondere Petrinetzen, verglichen (siehe Kapitel 4.1.3).

4. Modellierungstechnik

In diesem Kapitel werden Netzformalismen vorgestellt, die in dieser Arbeit als Modellierungstechniken einerseits für die Darstellung sozialwissenschaftlicher Theorien, andererseits für die Repräsentation der (Multi-)Agentensysteme verwendet werden sollen.

In Zusammenhang mit der Verwendung der Petrinetze als Basis für eine Modellierungssprache, die zwischen Modellen der Soziologie und Informatik vermitteln soll, wurde in [vLMV03] der Begriff *vermittelnde Petrinetze* (kurz: *vermittelnde Netze*) eingeführt. Vermittelnde Netze basieren auf dem formalen Modell der Petrinetze, schließen aber auch höhere Netzklassen wie “*Netze-in-Netzen*“ nach Valk ([Val98], [Val00a]) und *Referenznetze* nach Kummer ([Kum02]) mit ein. Indem die einfacheren Netzarten schrittweise in die höheren Modellierungstechniken überführt werden, wird bei der Modellierung die Darstellung des Modellgegenstandes stufenweise präzisiert.

In weiten Teilen dieser Arbeit (Kapitel 5 bis 7) wird eine äquivalente Vorgehensweise benutzt, um die Aussagen der sozialwissenschaftlichen Emotionstheorien und die Architekturen emotionaler Agenten mit der gleichen Modellierungstechnik darzustellen.

Im Abschnitt 4.1 werden daher zunächst die verschiedenen Netzformalismen dargestellt, die als vermittelnde Netze in dieser Arbeit Einsatz finden. Anschließend wird ein petrinetz-basierter Modellierungsansatz für MAS und seine konkrete Umsetzung im Multiagentensystem MULAN vorgestellt (Kapitel 4.2). Schließlich wird das methodische Vorgehen bei der Modellierung von sozialwissenschaftlichen Theorien erläutert (Kapitel 4.4). Wir benutzen den in [vLMV03] begründeten und in weiteren Arbeiten (z.B. [KR02], [KMRV05]) mehrfach angewendeten *sozialwissenschaftlichen Modellierungsansatz* für die Darstellung der sozialwissenschaftlichen Theorien mit informatischer Modellierungstechnik der Petri- bzw. Referenznetze.

4.1. Benutzte Formalismen

In dieser Arbeit werden Petrinetze - wie in der Informatik generell - zur Formulierung von Systemmodellen eingesetzt. Dabei werden je nach Modellierungsgegenstand unterschiedliche Petrinetz-Formalismen verwendet, insbesondere Platz- / Transitions-Netze (P/T-Netze) ([JV80]), gefärbte Netze ([Val93]), Objektnetze ([Val96], [Val98], [Val00c], [Val00b], [Val00a]), Referenznetze ([KLL⁺00], [Kum98], [Kum00] und [KWDC09]) und schließlich agentenorientierte Petrinetze ([MW97], [Röl99], [KMR01a]). In diesem Abschnitt werden grundlegende Eigenschaften der Petrinetze und einiger weiterer Netzarten, die in dieser Arbeit Anwendung finden, beschrieben.

4. Modellierungstechnik

4.1.1. Petrinetze

Petrinetze sind eine formal exakt definierte, operationale und visuelle Beschreibungsmethode für nebenläufige verteilte Systeme [Jen92]. Sie gehen auf grundlegende Arbeiten von Carl Adam Petri zurück und stellen eine bewusste Abwendung von klassischen Modellierungstechniken durch sequenzielle endliche Automaten und Turing-Maschinen zu einer kommunikationsorientierten Sicht dar [Pet62]. Bedingt durch die formale Semantik können Petrinetze formal analysiert und ihre jeweiligen relevanten Eigenschaften verifiziert werden. Wesentliche Konzepte der Verifikation wie Invarianten, Lebendigkeit oder Fairness, sind fest in der Petrinetzforschung verankert.

Ein Petrinetz ist ein gerichteter Graph, der aus passiven und aktiven Elementen (entspr. (*Plätzen*) und *Transitionen*) besteht. Ein Platz stellt in etwa einen Speicher bzw. eine Ablage für Ressourcen dar, während eine Transition eine Handlung bzw. einen Verarbeitungsschritt repräsentiert. Aufbauend auf dieser statischen Struktur wird das dynamische Verhalten durch die Belegung der Plätze mit Marken erzeugt. In einfachen P/T-Netzen sind die Marken nicht unterscheidbar [JV80], während höhere Petrinetzformalismen eine Typisierung der Marken erlauben (vgl. gefärbte Netze).

Eine spezielle Belegung aller Plätze eines Netzes mit Marken (*Markierung*) stellt einen *Zustand* des durch das Netz modellierten Systems dar. Der Übergang von einer Markierung zu einer anderen ist durch das *Schalten* (auch *Feuern*) einer Transition möglich: Eine Transition schaltet, falls alle Plätze ihres Vorbereiches mit passenden Marken versehen sind. In diesem Sinne kann die Markierung der Plätze des Vorbereichs als Erfüllung von Bedingungen bzw. als Vorhandensein von Ressourcen interpretiert werden. Die Marken werden beim Schalten abgezogen und alle Plätze aus dem Nachbereich der Transition mit neuen Marken belegt. Das Schalten einer Transition entspricht somit der Ausführung einer Handlung bzw. einem Prozessschritt, für den bestimmte Bedingungen erfüllt sein müssen bzw. Ressourcen benötigt werden, welche durch die Ausführung des jeweiligen Prozessschrittes “verbraucht” werden. Um das Schalten einer Transition zu ermöglichen, müssen zusätzlich die evtl. vorhandenen Schaltbedingungen (*guards*) erfüllt sein. Abb. 4.1 veranschaulicht das Schalten einer Transition mit drei Plätzen im Vorbereich (p1, p2, p3) und zwei Transitionen (p4, p5) im Nachbereich. In höheren Petrinetzformalismen, z.B. in Referenznetzen, können die abgezogenen und die neu hinzugefügten Marken zusätzlich durch Kanten- und Transitionbeschriftungen beeinflusst werden (s.u.).

Petrinetze wurden traditionell als eine Symbolsprache zur Darstellung von relationalen Beziehungen und speziell von prozessoralen Entitäten verstanden. Eingesetzt werden Petrinetze vorwiegend zur Fixierung einer operationalen Semantik von Systemmodellen. Eine Einführung in Petrinetze findet sich z.B. in [GV03].

4.1.2. Objektnetze, Referenznetze

Carl Adam Petri [Pet62] hat neben seiner *Allgemeinen Systemtheorie*, auf die die Theorie der Petrinetzprozesse zurückgeht, bereits einige weitere Ausprägungen von Petrinetz-Modellen erarbeitet. Ausgehend von dem Basismodell der Bedingungs-/ Ereignis-Netze (später *Elementare Netzsysteme* genannt) wurden die *Platz- / Transitions-Netze* [JV80] weltweit

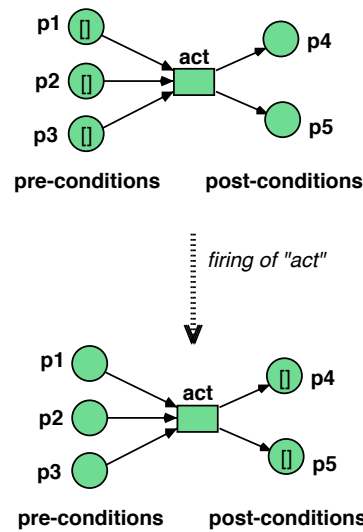


Abbildung 4.1.: Schalten eines einfachen P/T-Netzes (aus [GV03]).

bearbeitet. Durch die Einführung individueller Marken bei *Prädikat- / Transitions-Netzen* und *gefärbten Netzen* wurde ein weiterer Abstraktionsgrad erreicht. In diesen Petrinetzmodellen symbolisieren Marken Objekte in unterschiedlicher Abstraktion. In *Objektnetzen* wird der Gedanke dahingehend fortgeführt, dass dem Datenobjekt weitere Eigenschaften zugesprochen werden, wie zum Beispiel ein eigenes dynamisches Verhalten, das dem dynamischen Verhalten des zu modellierenden realweltlichen Objekts zugeordnet wird. Objekte werden hier als Marken in einem übergeordneten Systemnetz modelliert.

Abb. 4.2 veranschaulicht ein *Objektnetz*, das sich als Marke auf der Stelle p1 des *Systemnetzes* befindet. Die Anschrift on:channel(42) an der Transition t im Systemnetz ist der Downlink zu dem korrespondierenden Uplink :channel(x) an der Transition e im Objektnetz (vgl. Kapitel 4.4.2).

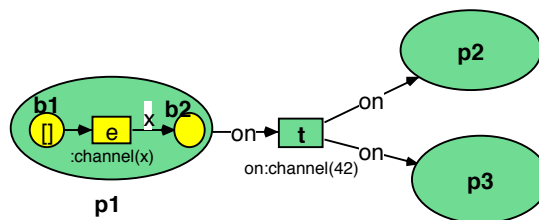


Abbildung 4.2.: Ein Objektnetz als Marke in einem Systemnetz (aus [KMRV05], S. 56).

Unterschiedliche Semantiken ([Val00c], [Val00b]) für verteiltes Bearbeiten beschreiben grundsätzliche Situationen: a) in der *Wertsemantik* können Kopien oder Repräsentanten desselben Objektes unabhängig und nebenläufig arbeiten, während b) in der *Referenzseman-*

4. Modellierungstechnik

tik Kopien nicht möglich sind, d.h. bei verteilter Bearbeitung auf ein Objekt referenziert wird.

Objektnetze mit der zuletzt genannten Semantik wurden in [Kum02] als *Referenznetze* intensiv untersucht. Referenznetze sind objektorientierte Petrinetze höherer Ordnung, welche das Konzept der “Netze in Netzen” nach Valk ([Val98], [Val00a]) weiterentwickeln und mit *synchronen Kanälen* nach Christensen [CH94] sowie der Anschriftssprache Java kombinieren. Für Referenznetze steht eine integrierte Entwicklungsumgebung RENEW (*Reference Net Workshop*, [KWD⁺04]) zur Verfügung, in der Netze grafisch dargestellt und simuliert (d.h. ausgeführt) werden können.

Für Referenznetze wurden zahlreiche theoretische Ergebnisse [Kum02] und Anwendungen in der Modellierung erarbeitet (z.B. [vdAMVW99]). Alle hier aufgeführten Petrinetzformalismen erreichen eine Mächtigkeit, die der von Turingmaschinen gleichwertig ist. Dies wurde in [Wie01] und [Ste02] gezeigt, indem die jeweiligen Formalismen dazu benutzt wurden, sich selbst zu modellieren. Weitere Details zu den benutzten Modellierungsfomalismen und deren Eigenschaften sind in [vLMV03] zu finden.

4.1.3. Agentennetze

Unter anderem werden Petrinetze bzw. Referenznetze zur Darstellung von Agentenarchitekturen eingesetzt. Erste Arbeiten zur Modellierung von Agenten durch Petrinetze gehen auf [MW97] zurück und wurden in Hinsicht auf ihre Strukturen und formalen Eigenschaften untersucht ([Röl99]; [KMR01a]). Auf der Basis von Objekt-Petrinetzen, und zwar speziell des RENEW-Werkzeugs, wurde eine konzeptionelle und technische Plattform CAPA (siehe Kapitel 4.1.4) zur Darstellung, Implementation und Simulation von verteilten Agenten entwickelt, die zudem den FIPA-Spezifizierungen genügt (vgl. [KR01] für die konzeptuelle und [DMR03] für die technische Grundlage).

Der Formalismus der Referenznetze hat mehrere Vorteile gegenüber anderen Ansätzen zur Agentenmodellierung wie z.B. AUML ([BOvD00]):

- **Operationale Semantik:** Referenznetze besitzen eine eindeutige Semantik. Anders als bei UML und den daraus abgeleiteten Techniken (wie AUML) kann ein durch ein Referenznetz dargestelltes Modell in einer eindeutigen Weise ausgeführt werden. Darüber hinaus gibt es zahlreiche anschlussfähige Arbeiten zur Verifikation von Petrinetzmodellen ([MV93], [Val93]).
- **Implementation durch Spezifikation:** Aus demselben Grund stellt ein fertiges Modell gleichzeitig die Implementation des spezifizierten Gegenstandes dar; die Übertragung in eine gesonderte Form (z.B. eine Programmiersprache) entfällt.
- **Einheitliche Form:** UML besitzt eine verwirrende Vielfalt an Darstellungstechniken und -formen, die teilweise äquivalente oder zumindest sich stark überschneidende Bedeutungen haben. Diese “babylonische” Vielfalt entfällt bei der Benutzung von Petrinetzen, die für viele Anwendungsbereiche den kleinstmöglichen Umfang an Darstellungsmitteln besitzen.

- **Einheitliche Darstellung:** Ein wichtiges Ziel der Werkzeugunterstützung ist die einheitliche Darstellung von ausgewählten Sichten auf die modellierten Sachverhalte.
- **Ausdrucksmächtigkeit:** Die Mächtigkeit der Referenznetze ist im Hinblick auf die für MAS wichtigen Konzepte höher und erlaubt deren unmittelbare und intuitive Darstellung.

4.1.4. CAPA

Die FIPA-kompatible Agentenplattform CAPA ([Duv02], [DMR03]) wurde in der Arbeitsgruppe Theoretische Grundlagen der Informatik (TGI) an der Universität Hamburg von Michael Duvigneau entwickelt. CAPA ist eine Agentenplattform und gleichzeitig ein Rahmenwerk zur Spezifikation und Ausführung von Agenten auf Basis von Referenznetzen.

Als FIPA-konforme Agentenplattform bietet CAPA zur Laufzeit einen Nachrichtentransportdienst nach FIPA-Standard, einen AMS-Agenten nach FIPA zur Erzeugung und Verwaltung von Agenten, einen DF-Agenten nach FIPA als Verzeichnisdienst und einen Viewer (MULANVIEWER) zur Anzeige und Steuerung laufender Agenten (vgl. Kapitel 3.4.3).

Als Rahmenwerk zur Modellierung und Ausführung von Agenten bietet CAPA einen erweiterbaren Standard-Agenten, der eine Wissensbasis mit Schlüssel-Wert-Paaren zur Verfügung stellt (die Werte sind Java-Objekte). Um mit CAPA eine Anwendung zu implementieren, sind für jeden Agenten eine initiale Wissensbasis, Protokollnetze sowie ggf. Entscheidungskomponenten und Java-Hilfsklassen zu definieren.

Die Protokollnetze sind Referenznetze, die mittels synchroner Kanäle die Schnittstellen für den Zugriff auf die Wissensbasis und für das Senden und Empfangen von Nachrichten nutzen (siehe auch Kapitel 4.3). Durch Protokollnetze wird das Verhalten des Agenten spezifiziert. Für die Modellierung von Protokollnetzen steht in RENEW ein Netzkomponenten-Plugin zur Verfügung, mit dem gleichartige, immer wiederkehrende Komponenten eines Protokollnetzes, wie z.B. das Versenden und Empfangen von Nachrichten, Kontrollsequenzen wie “if“, “cond“ etc. einfach eingebaut werden können.

In der Wissensbasis wird festgelegt, wie der Agent auf die ankommenden Nachrichten reagiert und welche Aktionen er proaktiv, also ohne einen Anstoß durch eine Nachricht, ausführt. Ferner kann der Agent anwendungsbezogene Informationen in seiner Wissensbasis speichern. Die initiale Wissensbasis kann entweder in einem einfach editierbaren Datei-Format angegeben werden oder mit einem grafischen Editor (KnowledgeBase Editor, “KBE“) für alle Agenten einer Anwendung spezifiziert werden.

Die oben erwähnten Hilfsklassen in Java können beliebige anwendungsspezifische Funktionen erfüllen, insbesondere steht die Ontologie, also das Begriffsgerüst der Anwendung in Form von Java-Klassen als Speicher- und Nachrichtenformat zur Verfügung.

4.2. Eine petrinetzbasierte Multiagentenarchitektur: Mulan

Die Multiagentenarchitektur MULAN (Multi-Agenten-Netze) ([Röl04]) ist eine allgemeine Architektur eines MAS, die vollständig in Petrinetzen implementiert ist. Dabei finden Referenznetze ([Kum02]) Anwendung. Sie erlauben neben den generellen Vorteilen grafischer Entwicklungsumgebungen die Umsetzung des “Netze in Netzen“-Gedankens in die Praxis, der sich besonders gut für die Modellierung und Darstellung hierarchischer (auch rekursiver) nebenläufiger Systeme eignet: Ein MAS gibt eine Struktur vor, in der sich innerhalb einer gewissen Umgebung Plattformen und Kommunikationsstrukturen befinden. Plattformen enthalten Agenten und Strukturen zu deren Unterstützung. Agenten wiederum enthalten Verhaltensprotokolle, die selbst wieder geschachtelt sein können. Referenznetze können jede dieser Ebenen mit allen Beziehungen zu darüber und darunter liegenden Ebenen intuitiv darstellen, was eine einheitliche Betrachtung des Gesamtsystems von der Makro- bis zur Mikroebene gestattet.

Bei der Erstellung von MULAN wurden FIPA-Standards (vgl. Kapitel 3.5) berücksichtigt. Diese schreiben ein einheitliches Nachrichtenformat (an KQML [Lab96] angelehnt) und bestimmte Minimaldienstleistungen vor, reglementieren aber den internen Aufbau der Agenten kaum. Viele Elemente des Multiagentensystems, wie die Plattform und die generelle Struktur eines Agenten, können für beliebige Agenten (wieder-) verwendet werden, MULAN ist also gleichzeitig ein Rahmenwerk zur Entwicklung von Agenten. Es ermöglicht eine einheitliche Modellierung des Agentenverhaltens auf verschiedenen Abstraktionsebenen sowie den Vergleich verschiedener Agentenarchitekturen.

Die petrinetzbasierte MULAN-Architektur definiert daher auf einer konzeptionellen Ebene die Strukturen eines Multiagentensystems.¹ Abb. 4.3 zeigt die vier Hierarchiestufen eines Multiagentensystems in MULAN: *Agentensystem*, *Agentenplattform*, *Agent* und *Protokoll*.²

Links oben ist die Architektur des *Agentensystems* dargestellt. Es besteht aus Stellen, die Plattformnetze als Marken beinhalten, und Transitionen, die die Kommunikationskanäle und damit die Infrastruktur des Agentensystems beschreiben. Diese Systeminfrastruktur muss für jedes konkrete System neu modelliert werden.

Der ZOOM-Pfeil zeigt die vergrößerte Darstellung einer Marke, also der *Plattform*, an. Die Plattform enthält Agenten(netze), die auf der zentralen Stelle **agents** liegen. Die Transitionen stellen Dienste der Plattform dar: Agenten können von der Plattform erzeugt (Transition **new**) und wieder gelöscht (Transition **destroy**) werden. Die Plattform ermöglicht die Migration der Agenten sowie ihre Kommunikation sowohl mit den Agenten auf der Plattform (Transition **internal communication**) als auch mit externen Agenten (Transition **external communication**).

Eine weitere Vergrößerung, diesmal der Agentenmarke, zeigt das *Agentennetz*. Ein Agent kommuniziert mit seiner Umwelt mittels Nachrichten, die er als Input von außen bekommt und verarbeitet oder ggf. verschickt. Er kann so über die Plattform mit anderen Agenten kommunizieren. Die Transitionen **incoming** und **outgoing** sind die einzigen Schnittstellen des

¹Die folgende Beschreibung basiert im Kern auf [KMRV05], S. 6f.); für die ausführliche Darstellung siehe [Röl04].

²In der Abbildung wurde die MAS-Architektur in Mulan vereinfacht dargestellt, z.B. wurden einige Aufschriften sowie alle synchronen Kanäle weggelassen.

4.2. Eine petrinetzbasierte Multiagentenarchitektur: Mulan

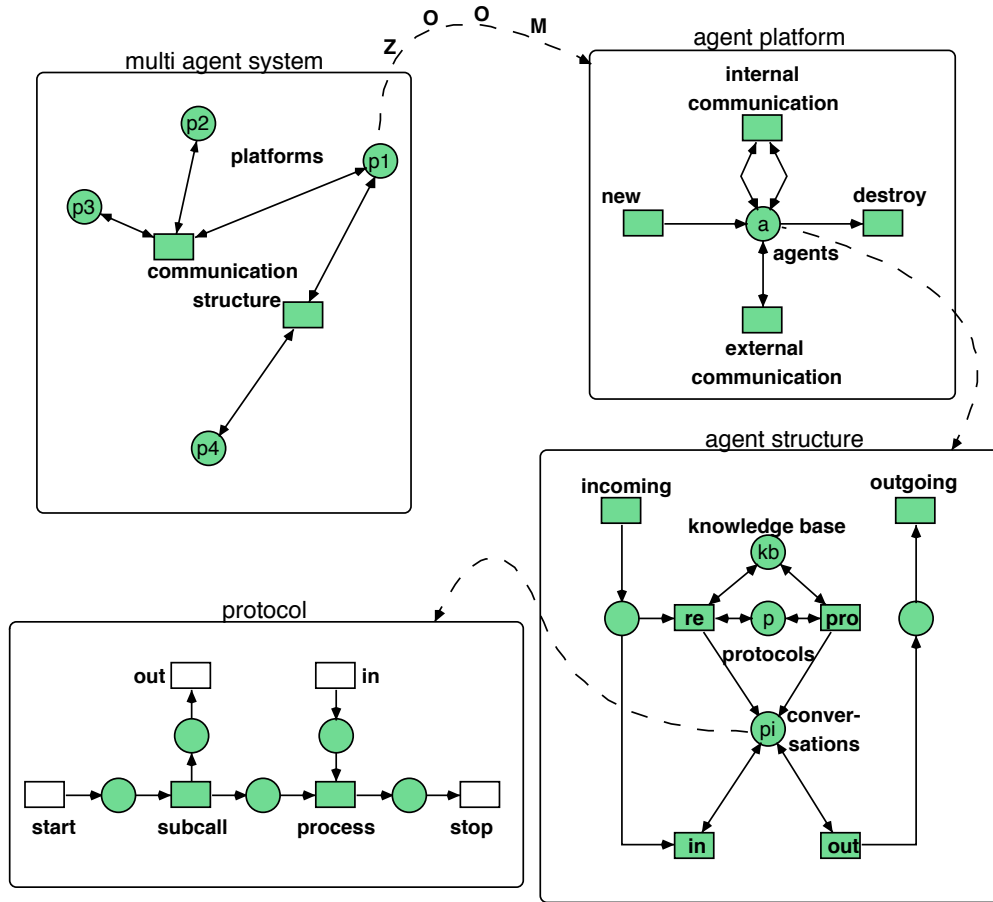


Abbildung 4.3.: Multiagentensystem in MULAN. Darstellung nach [KMRV05], S. 6, modifiziert nach ([Röl04], S. 143).

Agenten zum Rest des Agentensystems, alle anderen Aktivitäten sind in seinem Inneren gekapselt. Die Verhaltensweisen eines Agenten werden durch *Protokolle* modelliert, die als Marken an der Stelle **protocols** vorliegen. Die Wahl der Protokolle, also der auszuführenden Verhaltensweisen, ist die zentrale Aktivität eines Agenten. Die Protokolle können proaktiv (Transition **pro**) oder reaktiv (Transition **re**) ausgewählt werden. Mit dem Schalten der Transition **pro** kann ein Protokoll ohne Input von außen gestartet werden, z.B. als Folge agenteninterner Deliberationsprozesse. Die Transition **re** schaltet nur als Reaktion auf eine von außen ankommende Nachricht.

Ein ausgewähltes und aktiviertes Protokoll wird *Konversation* genannt, da es meist den Nachrichtenaustausch mit anderen Agenten (oder den internen Austausch) impliziert. Eine

4. Modellierungstechnik

neu erzeugte Konversation enthält ein eindeutiges Identifikationslabel, das in der Abbildung zwecks besserer Lesbarkeit weggelassen wurde. *Konversationen* können Nachrichten entgegennehmen (Transition *in*), Nachrichten versenden (Transition *out*) und auf die Wissensbasis zugreifen. Nach der Abarbeitung werden die Protokolle, d.h. die abgeschlossenen Verhaltensweisen, von der Stelle *conversations* abgezogen.

Protokolle sind spezielle Referenznetze, deren Schnittstelle eindeutig durch das Agentennetz vorgegeben ist. Ein Protokoll (rechts unten) steuert den Agenten von innen und kommuniziert mit anderen Agenten über Transitionen, die der Agent dem Protokoll als Dienst anbietet: Nachrichteneingang und -ausgang (Transitionen *in* und *out*). Protokolle implementieren das Agentenverhalten und können sowohl einfache serielle Abläufe als auch komplexe dynamische Workflows darstellen.

In [Ree10] werden die minimal notwendigen Elemente für die Implementation einer Anwendung mit MULAN / CAPA beschrieben (vgl. [Ree10], S. 85). Diese beinhalten u. a.

- **Wissensbasen:** Für jeden Agenten wird eine initiale Wissensbasis in Form einer Konfigurationsdatei (wis-Format oder xml-Format) angegeben. Darin werden die Dienstbeschreibungen, das reaktive und proaktive Verhalten und initiales Wissen hinterlegt.
- **Protokollnetze:** Das Verhalten des Agenten und die Verwendung der Wissensbasis werden durch eine Reihe von Protokollnetzen in Form von Referenznetzen implementiert.
- **Entscheidungskomponenten:** Entscheidungskomponenten (*Decision Component, DC*) implementieren, wie auch die Protokollnetze, anwendungsspezifisches Verhalten, jedoch geben sie den ansonsten statischen und Workflow-artigen Protokollnetzen zusätzliche Flexibilität, indem sie bestimmte Komponenten / Kontrollstrukturen von diesen übernehmen und von Protokollnetzinstanzen bezüglich der auszuwählenden Entscheidungen / Verhalten angefragt werden. Entscheidungskomponenten können auch proaktives Verhalten auslösen, indem sie die Instanziierung eines Protokollnetzes erzwingen. Auf diese Weise kann eine Planungskomponente im Sinne der KI als Entscheidungskomponente in einen Agenten eingebunden werden.
- **Ontologie:** Die Ontologie liegt in Form von Java-Klassen vor. Die Java-Klassen bieten Hilfsmethoden zum Umgang mit Nachrichten. Die Ontologie ist im Agentennetz nicht abgebildet. Java-Klassen können für die Implementierung der initialen Wissensbasis, für die Protokollnetze und für die Entscheidungskomponenten verwendet werden.
- **Anwendungsstart:** Beim Start einer Anwendung werden der Java-Klassenpfad und der RENEW-Netzpfad sowie diverse Parameter für CAPA konfiguriert. Der Anwendungsstart kann in einem Startskript (als Shellskript), in einem Referenznetz, in Form einer speziellen Multiagentensystem-Konfiguration oder über Menüeinträge von CAPA in der RENEW-GUI gestartet werden.

4.3. PAOSE-Ansatz

Die Petrinetz, Agenten und Organisations-orientierte Software-Entwicklung (PAOSE) ist ein systematischer Ansatz, der von der Arbeitsgruppe Theoretische Grundlagen der Informatik (TGI) an der Universität Hamburg für die Konstruktion von Multiagentenanwendungen entwickelt wurde. Im Kontext dieses Ansatzes wurde eine Reihe von Konzepten, Methoden und Modellierungstechniken vorgestellt, die die systematische Konstruktion von Multiagentenanwendungen, auf der Grundlage des MULAN / CAPA-Rahmenwerks ermöglichen (siehe z.B. [Cab10], [Mol05], [Mol06]).

Die wichtigen Werkzeuge für PAOSE sind RENEW mit diversen Plugins zur Unterstützung grafischer Modellierung und CAPA als Rahmenwerk für die Agentenimplementierung. Für diese Arbeit wichtige PAOSE-Modelldiagramme sind die UML-Use-Case-Diagramme sowie Agenteninteraktionsprotokolle (AIP), die zur Darstellung agentenübergreifender Prozesse verwendet werden. Das Vorgehen, um von ermittelten Anforderungen zu einem laufenden System zu gelangen, wird in PAOSE in die iterierten Phasen Grobentwurf, Umsetzung in drei parallelen Schritten (Verhalten, Agenten und Ontologie) und Integration gegliedert (vgl. [Ree10]). Die Einzelheiten zu den Modelldiagrammen und zum grundsätzlichen Vorgehen werden in Kapitel 9 vorgestellt, wo der PAOSE-Modellierungsansatz für die Entwicklung der Multiagentenanwendungen auf Basis eines Referenzmodells für die Emotionsmodellierung E-MULAN verwendet wird.

4.4. Petrinetze als sozionische Darstellungstechnik

In [vLMV03] wurde gezeigt, wie organisationssoziologische und sozialtheoretische Konzepte mit Hilfe von Petrinetzen modelliert und reformuliert werden können. Diese Vorgehensweise ist eine Möglichkeit der Präsentation soziologischer Theorien in informatischer Darstellungssprache. Der sozionische Modellierungsansatz ist für die vorliegende Arbeit von grundlegender Bedeutung, da er als ein etabliertes methodisches Werkzeug zur Darstellung der sozialwissenschaftlichen Theorien (d. h. Theorien, die aus Soziologie, Psychologie, Neurowissenschaften usw. kommend in die Informatik Eingang finden - vgl. 1.5 zur Terminologie) mit informatischer Modellierungstechnik die hier verwendete Art der Theoriendarstellung begründet.

4.4.1. Sozionische Theorienmodellierung mit Petrinetzen

Im Bereich der Theorienmodellierung hat die Evaluierung der speziellen Petrinetz-Formalismen einige ihrer wichtigen Vorteile bei der Modellierung soziologischer Theorien gezeigt. In [vLMV03] wird insbesondere die schwache Ausdrucksmächtigkeit bisheriger Modellierungsansätze, die unpräzise Semantik, die nicht vorhandene Werkzeugunterstützung und die fehlende Berücksichtigung der Änderung von Strukturen, die üblicherweise als statisch und damit unveränderbar angesehen werden, kritisiert ([vLMV03], S. 4). Mit Petrinetzen steht eine exakte formale Modellierungstechnik zur Verfügung, die eine adäquate Modellierung von soziologischen Konzepten und Strukturen ermöglicht. In [vLMV03] wurden Petrinetze

4. Modellierungstechnik

als Denkwerkzeug, als vermittelnde Darstellungstechnik und als Basis für die operationale Semantik der Agentensysteme eingesetzt.

Allgemeine Vorteile der Petrinetzmodellierung

Die Modellierung mit Petrinetzen kann grundsätzlich sowohl als Beitrag zur konzeptionellen Begriffsbildung als auch für konkrete Computersimulationen genutzt werden und weist die potenzielle Implementierbarkeit eines Modells nach. In [vLMV03] werden an einem konkreten Beispiel (Garbage Can-Modell [CMO72]) insbesondere folgende Vorteile der Modellierung mit Petrinetzen illustriert:

- die Modellierung ist hierarchisch,
- die modellierten Objekte sind direkt identifizierbar,
- es werden verschiedenen Abstraktionsebenen sowohl bei der Modellierung als auch bei der Simulation angeboten,
- das Modell ist modifizierbar (es können neue Modellkomponenten hinzugefügt, Komponenten entfernt und verändert werden),
- es erlaubt dynamische (Struktur-)Änderungen,
- Nebenläufigkeit und Nichtdeterminismus sind direkt darstellbar und
- Modellierung von Verhalten über die Zeit und Darstellung von Ressourcen wird unterstützt.

Eine weitere, im Hinblick auf Theorienmodellierung besonders wichtige Eigenschaft ist die operationale Semantik der Petrinetze: Die Petrinetzmodelle beschreiben sowohl die statische Sicht auf ein Modell als auch die resultierende Dynamik.

Das Baukastenprinzip

Nach [vLMV03] besteht der Ertrag der sozionischen Modellierung von sozialwissenschaftlichen Theorien mit Petrinetzen einerseits darin, dass die zentralen Aspekte einer Theorie in Petrinetzmodellen dargestellt werden, die prinzipiell als Programmierbausteine in ein ausführbares Multiagentensystem integriert werden können. Andererseits ermöglicht diese Art der Modellierung einen direkten Vergleich und ggf. die Integration der modellierten theoretischen Ansätze, da die Modellierung mit Petrinetzen dem *Baukastenprinzip* folgt und dadurch die Kombination und Integration von Modellen höchst unterschiedlichen theoretischen Ursprungs und Abstraktionsgrades unterstützt. Wie in Kapitel 4.5 diskutiert wird, ist gerade vor dem Hintergrund der Interdisziplinarität und Fragmentierung der Emotionserforschung die Möglichkeit der Integration verschiedener emotionstheoretischer Ansätze zu einem allgemeinen Modell besonders interessant.

Bei der Modellierung der sozialwissenschaftlichen Theorien ist es eine besondere Herausforderung, den theoretischen Aussagen in ihrer Komplexität treu zu bleiben. Die wichtige

4.4. Petrinetze als sozionische Darstellungstechnik

Aufgabe besteht darin, einen adäquaten Abstraktionsgrad zu finden: Die entstehenden Modelle sollten einerseits abstrakt genug sein, um die übersichtliche Darstellung des Modellgegenstandes ohne übermäßige Details zu liefern. Andererseits sollte nicht über konzeptuell wichtige Sachverhalte hinwegformalisiert werden, um nicht dünne und wenig aussagekräftige Modelle zu erhalten. Das Baukastenprinzip der Petrinetzmodellierung ermöglicht es, dieses Modellierungsdilemma zu umgehen: Petrinetze als Darstellungstechnik sind in der Lage, sowohl Komplexität zu erfassen, als auch abstrakte Überblicksmodelle zu erzeugen. Komplexe Zusammenhänge können bis auf die Ebene ihrer Bestandteile heruntergebrochen und in einer Menge sehr einfacher Netze formuliert / dargestellt werden. Diese Netze können jedoch jederzeit rekombiniert und zu komplexen Überblicksnetzen zusammengefasst werden. Die Überblicksnetze kann man wiederum vereinfacht darstellen, indem man von im gegebenen Zusammenhang unrelevanten Theorie- bzw. Modellteilen abstrahiert. Damit ist ein beliebiger Detailliertheits- bzw. Abstraktionsgrad der Modelle erreichbar, wobei die verschiedenen Darstellungen gleichzeitig Verwendung finden können.

Vorgehensweise bei der sozionischen Modellerstellung

Bei der Modellierung der Theoriebausteine wird schrittweise vorgegangen: Ausgehend von Einzelaussagen über Abhängigkeiten und Kausalitäten werden mehrstufige Zusammenhänge aufgedeckt und Wechselwirkungsketten identifiziert. Die abstrakten Petrinetzmodelle, die zuerst nur einfache Abhängigkeitsketten darstellen, werden schrittweise verfeinert und weiterentwickelt. So wird im ersten Schritt eine textuelle Darstellung der Theorie in eine oder mehrere Netze umgewandelt, die die in der Theorie spezifizierten Elemente und ihre kausalen Beziehungen beschreiben. Durch die Analyse dieser Netze wird in einem zweiten Schritt versucht, eine Modellstruktur zu erkennen, die durch die Wechselwirkungen der Modellelemente konstituiert wird und die Dynamik dieser Wechselwirkungen wiedergibt. Wie in ([vLMV03], S. 101) festgestellt wurde, macht erst die Identifizierung eines vollständigen und in sich geschlossenen Wechselwirkungs- bzw. Reproduktionskreislaufs das Modell operational: Aktionen von Akteuren führen beispielsweise zur Etablierung von Gesellschaftsstrukturen, die nachfolgende Aktionen beeinflussen (können).

Ertrag sozionischer Theoriemodellierung

Bei der Analyse der durch Petrinetz-Modellierung gewonnenen Prozess- und Strukturdarstellungen der Systemabläufe können oft (wechselseitige) Referenzen bzw. Kausalitäten zwischen Theorieteilern festgestellt werden, die aus der textuellen Beschreibung nicht ohne Weiteres ableitbar wären. So werden vielerorts abstraktere Strukturdarstellungen möglich bzw. Teilmodelle können zu einem allgemeineren Modell zusammengefasst werden.

Da zur Darstellung mittels informatischer Modellierungstechniken präzise Repräsentationen notwendig sind, können bei der Modellierung von Theorieaussagen ggf. unspezifizierte Stellen in der textuellen Darstellung aufgedeckt werden. So sind insbesondere "blinde Flecken" der Theorie bzw. Inkonsistenzen bei der Beschreibung der Wechselwirkungen der Strukturelemente aus der textuellen Theoriedarstellung besonders schwer herauszuarbeiten.

4. Modellierungstechnik

Neben den internen Referenzen können anhand der Modelle auch externe Referenzen auf andere Theorien hergestellt werden, denn die Modelle erleichtern den Vergleich der fundamentalen Strukturen und Abläufe, die in einer sozialwissenschaftlichen Theorie thematisiert werden, so dass strukturelle Analogien und mögliche Anknüpfungspunkte deutlicher erkannt werden.

Dies kann einerseits durch Analyse der statischen Modellstrukturen geschehen, andererseits und insbesondere jedoch bei der Modellsimulation. Auf diese Weise können aus der modellierten Theorie sogar implizite Aussagen abgeleitet werden, die in der textuellen Theoriedarstellung nicht thematisiert wurden.

4.4.2. Vermittelnde Netze: Grundlagen und allgemeine Modellierungsprinzipien

Im Folgenden werden grundlegende Elemente von vermittelnden Netzen nach [vLMV03] beschrieben und die allgemeine Vorgehensweise bei der sozionischen Theorienmodellierung vorgestellt.

Netzelemente

Für die Darstellung der soziologischen Theorien in einer textnahen Form verwenden [vLMV03] in den vermittelnden Netzen folgende Elemente:

- **Prozesshafte Objekte der Beschreibung.** Dies sind zustandsbehaftete Einheiten (Akteure, Strukturen usw.). Sie werden durch einen Platz mit Doppelkanten verbunden dargestellt (vgl. Abb. 4.4).
- **Bedingungen.** Damit sind in [vLMV03] sowohl Ressourcen als auch allgemeine Konzepte wie beispielsweise Emotionen gemeint (vgl. [vLMV03], S. 383). Ressourcen werden durch das Schalten der Transition “verbraucht“, was mit einfachen Eingangskanten notiert wird (vgl. Abb. 4.4).
- **Ergebnisse.** Ebenso wie bei Bedingungen, werden damit Ressourcen und Konzepte bezeichnet, die durch das Schalten der Transition “erzeugt“ werden (vgl. Abb. 4.4).
- **Nebenbedingungen.** Damit werden Objekte modelliert, die sich selbst im Laufe der Aktion nicht verändern, d.h. nicht verbraucht werden, aber für die Durchführung der Aktivität unbedingt erforderlich sind. Sie werden durch Testkanten (s.u.) notiert.

Grafische Elemente

Neben den zwei klassischen Arten von Kanten, den *Eingangs- und Ausgangskanten*, die Marken von einem Platz entfernen bzw. hinzufügen, kennt der Referenznetzformalismus nach [Kum02] weitere Kantenarten: Reservierungskanten, Testkanten und flexible Kanten.

4.4. Petrinetze als sozionische Darstellungstechnik

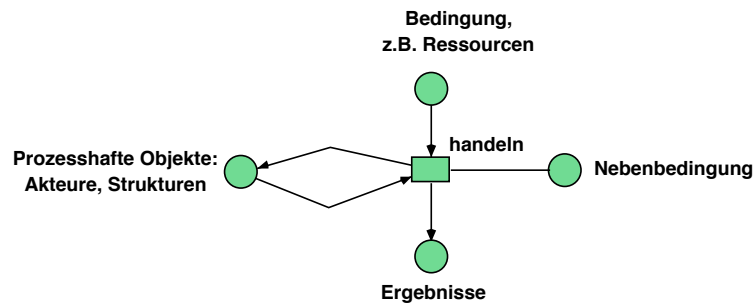


Abbildung 4.4.: Netzelemente von vermittelnden Netzen.

- **Reservierungskanten** stellen eine abkürzende Notation für eine Ein- und eine Ausgabekante von einem Platz zu einer Transition dar, die beide dieselbe Beschriftung tragen. Sie reservieren eine Marke für die Zeit des Feuerns einer Transition (vgl. Abb. 4.5, links). Eine Reservierungskante zieht beim Schalten der Transition eine Marke ab und legt sie gleich wieder zurück.
- **Testkanten** prüfen, ob eine Marke auf einem Platz liegt. Eine einzelne Marke kann von mehreren Testkanten gleichzeitig auf ihre Existenz (oder ihren Wert) getestet werden (vgl. Abb. 4.5 Mitte).
- **Flexible Kanten** sind in der Lage, eine dynamisch festgelegte Anzahl von Marken gleichzeitig zu transportieren (vgl. Abb. 4.5, rechts).

Abb. 4.5 zeigt die drei Kantenvarianten in der Darstellungsweise des Petrinetzwerkzeugs RENEW.

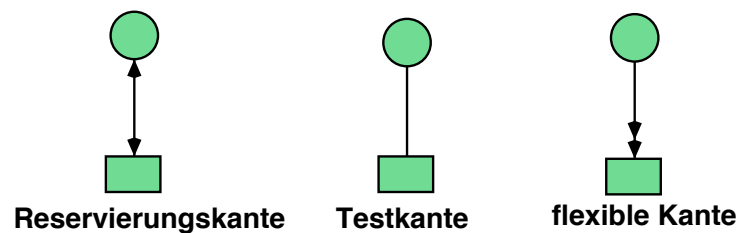


Abbildung 4.5.: Kantenarten in Referenznetzen.

In der Literatur werden noch weitere Kantenarten beschrieben: So werden *Inhibitor-Kanten* (*inhibitorische Kanten*) dazu genutzt, um eine Transition zu deaktivieren, wenn sich Objekte auf der verbundenen Stelle befinden. Das Schalten einer Transition hat keinen Einfluss auf die verbundene Stelle.

Um die Übersichtlichkeit innerhalb eines Netzes durch Vermeidung unnötiger Kanten zu verbessern, kann eine Kopie eines Platzes angelegt werden, die mit vielen Transitionen

4. Modellierungstechnik

verbunden ist. Diese sogenannten *virtuellen Plätze* tragen zu jedem Zeitpunkt dieselbe Markierung wie der ursprüngliche Platz, sind also mit Fusionsplätzen [Jen92] zu vergleichen. Die virtuellen Plätze sind jedoch nur innerhalb einer Netzinstanz gültig.

Netzinstanzen und synchrone Kanäle

Netzinstanzen stellen Ausprägungen eines Netzes dar, ähnlich wie Objekte einer objektorientierten Sprache Instanzen einer Klasse sind. Instanzen desselben Netzes können unterschiedliche Zustände einnehmen und sind auch sonst voneinander unabhängig. *Synchrone Kanäle* ermöglichen die Synchronisation (z.B. von zwei Netzinstanzen) durch das zeitgleiche Feuern von zwei oder mehr Transitionen mit der Möglichkeit des Datenaustausches. Die Synchronisation erfolgt über den Namen des Kanals und die Übergabeparameter. Eine der zu synchronisierenden Transitionen stellt dem Kanalaufruf den Namen der aufzurufenden Netzinstanz voran. Innerhalb dieser Netzinstanz wird daraufhin nach einer Transition gesucht, die einen mit dem Namen und den Parametern übereinstimmenden Kanalaufruf trägt. Durch Verwendung des Schlüsselwortes “this” kann die Synchronisation auf die aufrufende Netzinstanz beschränkt werden. So lassen sich synchrone Kanäle auch innerhalb einer Netzinstanz verwenden. Grundsätzlich ist jedoch die Synchronisation von beliebigen Netzinstanzen möglich. Abb. 4.6 illustriert die Verwendung eines Kanals innerhalb einer Netzinstanz.

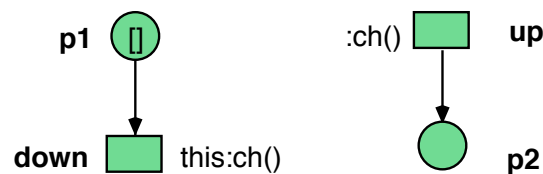


Abbildung 4.6.: Synchroner Kanal innerhalb eines Netzes.

Falls ein Referenznetz den reservierten Kanal “new“ aufweist, wird dieser für die Erzeugung einer neuen Instanz dieses Netzes aufgerufen. Die Parameter dieses Kanals können von der neuen Netzinstanz zur Initialisierung genutzt werden. Zur Darstellung der genauen Syntax bei der Benutzung von synchronen Kanälen siehe z.B. [Kum02], [Duv03].

Hierarchisierung

Durch Kanäle verbundene Netzinstanzen bilden nicht notwendigerweise eine Netzhierarchie (vgl. [Jen92]). Eine streng hierarchische Beziehung zwischen Instanzen ergibt sich aus der Modellierung einer Verbindung zweier Netze, bei der die Identität einer Netzinstanz der anderen unbekannt ist. Eine saubere hierarchische Beziehung zwischen einem *Systemnetz* (Netzinstanz mit Downlink) und einem *Objektnetz* (Netzinstanz mit Uplink) ergibt sich, wenn das Objektnetz aus Sicht des Systemnetzes eine Transitionsverfeinerung darstellt (vgl. Abb. 4.2).

Für die sozionische Theorienmodellierung stellt Hierarchisierung eine wichtige Darstellungstechnik dar. Durch das Netze-in-Netzen-Paradigma können die Strukturen der niedrigeren Hierarchieebenen als Marken in Systemnetzen abgebildet werden, welche die Strukturen der höheren Hierarchieebene repräsentieren. Auf diese Weise lassen sich die in den sozialwissenschaftlichen Theorien beschriebenen hierarchischen Beziehungen bis zu einer beliebigen Tiefe modellieren.

Abstraktion

Auf der allgemeinen Ebene werden in [vLMV03] zwei Abstraktionsarten unterschieden: Zum einen die *Verfeinerung* bzw. *Vergrößerung*, bei der Sachverhalte auf das Gemeinsame reduziert werden; zum anderen die *Faltung*, bei der gleichartige Prozesse bzw. Strukturen zusammengefasst werden.

Netztechnisch wird als Abstraktion bei einem stellenberandetem Netz eine Stelle und bei einem transitionsberandetem Netz eine Transition betrachtet.

Verfeinerung / Vergrößerung

Die *Vergrößerung* ist eine lokale Abstraktion. Es werden mehrere Knoten zu einer Stelle oder Transition zusammengefasst, die diese Knotenmenge ersetzt. Dabei darf eine Menge von Knoten nur zusammengefasst werden, falls alle Knoten am Rand vom gleichen Typ sind. Und genau von diesem Typ muss dann auch der Knoten sein, der diese Menge ersetzt.

Die *Verfeinerung* wird benutzt, um die innere Struktur einer Transition oder Stelle genauer zu konkretisieren. Ob und wann man eine Verfeinerung bzw. eine Vergrößerung einsetzt, hängt jeweils vom Modellierungsgegenstand und dem gewünschten Abstraktionsgrad ab. Grundsätzlich hat sich bei der sozionischen Modellierung mit vermittelnden Netzen die Vorgehensweise bewährt, bei der die zu modellierende Theorie am Anfang möglichst grob modelliert und später Schritt für Schritt verfeinert wird, bis das Gesamtmodell alle wichtigen Aspekte der Theorie beinhaltet.

In Abb. 4.7 wird auf der linken Seite ein Beispiel der stellenberandeten Abstraktion bzw. die Verfeinerung einer Stelle illustriert. Diese Abstraktion verändert offensichtlich das Verhalten des Netzes. Auf der rechten Seite ist ein Beispiel für die Vergrößerung eines transitionsberandeten Teilnetzes beziehungsweise die Verfeinerung einer Transition dargestellt.

Faltung

Bei einer als *Faltung* bezeichneter Abstraktionsart kann ein Netzgraph, der aus gleich strukturierten Teilnetzen besteht, dahingehend vereinfacht werden, dass man die gleichen Teilnetze einfach "zusammenfasst". Von dieser Abstraktion wird beim Entwurf des Referenzmodells E-MULANim Abschnitt 8.2 Gebrauch gemacht. Es dürfen im Rahmen der Faltung nur Knoten gleichen Typs übereinandergelegt werden. Des Weiteren müssen alle Relationsbeziehungen erhalten bleiben und die Teilnetze isomorph sein. Jedoch müssen die Teilnetze hier nicht wie bei der Vergrößerung am Netzrand Knoten gleichen Typs aufweisen.

4. Modellierungstechnik

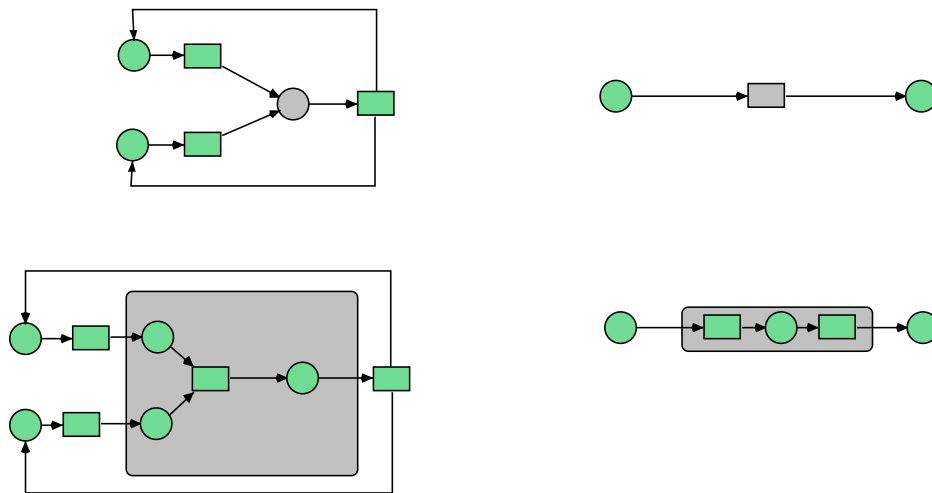


Abbildung 4.7.: Verfeinerung / Vergröberung.

Entsprechend wird das Verdoppeln von Teilnetzen als *Entfaltung* bezeichnet, wobei zwei gleich strukturierte Teilnetze entstehen. In dem Beispiel in Abb. 4.8 werden die beiden Transitionen t1 und t2 zu einer Transition t12 und die zwei Stellen s2 und s3 zu einer Stelle s23 zusammengelegt.

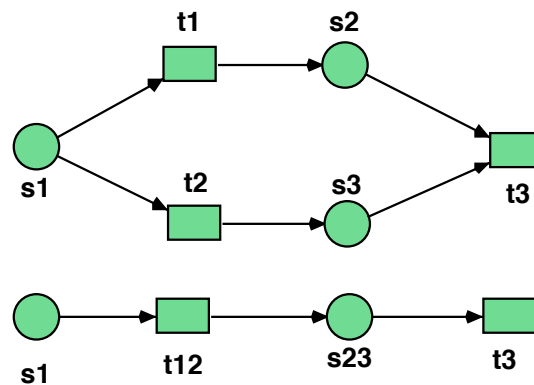


Abbildung 4.8.: Faltung und Entfaltung.

Zusammenfassung und Bewertung Die Vorteile der Petrinetze als Modellierungstechnik für sozialwissenschaftliche Zusammenhänge wurden bereits in [vLMV03] aufgeführt und diskutiert. Durch die operationale Semantik und das weite Spektrum der Analysemethoden erlauben Petrinetze die Evaluation und formale semantische Überprüfung der erstellten Modelle. Zusätzlich ermöglicht später die formale Repräsentation der Emotionsmodelle ihre leichtere Integration in die informatische Anwendungsdomäne. Die grafische Repräsentation

4.5. Anwendung des sozionischen Ansatzes bei der Modellierung von Emotionstheorien

hat sich insbesondere im interdisziplinären Kontext bewährt, wo sie eine gemeinsame Kommunikationssprache bildet, z.B. bei gemeinsamen Forschungsarbeiten der Informatik und der Soziologie im Rahmen von Sozionik-Projekten (vgl. [vLMV03], [KMRV05]). Im Unterschied zu anderen Automaten-Formalismen erlauben Petrinetze eine direkte Modellierung der Nebenläufigkeit. Petrinetze höherer Ordnung (Referenznetze, Objektnetze) ermöglichen die Repräsentation rekursiver Strukturen und emergenter Prozesse und sind in der Lage, sowohl strukturelle als auch prozessuale Modellperspektiven gleichzeitig darzustellen. Ferner existieren diverse Erweiterungen des Basisformalismus und aufwendige Werkzeuge für die meisten dieser Erweiterungen.

Auf der Erfahrung aus der sozionischen Modellierung von soziologischen Theorien aufbauend, schlagen wir die Anwendung der Petrinetze / Referenznetze für die Modellierung der Emotionstheorien aus Psychologie, Soziologie und Neurowissenschaft vor.

4.5. Anwendung des sozionischen Ansatzes bei der Modellierung von Emotionstheorien

In diesem Abschnitt wird die Anwendung des sozionischen Modellierungsansatzes für die Darstellung der Emotionstheorien aus verschiedenen wissenschaftlichen Disziplinen diskutiert.

Traditionell gehört die Untersuchung von Emotionen zu den klassischen Forschungsgebieten der Psychologie. Seit Mitte der 1970er Jahre ist sie zu einem expliziten Forschungsgegenstand der Soziologie geworden [Fla03] und seit Kurzem auch der Neurobiologie. Jede dieser Wissenschaften entwickelte mehrere Emotionstheorien, welche sich teilweise ergänzen, meistens jedoch parallel zueinander verschiedene Emotionsaspekte beleuchten bzw. dieselben emotionalen Phänomene aus verschiedenen Perspektiven beschreiben. Bedingt durch die unterschiedlichen Wissenschaftskulturen in den verschiedenen Disziplinen und evtl. auch durch sein überdurchschnittlich schnelles Wachstum in den letzten Jahrzehnten, ist das interdisziplinäre Forschungsfeld *Affective Science* sehr fragmentiert: Es gibt zahlreiche Emotionstheorien, -konzepte und -modelle, ohne dass die Integration der Grundlagen in eine allgemein akzeptierte Emotionstheorie erfolgt ist. Klaus R. Scherer spricht in diesem Zusammenhang von einem "regelrechten Wildwuchs von Theorievorschlägen" ([Vaa03]).

4.5.1. Motivation

Die Analyse der aktuellen Emotionstheorien in [ZS00] hat gezeigt, dass viele der gängigen Emotionstheorien lediglich als Teiltheorien anzusehen sind, da sie sich auf bestimmte Aspekte des Emotionsphänomens spezialisiert haben und nur innerhalb dieses Geltungsbereichs ihren Erklärungsanspruch sehen. "Eine umfassende theoretische Erklärung des Emotionsphänomens [...] erfordert jedoch *integrative Theorien*, die der Multikomponentalität des Phänomens Rechnung tragen und die insbesondere konkrete Vorhersagen über Auslösung und Differenzierung von Emotionsprozessen als auch über die hierbei auftretenden Reaktionsmuster in den verschiedenen [System-]Komponenten machen können." ([ZS00], S. 157).

4. Modellierungstechnik

Auch in der Informatik ist in den letzten Jahrzehnten eine Vielzahl von Emotionsmodellen entstanden, welche sich in ihren grundlegenden Annahmen, ihrer Terminologie, ihren Gültigkeitsgrenzen und ihrem Erklärungspotenzial sehr stark unterscheiden. Aufgrund dieses starken Wachstums ist die informatische Emotionsforschung mit vielen Problemen konfrontiert (vgl. Kapitel 1.2), darunter die starke Forschungsfragmentierung aufgrund der Schwierigkeit der Analyse und des Vergleichs der verschiedenen Architekturen: “An investigation into the computational modeling of emotion leads the developer to many publications of independent work conducted over at least the last two decades. Some of this research is fragmented, leading to little further work or potential for subsequent development, but can still add value in the pursuit of computational models of emotions for developers looking to model specific aspects of emotion“ ([SMBB08], S. 203).

Der Vergleich verschiedener Emotionsarchitekturen wird insbesondere durch das Fehlen einer allgemein anerkannten Terminologie erschwert und letztendlich ist die konzeptuelle Verwirrung und Forschungsfragmentierung der zeitgenössischen informatischen Emotionsforschung (vgl. z.B. [SS02]) u.a. auf das Fehlen einer integrativen theoretischen Grundlage zurückzuführen. In dieser Arbeit wird u.a. das Ziel verfolgt, die existierenden informatischen Emotionsmodelle auf einer abstrakten, architekturunabhängigen Ebene zu beschreiben und zu analysieren, wobei eine universelle Modellierungstechnik, mit der sich sowohl sozialwissenschaftliche Emotionstheorien als auch informatische Emotionsarchitekturen beschreiben lassen, die Basis für die Analyse und den Vergleich der Emotionstheorien und -modelle bildet (vgl. Kapitel 1.3).

4.5.2. Integration auf Basis der Modellierungstechnik

Vor dem Hintergrund der Fragmentierung im Bereich der Emotionserforschung und -modellierung ist die Integration verschiedener emotionstheoretischer Ansätze zu einem allgemeinen Modell besonders wünschenswert. Eine geeignete Repräsentationssprache könnte nach der hier vertretenen These den Vergleich verschiedener Emotionstheorien und -modelle erleichtern und ihre Integration in ein gemeinsames Referenzmodell ermöglichen. Dazu sollte die Modellierungstechnik verschiedene emotionale Konzepte repräsentieren können und einen Übergang zur operationalen Darstellung ermöglichen, wenn die Modelle in konkrete (Simulations-)Architekturen und Prozesse umgesetzt werden. Das heißt, dass von der Modellierungstechnik einerseits gefordert wird, dass sie emotionale Konzepte auf abstrakter, aber operationaler Ebene in Beziehung setzt, so dass die entstehenden Modelle ausführbar sind. Andererseits sollte die Modellierungstechnik emotionale Architekturen von Agenten und Agentensystemen modellieren und emotionale Prozesse auf verschiedenen Ebenen abbilden können. Schließlich sollte sie auch zur Darstellung sozialwissenschaftlicher Emotionstheorien anwendbar sein.

Mit derselben Modellierungstechnik sollten sich ferner sowohl emotionale Abläufe der Makroebene in soziologischen Strukturen als auch psychologische und neurobiologische Prozesse der Mikroebene darstellen lassen. Gleichzeitig muss diese Repräsentationssprache eine formale Spezifikation der emotionalen Abläufe und ihre Abbildung in einem informatischen System ermöglichen. Ferner sollte sie intuitiv und leicht verständlich sein, um zur Vereinfachung interdisziplinärer Forschungsdiskussion, der Überwindung der Terminologie-

4.5. Anwendung des sozionischen Ansatzes bei der Modellierung von Emotionstheorien

unschärfe und zur Integration verschiedener Emotionstheorien beitragen zu können (vgl. [SS02], [Slo04], [MFvLvS06]).

4.5.3. Zusammenfassung und weiteres Vorgehen

In Abschnitt 4 wurden die Agentennetze nach [Röl04] als Instrument zur Modellierung von Agenten und Multiagentensystemen vorgestellt. Ferner wurde der in [vLMV03] spezifizierte Ansatz zur sozionischen Theorienmodellierung mit Hilfe von Petrinetzen beschrieben.

Im Folgenden soll dieses Verfahren zur Repräsentation der Emotionstheorien aus Psychologie, Neurowissenschaft und Soziologie angewendet werden (siehe Kapitel 6). Ausgewählte emotionstheoretische Ansätze aus Psychologie, Soziologie und Neurowissenschaften werden zusammengetragen und unter Verwendung der sozionischen Modellierungsverfahren nachgebildet. Die Möglichkeiten der Integration verschiedener Ansätze werden aufgezeigt und diskutiert.

Das hier gewählte Vorgehen der grafischen Darstellung und schrittweisen Verfeinerung der theoretischen Konzepte mit dem *operationalem* Modellierungsformalismus ist besonders vorteilhaft, da die erstellten Modelle gleichzeitig eine Implementierung des modellierten Gegenstandes darstellen.

Die Integration der psychologischen bzw. neurowissenschaftlichen und der soziologischen Perspektiven auf Emotion erfolgt auf Basis einer generellen Modellierungstechnik der Petrinetze, welche die Entwicklung eines integrativen Modells ermöglicht, die (a) im MAS-Kontext einsetzbar ist und (b) eine theoretische Grundlage für die Implementierung der Emotion in (verteilte) künstliche Systeme liefert.

Die Übersicht über die Emotionstheorien soll die Entstehung und Funktion der Emotion bei einem Akteur, einer Gruppe bzw. einem natürlichen Organismus in der für einen Informatiker leicht zugänglichen Beschreibungssprache der Petrinetz-Modelle darstellen und dadurch den Transfer der modellierten sozialwissenschaftlichen Konzepte in die Informatik-Domäne erleichtern. Ziel ist es dabei, der konzeptuellen und terminologischen Verwirrung und Forschungsfragmentierung auf dem Gebiet der informatischen Emotionsmodellierung entgegenzuwirken und die emotionstheoretische Fundierung der Emotionsmodelle der Informatik langfristig zu verbessern.

Der besondere Schwerpunkt wird, entsprechend dem Gesamtziel der Arbeit (vgl. Kapitel 1.3), auf die Aufbereitung der emotionstheoretischen Grundlage für Emotionsmodellierung in Agentensystemen gelegt. Deshalb werden in erster Linie solche sozialwissenschaftlichen Theorien untersucht, die die Erfassung funktionaler und sozialer Emotionskomponenten ermöglichen bzw. Ansatzpunkte für die Integration akteursexterner soziologischer und akteursinterner psychologischer Sichtweisen im Hinblick auf die Strukturation und Kontrollierbarkeit künstlicher Agentengesellschaften bieten. Damit wird die bevorzugte Betrachtung soziologischer und neurowissenschaftlicher Ansätze begründet, welche schwerpunktmäßig die Problematik der Gesellschaftsstrukturation und Normetablierung durch Emotion betrachten.

Im Anschluss an die Betrachtung sozialwissenschaftlicher Emotionstheorien werden in Kapitel 7 die informatischen Architekturen für die emotionalen Agenten- und Multiagentensysteme vorgestellt und mit der gleichen Technik modelliert. Die Architekturen werden

4. Modellierungstechnik

im Hinblick auf ihre psychologische, sozialwissenschaftliche bzw. neurowissenschaftliche Fundierung sowie die verwendete Terminologie und Validität ausgewertet.

Teil II.

Theorien und Modelle

5. Emotionsdefinition

Der Terminus Emotion beinhaltet eine Reihe von komplexen Phänomenen, die nur schwer in einer gemeinsamen Definition zusammengefasst werden können. Das Emotionsphänomen kann aus verschiedenen Perspektiven erklärt werden, die jeweils neurobiologische, physiologische, kognitive, kommunikative oder soziale Komponenten der Emotion erfassen. Ja nach Forschungsperspektive geraten verschiedene Fragen in den Vordergrund: Welche Einflussfaktoren spielen bei der Entstehung von Emotion eine Rolle? Kommen hierbei hauptsächlich biologische oder kulturelle Faktoren zur Geltung? Sind vor allem individuelle Empfindungen oder soziale Normen ursächlich für die Emotionsentstehung? Welche Funktion(en) hat Emotion? Ist ihre Auswirkung auf das Individuum begrenzt oder wird auch das soziale System, die Gesellschaft von ihr beeinflusst?

“Angesichts der Tatsache, dass die psychologische Emotionsforschung auf eine etwa hundertjährige Tradition zurückschauen kann, sollte man erwarten, dass inzwischen Einigkeit darüber besteht, was unter Emotionen verstanden wird. [...] Dennoch ist “Emotion“ eines der am schlechtesten definierten Konzepte in der Psychologie.“ ([SA81], S. 24ff.)

Je nachdem, welche Forschungsfrage im Fokus der Betrachtung steht, wird der Begriff *Emotion* jeweils anders definiert [ZS00]. Aus diesem Grunde gibt es auch keine interdisziplinär akzeptierte Definition. Stattdessen findet man in der einschlägigen Literatur eine große Zahl unterschiedlicher expliziter und impliziter Definitionen. In [KK81] kamen Autoren nach einer eingehenden Analyse der vielfältigen Emotionsdefinitionen zu dem Ergebnis, dass jede Emotionsdefinition im Wesentlichen durch die Vorannahmen der Forscher bestimmt und daher nicht konsensfähig ist. Auch noch fünfzehn Jahre später stellten [SA96] fest: “Bislang ist kein Konsens festzustellen, was man unter einer Emotion zu verstehen hat“ ([SA96], S. 18).

In diesem Abschnitt werden die Emotionsdefinitionen von Forschern aus der Emotionspsychologie und Neurobiologie zusammengetragen mit dem Ziel, die Abgrenzung von Emotion zu Begriffen wie “Reflexe“, “Triebe“, “Affekte“, “Stimmungen“ und “Gefühle“ zu finden und die Rolle der Emotionen für das Funktionieren eines Organismus aufzuzeigen.

Die Emotionsdefinitionen werden im Hinblick auf die Umsetzbarkeit und Verwendbarkeit für die informatische Modellierung bewertet. Ausgewählte Emotionsdefinitionen werden mit Petrinetzen illustriert, einerseits um die konzeptuellen und terminologischen Zusammenhänge zu veranschaulichen, andererseits um verschiedene Definitionen zu vergleichen bzw. zueinander in Beziehung setzen zu können. Die Petrinetz-Modelle dienen hierbei lediglich der Veranschaulichung der Emotionsdefinitionen bzw. einzelner Theorien-Aussagen und erheben keinesfalls den Anspruch, als semantisch korrekte Modelle der Theorien bzw. als “Modelle der Emotion“ verstanden zu werden. Schließlich wird aus den erstellten Netzmodellen eine dieser Arbeit zu Grunde liegende Emotionsdefinition für technische Systeme abgeleitet.

5. Emotionsdefinition

Motivation für eine explizite Emotionsdefinition aus Sicht der informatischen Modellierung

Für die informatische Modellierung stellt das Fehlen einer gemeinsamen Emotionsdefinition ein grundlegendes Problem dar, denn informatische Modelle, die auf verschiedenen Emotionsdefinitionen gründen, sind nur bedingt miteinander vergleichbar; die informatische Diskussion wird durch stark divergierende Emotionskonzepte verschiedener Forscher sehr erschwert.

In [Slo04] wurde kritisiert, dass die Theorien sowie die Fragen, die sie zu beantworten suchen, sich als zumindest ambivalent herausstellen, wenn die Konzepte, die theoretische Annahmen und Aussagen formulieren sollen, mehrdeutig sind.

“For instance, if our concept of “emotion“ is ambiguous then so also will many questions about emotions be ambiguous, e.g. how emotions evolved, what their functions are, which animals have them, which brain mechanisms produce them, what types there are, whether they are needed for intelligence, whether a foetus has them, etc.“ ([Slo04], S. 1).

Der alltägliche Sprachgebrauch durchmischt das Verständnis von “Emotion“ mit dem von “Motivation“, “Einstellung“, “Stimmung“ und anderen Bezeichnungen für affektive Zustände und Prozesse und ist deshalb zu mehrdeutig und unklar, um einen Ausgangspunkt für wissenschaftliche Fragenstellungen und Theorienbildung, geschweige denn für die Umsetzung in einem Computersystem zu bilden. [Slo04] kritisiert: “But that does not prevent scientists assuming that these concepts can be used to formulate clear, unambiguous scientific questions or theories. For instance, sometimes people who argue that emotions are needed for intelligence are merely defending the truism that motivation is needed for action, and preferences are needed for selecting between options“ ([Slo04], S. 2).

Eine solche Übergeneralisierung des Begriffs Emotion führt dazu, dass viele informatische “Modelle der Emotion“ alle möglichen kognitiven Mechanismen modellieren, die mit der Emotion in natürlichen intelligenten Systemen nicht viel gemeinsam haben. Folglich erfüllen Emotionen in informatischen Modellen sehr unterschiedliche Funktionen, die jedoch modellspezifisch und nicht auf andere Modelle übertragbar sind, da der Emotionsbegriff und deshalb auch die Rolle der Emotion innerhalb des Systems in verschiedenen Modellen sehr unterschiedlich gefasst wird.

Wenn die Betrachtung von Emotionen beispielsweise so weit gefasst wird, dass sie Ziele und Präferenzen mit einschließt, können ganz andere Aussagen etwa über ihre Rolle bei der Entscheidungsfindung getroffen werden, als wenn man Emotionen als Prozesse spezifiziert, bei denen ein Subsystem in das Funktionieren eines anderen eingreift [Slo04]. Im zweiten Fall trifft die Aussage, dass Emotionen für Intelligenz erforderlich sind, nicht zu: “[O]n the contrary, emotions of that sort can get in the way of sensible decisions and actions“ ([Slo04], S. 2). Sloman plädiert deshalb für eine genaue Definition des Emotionsbegriffs und eine subtile Unterscheidung der verschiedenen Emotionsarten:

“We should not put both (i) a general preference for saving effort, and (ii) fear produced by a stampeding herd, in the same conceptual basket when they have

so very many differences, including the (relative) permanence of the first and the transience of the second“ ([Slo04], S. 2).¹

Funktionale Emotionsdefinition

Einige Forscher ([Fri09], [Slo01]) weisen auf eine grundsätzliche Unmöglichkeit der konzeptionellen Trennung von emotionalen und kognitiven Prozessen in einem intelligenten System hin. So gibt es für Frijda z.B. keine scharfe Abgrenzung zwischen Emotionen und anderen mentalen Phänomenen wie Zielen, Affekten, Motivationen usw. Er führt aus, dass das Wort “Emotion“ sich nicht auf eine “natürliche Klasse“ bezieht und deshalb nicht auf eine wohldefinierte Klasse von Phänomenen verweisen kann, die klar unterscheidbar ist von anderen mentalen und Verhaltensprozessen. Aus diesem Grunde sei eine *funktionale Beschreibung* bei der Begriffsbestimmung der konzeptuellen Definition vorzuziehen [Fri86], [Fri09].

Grundsätzlich verlangt die Emotionsmodellierung im Sinne der technischen Systemoptimierung nach einer *funktionalen* Definition der Emotion: “Our best hope is to define emotions in terms of *their functional role*[Hervorh. i.O.]“ ([Sch02], S. 3). Informatik ist also an einer funktionalen Beschreibung von Emotion interessiert, welche die Emotionsentstehungsprozesse (Emotionsgenerierung) und Auswirkungen der Emotion auf das Systemverhalten spezifiziert. Wir suchen deshalb nach ganzheitlichen Ansätzen, manchmal als “Funktionsmodelle der Emotion“ [McD69] bezeichnet, die Emotion als Reaktion auf Ereignisse im Kontext der Zielverfolgung betrachten und insbesondere die Implikationen von Emotionen bei kognitiven Prozessen und bei der Handlungssteuerung zum Forschungsgegenstand machen.

Abb. 5.1 aus [RE09] veranschaulicht die funktionale Beziehung zwischen Emotion, Motivation und Handeln. Der Emotionsprozess [Fri86] schließt demnach sowohl *Emotionsentstehung* (durch Bewertung der Umweltereignisse und Handlungsergebnisse), als auch *Emotionsfolgen* (Generierung von Handlungsimpulsen, Verhaltenssteuerung usw.) ein. Es fällt auf, dass sowohl für die Bewertung der Situation als auch für die Generierung von Motivationen bzw. Handlungen individuelle Ziele, Anforderungen und Ressourcen eine Rolle spielen. Es werden demnach die gleichen Einflussgrößen betrachtet und teilweise die gleichen Begriffe benutzt, die auch bei der Beschreibung von Planungs- und Handlungswahlprozessen eines autonomen (BDI-)Agenten gebraucht werden (siehe Kapitel 3.2.5).

Eine solche Sichtweise auf Emotion kommt dem Wunsch der Informatik nach einer funktionalen Beschreibung des Emotionsmechanismus sehr nahe und bietet daneben eine gute Anbindungsmöglichkeit an die Agenten-Konzepte. In diesem Kapitel werden wir nach Emotionsdefinitionen und theoretischen Ansätzen suchen, welche den in Abb. 5.1 spezifizierten Emotionsprozess ganz oder teilweise abdecken und die einzelnen Prozessschritte näher spezifizieren.

In Abb. 5.2 wird die gesuchte funktionale Beschreibung des Emotionsprozesses, die wir im Folgenden als “funktionale Emotionsdefinition“ bezeichnen, mit einem Petrinetz-Modell

¹[Slo04] schlägt in diesem Zusammenhang die Einführung des Begriffes “affect“ mit Unterbegriffen “emotion“, “desire“, “preference“, “attitude“, “value“, “mood“, usw. vor, um die Bedeutung des Emotionsbegriffs in verschiedenen Zusammenhängen genauer spezifizieren zu können ([Slo04], S. 2).

Emotion, Motivation und Handeln: funktionale Beziehungen

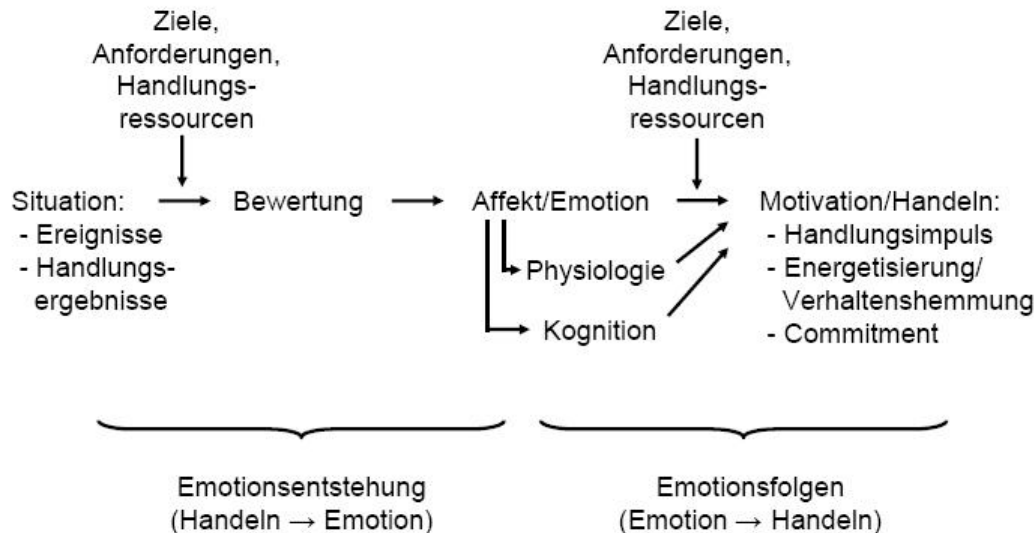


Abbildung 5.1.: Wechselseitige Beeinflussung von Emotion und Handeln (nach [RE09], S. 685).

veranschaulicht. Dieses Modell wird uns im Folgenden als Referenzmodell bei der Auswertung der verschiedenen Emotionsdefinitionen im Hinblick auf die Emotionsmodellierung in den informatischen Systemen dienen.

Jede der im Folgenden betrachteten Emotionsdefinitionen ist meist Teil einer umfassenden Emotionstheorie, auf die jedoch in diesem Kapitel noch nicht eingegangen werden soll.

Vielmehr soll hier das Ziel sein, den Emotionsbegriff zu anderen mentalen Phänomenen abzugrenzen und aus der Fülle der einschlägigen Emotionskategorisierungen und Sichtweisen auf Emotionen solche herauszufiltern, die im Kontext der künstlichen intelligenten Systeme anwendbar sind. Die zugehörigen Emotionstheorien werden im Kapitel 6 näher betrachtet; dabei werden einige der hier vorgestellten Emotionsdefinitionen und ihre Netzmodelle unter Hinzuziehung der zugehörigen Emotionstheorien konkretisiert und weiter ausgebaut.

5.1. Abgrenzung von Emotionen zu verwandten mentalen Phänomenen

Die Begriffe *Affekt*, *Gefühl* und *Stimmung* werden in der Emotionspsychologie ähnlich mehrdeutig verwendet wie der Begriff *Emotion*. Bei den Überlegungen bezüglich der gegenseitigen Abgrenzung herrscht in der psychologischen Forschung noch keine Einigkeit. Auch in der sozialwissenschaftlichen Literatur werden diese Begriffe nicht scharf getrennt und oft synonym verwendet (vgl. z.B. [Fla03], [Bar06]).

5.1. Abgrenzung von Emotionen zu verwandten mentalen Phänomenen

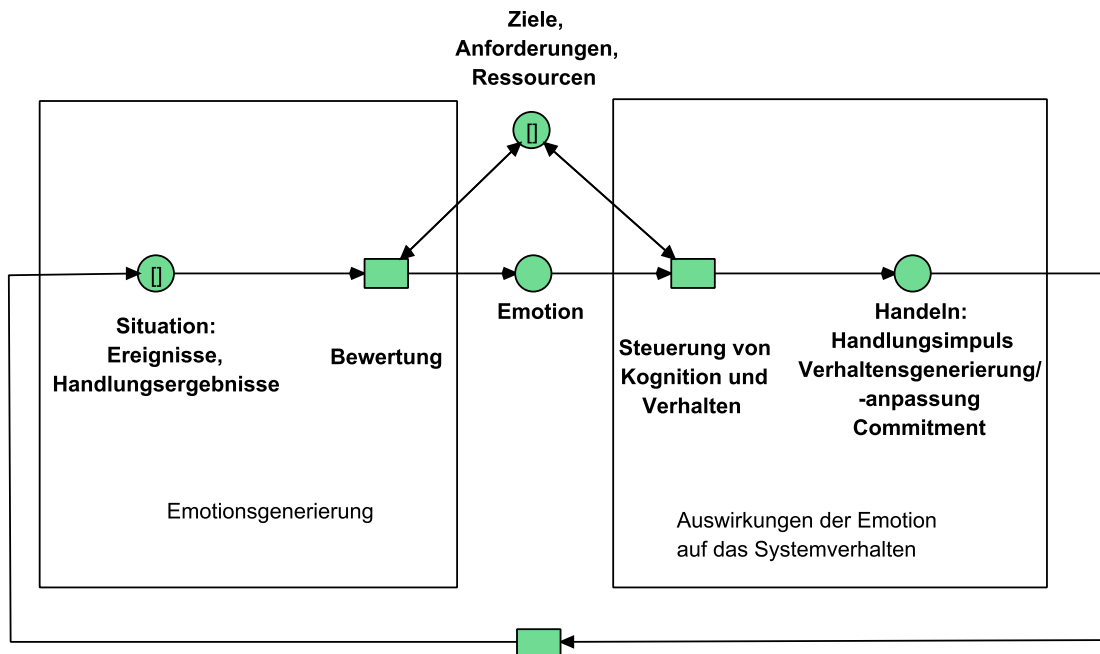


Abbildung 5.2.: Funktionale Emotionsdefinition.

In diesem Abschnitt soll der Begriff “Emotion“ zu verwandten, oft gleichbedeutend verwendeten Begriffen abgegrenzt werden. Von verschiedenen Autoren vorgeschlagene Abgrenzungsmöglichkeiten werden hierzu vorgestellt.

Emotion vs. Affekt Intensive, heftige, kurzzeitig auftretende Emotionen mit desorganisierender bzw. einengender Wirkung auf das Verhalten und Erleben werden im Deutschen oft als *Affekte* bezeichnet. Dieser Begriff spielt in der gegenwärtigen Emotionspsychologie kaum eine Rolle und findet eher in der Psychiatrie zur Kennzeichnung kurzfristiger und besonders intensiver Emotionen, die oft mit einem Verlust der Handlungskontrolle einhergehen, Verwendung [OEM00].

Im Englischen wird der Begriff “affect“ als Synonym für Emotion gebraucht (z.B. “affective state“, “affective computing“), aber auch als übergeordneter Begriff für Emotion und Stimmung oder als Bezeichnung für das subjektive Erleben. In [MF00] wird Affekt als “the feeling of emotion“ definiert ([MF00], S. 60): Es soll dem Organismus eine erste schnelle Einordnung der Situation ermöglichen.

Emotion vs. Stimmung Längerfristige emotionale Tönungen des Erlebens ohne klaren Reiz-, Situations-, Tätigkeits- und Bedürfnisbezug heißen Stimmungen [Vaa03]. Stimmungen können von diversen Faktoren, wie z.B. Tages- oder Jahreszeit abhängig sein [WR20].

Silberer (1999) definiert Stimmung als eine “momentan subjektiv erfahrene Befindlichkeit“ (Silberer 1999, S. 32).

5. Emotionsdefinition

Nach Mandl und Huber (1983) schaffen Stimmungen “ein dauerhaftes, über längere Zeiträume relativ unverändertes Bezugssystem für alle informationsverarbeitenden Prozesse - beispielsweise eine gehobene oder depressive Stimmung“ ([Man83], S. 5).

Nach Watson und Clark (1994) haben Stimmungen keine so spezifische gefühlsbezogene Griffigkeit wie prototypische Emotionen, sondern bestehen eher in einem angenehmen oder unangenehmen Gefühl, das in seiner Intensität variieren kann.

Morris (1989) veranschaulicht die Beziehung zwischen Stimmungen und Emotionen anhand einer Figur-Grund-Beziehung: Der Emotion (Figur) wird die meiste (bewusste) Aufmerksamkeit gewidmet, der Stimmung (Grund) kommt dagegen wenig Aufmerksamkeit zu. Jedoch kann auch der Stimmung Aufmerksamkeit geschenkt werden, besonders wenn ihre Intensität im Vergleich zur Emotion relativ hoch ist oder wenn die Aufmerksamkeitsanforderungen niedrig sind.

Im Unterschied zu Emotionen haben Stimmungen geringere Intensität, längere Dauer und sie sind nicht objektgerichtet ([Mor89], [Sch99a], [SA96]). Die Unterscheidung wird jedoch nur selten expliziert, ist bei verschiedenen Autoren unterschiedlich ausgeprägt und beruht hauptsächlich auf quantitativen Kriterien wie Intensität und zeitliche Ausdehnung, ohne eine qualitative Abgrenzung vorzunehmen. Oft werden weitere Unterscheidungskriterien postuliert (z.B. vorausgehende Bedingungen, beteiligte biologische Prozesse), die noch weniger konsensfähig sind (vgl. Davidson und Ekman, 1994).

In [MSR01] und [OEM00] werden alle vorgeschlagenen Unterscheidungskriterien für fragwürdig erklärt. Die Abgrenzung von Stimmung und Emotion wird als ein Problem gesehen, das erst - wenn überhaupt - durch die zukünftige Forschung zu lösen ist: “[U]ns erscheint es zum gegenwärtigen Zeitpunkt nicht sinnvoll zu sein, eine weitergehende, prinzipielle Unterscheidung von Stimmung und Emotion vorzuschlagen“ ([OEM00], S. 14).

Emotion vs. Gefühl Mit dem Begriff Gefühl ist im deutschen Sprachgebrauch oft eine subjektive Erlebensqualität einer sich auch anders zeigenden Emotion gemeint (vgl.[OEM00]). Nach dem Reaktionstrias-Konzept (Eysenck (1965), Lang (1968), [LAO70]; siehe auch Kapitel 5.3) ist Gefühl ein subjektiv erlebter Teilaspekt einer Emotion. Die Bezeichnung “Emotion“ steht im Deutschen üblicherweise für eine weitere Auffassung und einen Oberbegriff, der neben dem körperlichen Zustand und dem Ausdrucksverhalten auch das Gefühl als Komponente mit einschließt [OEM00].

Die missverständliche Verwendung der Termini hat sowohl in der Vergangenheit als auch in der Gegenwart heftige Debatten ausgelöst. James (1894) gebraucht die Begriffe “Gefühl“ und “Emotion“ synonym. Bischof (1989) versteht den emotionalen Zustand als physiologisch und das erlebte Gefühl als psychisch konnotiert.

Die meisten Autoren (z.B. [DLD⁺01], [OCC88], [Dam03]) sind sich darin einig, dass das Erleben eines Gefühls das Bewusstwerden des emotionalen Erlebens voraussetzt. Jedoch stellt eine solche Definition keine klare Spezifizierung dar, da die Präsenz des “subjektiven emotionalen Erlebens“ objektiv nicht überprüfbar ist.

Künstliche (und auch primitive natürliche) intelligente Systeme können demnach keine Gefühle “empfinden“, solange sie kein Bewusstsein besitzen (vgl. z.B. [OCC88]): “Our view

5.1. Abgrenzung von Emotionen zu verwandten mentalen Phänomenen

is that the subjective experience of emotion is central, and we do not consider it possible for computers to experience anything until and unless they are conscious“ ([OCC88], S. 182).

Einen gangbaren Mittelweg scheint Damasio's Abgrenzung von Emotion und Gefühl darzustellen. Demnach ist Emotion ein Prozess, der von außen in Form von Aktionen oder Bewegungen wie Gesichtsausdrücken, Stimmchwankungen und spezifischen Verhaltensweisen beobachtbar ist bzw. durch Hormonspiegelmessungen oder elektrophysiologische Untersuchungen sichtbar gemacht werden kann. Dagegen sind Gefühle versteckt und keineswegs für Außenstehende beobachtbar. Emotionen spielen sich im Körper ab, Gefühle im Geist [Dam03].

Reflexe und physiologische Triebe Lazarus (1991) sieht Reflexe, Triebe und Emotionen als drei verschiedene Anpassungsmechanismen, die zur optimalen Anpassung des Individuums an seine Umwelt beitragen sollen. Alle drei Mechanismusklassen sind nach Lazarus darauf spezialisiert, ein spezifisches Anpassungsproblem zu entdecken und Verhalten hervorzurufen, das das Anpassungsproblem löst. Emotionen haben in den einfacheren Klassen von Anpassungsmechanismen, nämlich in den Reflexen und den physiologischen Trieben, ihren phylogenetischen Ursprung ([SL90], S. 612).

Die Reflexe und Triebe grenzt Lazarus von den Emotionen folgendermaßen ab:

- Reflexe stellen einen effektiven Anpassungsmechanismus dar, wenn ein Anpassungsproblem in zuverlässiger und eindeutiger Weise von einem spezifischen Reiz signalisiert wird und eine stereotype Reaktion zur erfolgreichen Bewältigung dieser Bedingung genügt. Zur Bewältigung der komplexeren Anpassungsprobleme reicht die starre Verbindung von Reiz und Reaktion nicht mehr aus.
- Die physiologischen Triebe zur Befriedigung von grundlegenden Bedürfnissen wie z.B. Hunger und Durst stellen nach Lazarus einen ersten evolutionären Schritt in Richtung der Flexibilisierung von Anpassungsmechanismen dar. Zwar werden die physiologischen Triebe ähnlich wie die Reflexe von vergleichsweise spezifischen Reizen ausgelöst (wie z.B. dem Blutzuckerspiegel im Fall von Hunger), sie unterscheiden sich jedoch von den Reflexen durch eine höhere Flexibilität im Verhalten insofern, als sie eine ganze Klasse von Verhaltensweisen hervorrufen können. Diese gegenüber den Reflexen deutlich höhere Flexibilität des Verhaltens ermöglicht dem Organismus eine bessere Anpassung an unterschiedliche äußere Gegebenheiten, als es nur mit Hilfe der starren Reiz-Reaktions-Verbindung der Reflexe möglich wäre.

Entwicklungsgeschichtlich gesehen, entstanden Reflexe und physiologische Triebe als “Mechanismen, die das Entdecken von überlebensrelevanten Bedingungen mit dem Erzeugen von überlebenssteigernden Verhalten verbinden“ ([Laz91], S. 612). Nach Lazarus haben sich Emotionen bei höheren Organismen in einem weiteren evolutionären Schritt zur Bewältigung komplexer Anpassungsprobleme entwickelt, und zwar jede Emotion für ein ganz bestimmtes Problem. Im Unterschied zu Reflexen und Trieben lösen Emotionen nur grobe Handlungsimpulse oder Handlungstendenzen aus, deren Realisierung die Bewältigung des für die jeweilige Emotion spezifischen Anpassungsproblems begünstigt. Die konkrete

5. Emotionsdefinition

Gestaltung des Verhaltens wird durch solche Impulse allerdings laut Lazarus nur gering oder gar nicht eingeschränkt.

Elementare Affektzustände vs. Emotionen Bei den Versuchen, Emotionen zu definieren, greifen manche Autoren auf die physiologischen Begleiterscheinungen von emotionalen Prozessen zurück und betrachten sie als alleinige objektive Kriterien, um Emotionen zu identifizieren und zu klassifizieren (z.B. James, James-Lange, siehe Kapitel 6.1.1). Solche beobachtbaren physiologischen Emotionsmanifestationen sind z.B. Veränderungen des Gesichtsausdrucks, der Gestik und Körperhaltung, des Blutdrucks und der allgemeinen Muskelanspannung. Außerdem werden emotionale Prozesse von unzähligen Veränderungen in den internen homeostatischen Prozessen begleitet, z.B. hormonelle Veränderungen, die aber den einfachen Messmethoden nicht zugänglich sind.

Jedoch werden nicht alle Emotionen von physiologischen Veränderungen begleitet, so dass durch diese Herangehensweise nur eine Gruppe der sog. *elementaren oder primären Emotionen* (z.B. in der Definition von [Laz91]) erfasst wird, die oft als *elementare Affektzustände* bezeichnet werden.

[Pan98] folgend, sieht [Rot99] die gemeinsame Basis für die Emotionsidentifikation in der Unbestreitbarkeit der Existenz von Erlebniszuständen, die zusammen mit biologischen Grundbedürfnissen und ihrer Befriedigung auftreten, z.B. im Zusammenhang mit Nahrungs- und Flüssigkeitsaufnahme, Wachen und Schlafen, Wärmehaushalt, Fortpflanzung, Angriffs- und Verteidigungsverhalten, Verletzungen und Sozialverhalten. Nach [Pan98] sind solche elementaren Affektzustände angeboren und - soweit untersucht - allen Säuglingen gemeinsam.

Die meisten physiologisch orientierten Autoren (z.B. Ciompi, Ekman, Panksepp, Tomkins) schließen zumindest einige elementare Affektzustände in den Begriff der Emotion ein, z.B. Aggression, Wut, Gefühle der Verlassenheit, Geborgenheit und Fürsorge, während sie andere wie Hunger, Durst, Müdigkeit in aller Regel nicht als Emotionen ansehen.

Andere Autoren (z.B. [OCC88]) beschränken Emotionen auf Zustände wie Furcht, Freude, Glück, Verachtung, Ekel, Neugierde, Hoffnung, Enttäuschung und Erwartung, die durch positive und negative Erfahrungen in stärkerem Maße veränderbar sind und eine Form der Bewertung lebens- und überlebensrelevanter Ereignisse (besonders im sozialen Bereich) darstellen. Bei dieser in der Emotionspsychologie verbreiteten Betrachtungsweise werden die meisten elementaren affektiven Zustände aus dem Kreis der Emotionen ausgeschlossen.

In einigen Emotionsdefinitionen werden die elementaren Empfindungen des Schmerzes und der Lust ebenfalls als Emotionen oder als emotionale Dimensionen dargestellt (vgl. [Dam03], [Rol05]). Andere Emotionsforscher widersprechen: "However one can be enjoying a meal, a view, or a conversation, without being at all emotional about it" ([Slo03], S. 70).

Damasio (2003) sieht elementare Affektzustände als unvorhersagbare Ergebnisse des Zusammenspiels mehrerer niedriger regulatorischer Prozesse (metabolische Prozesse, Reflexe, aktuell relevante Triebe, Umweltsituation usw.). In seiner Emotionstaxonomie bezeichnet er diese Zustände als *Hintergrundemotionen* und betrachtet sie als Basis für die "eigentlichen Emotionen" (*emotions-proper*) (vgl. Kapitel 5.5).

Die Abgrenzung der Emotionen, insbesondere der Basisemotionen, gegenüber der physiologischen Bedürfnissen auftretenden Wahrnehmungen und Erlebnissen (Hunger, Durst,

Müdigkeit) ist nach wie vor sehr wage und stellt eine der kritischen Stellen der Emotionsdefinition dar, über die in der Forschung noch keinerlei Einigkeit herrscht.

Auswertung für die informatische Modellierung In der aktuellen emotionspsychologischen Forschung wird üblicherweise keine strukturelle bzw. qualitative Unterscheidung zwischen *Affekt*, *Stimmung* und Emotion gemacht. Die vorgeschlagenen Unterscheidungskriterien sind meist quantitativer Art. Die Umsetzung dieser Konzepte in informatischen Modellen zusätzlich zu Emotionen erscheint daher nicht sinnvoll. Auch eine Veranschaulichung der betrachteten Definitionen mit Petrinetzmodellen wird nicht vorgenommen, da sie wegen der rein quantitativen Beschreibungen offensichtlich keine neuen Erkenntnisse bezüglich der strukturellen Unterschiede der diskutierten Konzepte zum Emotionskonzept erwarten lässt.

Gefühl wird hingegen von den meisten Autoren als Teilkomponente einer Emotion angesehen, entsprechend dem Reaktionstrias-Konzept, welches im nächsten Abschnitt näher erläutert und durch ein Petrinetzmodell veranschaulicht wird.

Reflexe und *physiologische Triebe* werden aus der evolutionspsychologischen Perspektive als Vorstufen der Emotion gesehen. Alle drei Mechanismen dienen der optimalen Anpassung an sich verändernde Umweltbedingungen und unterscheiden sich durch die zunehmende Flexibilisierung der Reaktionen auf spezifische Reize: Reflexe stellen starre Reiz-Reaktions-Muster dar, während Emotionen grobe Handlungsimpulse (Handlungstendenzen) vorgeben, ohne das Verhalten wirklich einzuschränken.

Als *elementare Affektzustände* werden solche Emotionen bezeichnet, die im Zusammenhang mit biologischen Grundbedürfnissen und ihrer Befriedigung auftreten und mit physiologischen Veränderungen einhergehen. Bezüglich der Abgrenzung der Emotionen, insbesondere der Basisemotionen, zu den elementaren Affektzuständen besteht in der aktuellen psychologischen und neurophysiologischen Forschung kein Konsens. Die Entstehungsmechanismen der elementaren Affektzustände werden nicht genau spezifiziert, was die Modellierung dieser Mechanismen kaum möglich macht.

Aus der informatischen Perspektive der Anwendung von Emotionen für die funktionale Systemoptimierung erscheinen daher die Emotionsdefinitionen, die auf physiologische Emotionsmanifestationen zur Spezifizierung des Emotionsphänomens zurückgreifen, als nicht besonders wertvoll. Der zentrale Nachteil solcher Definitionen für unsere Zielsetzung ist die mangelnde Beleuchtung der funktionalen Rolle der Emotionen und ihrer Wechselwirkungen mit anderen kognitiven Mechanismen des intelligenten Systems, wie z.B. ihr Einfluss auf Perzeption, Planungsprozesse, Aufmerksamkeit, Motivation und Verhalten.

5.2. Emotionsklassifikationen

Wie lassen sich Emotionen kategorisieren? Gibt es Grunddimensionen, auf die sich Emotionen reduzieren lassen? Mit diesen Fragen beschäftigt sich die Emotionspsychologie schon seit längerer Zeit.

In diesem Abschnitt geben wir eine Übersicht über die verschiedenen Möglichkeiten der Emotionsklassifizierung. Zum einen wird die *kategoriale* Klassifikation von Emotion

5. Emotionsdefinition

vorgestellt, die *Basisemotionen* und *zusammengesetzte* Emotionen unterscheidet. Diese Herangehensweise impliziert die Annahme, dass eine beliebige Emotion sich aus Basisemotionen zusammensetzen lässt.

Bei der *dimensionalen* Klassifikation werden Emotionen im Gegensatz dazu als Ausprägung in bestimmten Dimensionen gesehen. Es wird versucht, diese grundlegenden Beschreibungsdimensionen aufzudecken.

5.2.1. Kategoriale Klassifikation

Bei der kategorialen Klassifikation geht man von der Annahme aus, dass sich eine begrenzte Anzahl von Emotionen identifizieren lässt, die “angeboren“, d.h. kulturell und sozial universell sind, und auf spezifischen neuronalen Aktivitäten beruhen. Aus diesen sog. *Basisemotionen* - manchmal auch *Primäremotionen* genannt - lassen sich alle anderen Emotionen zusammensetzen (sog. *zusammengesetzte* Emotionen).

Basisemotionen

Der zentrale Untersuchungsgegenstand der Basisemotionsforschung ist die Frage nach der Existenz einer begrenzten Anzahl von angeborenen diskreten Emotionen, die für die Auslösung und Differenzierung aller anderen Emotionen verantwortlich ist. In der Formulierung von Roth wird danach gefragt, “ob es beim Mensch (und eventuell bei Tieren) eine Grundausstattung mit einer gewissen Zahl unabhängig voneinander existierender Affekte bzw. Emotionen gibt“ ([Rot99], S. 265).

Die Kriterien für die Ernennung einer Emotion zur Basisemotion sind im Allgemeinen nicht eindeutig bestimmt. Als Ausgangspunkte dienen verschiedene Aspekte wie z.B. der Nachweis einer universellen, kulturunabhängigen Mimik (und Lautäußerungen) bei der Emotionsexpression, von emotionsspezifischen körperlichen Veränderungen, die Erregung verschiedener Gehirnareale bei unterschiedlichen Basisemotionen oder die Existenz elementarer Affektzustände, die nicht erlernt werden müssen, sondern angeboren und allen Säuglingen gemeinsam sind (z.B. Bück, 1985; Izard, 1977; Johnson- Laird and Oatley, 1989; Tomkins, 1962/1980; Lazarus, 1991; Panksepp, 1982; Plutchik, 1962/1980).

Die wichtigsten in der Literatur genannten Kriterien, anhand derer eine Emotion als Basisemotion klassifiziert werden kann, werden in [Buc00] zusammengefasst:

1. Atomarität: Eine Basisemotion ist elementar, sie enthält keine weiteren Emotionen als Bestandteile und kann nicht weiter reduziert werden. Im Unterschied dazu bestehen komplexe bzw. zusammengesetzte Emotionen aus mehreren Komponenten.²

²An dieser Stelle wird keine Aussage darüber getroffen, ob komplexe Emotionen auf Basisemotionen zurückgeführt werden können, bzw. worauf Basisemotionen ihrerseits reduzierbar sind. Nach einer *starken Hypothese* verfügen die Basisemotionen über einen monadischen Charakter und damit über eine “intrinsische Affektivität“ ([Rei00b], S. 211) auf; die *schwache Hypothese* besagt, dass Basisemotionen auf nicht affektive (kognitive) Komponenten zurückgeführt werden können (vgl. z.B. Dörner (1993), der Emotionen ausschließlich auf kognitive Mechanismen reduziert).

2. Universalität: Basisemotionen sind interkulturell und zwischen heterogenen sozialen Gruppierungen durchwegs beobachtbar. Dabei können sie jedoch kulturspezifisch ausgeformt sein, d.h. dass interkulturelle Unterschiede in der Ausprägung und Intensität der universellen emotionalen Grundmuster bestehen können. Im Unterschied dazu sind bestimmte komplexe bzw. zusammengesetzte Emotionen in einigen Kulturen nicht bekannt. Autoren, die die Existenz von Basisemotionen postulieren, berufen sich vor allem auf die kulturelle Universalität der Emotionsexpressionen ([EF75], [Ekm89]).
3. Distinkter Gesichts- und Körperausdruck, der von anderen spontan richtig gedeutet wird: Umfangreichen interkulturellen Untersuchungen von Paul Eckman zufolge sind mindestens sechs emotionale Zustände - Glück, Überraschung, Furcht, Ekel, Trauer und Zorn - durch Gesichtsausdrücke und Lautäußerungen charakterisierbar und werden von der großen Mehrheit aller Menschen in 21 Völkern der Erde spontan richtig gedeutet ([EF75], [Ekm82], [Ekm89], [Ekm92]). Allerdings räumt Ekman ein, dass es keinerlei Daten darüber gibt, wie viele Ausdrucksformen für eine Emotion universal sind. Auch gibt es keine Eins-zu-eins-Relation, d.h. die spezifischen Gesichtsausdrucksformen können nicht eindeutig auf Emotionen abgebildet werden (und vice versa). In [Rus94] bzw. [Rus03] werden die Untersuchungsergebnisse von Ekman in Frage gestellt.
4. Angeborenes Affektprogramm: Joseph LeDoux [LeD96a] und Jaak Panksepp ([Pan98], [Pan06]) postulieren die Existenz (semi)automatischer Reaktionssysteme ("response systems" oder "response components") phylogenetischen Ursprungs, die den Basisemotionen zugrunde liegende neurobiologische Mechanismen repräsentieren. Diese durch neuronale Verschaltungen fest verankerten Verhaltensprogramme, die nicht nur beim Menschen, sondern auch bei anderen Säugetieren nachgewiesen werden können, lösen als automatische Antwort auf einen spezifischen Reiz elementare emotionale Reaktionen aus. Damit kann jede Basisemotion als Ergebnis eines angeborenen Affektprogramms aufgefasst werden. Mit dieser Auffassung stellen die Fürsprecher der automatischen Affektprogramme ein zentrales Konzept physiologisch bedingter Basisemotionen zur Verfügung.

Panksepp geht von der Annahme deutlich abgrenzbarer affektiv-emotionaler Grundzustände aus, die auf unterschiedlichen neuronalen "Schaltkreisen" im Gehirn beruhen und deshalb als universell anzusehen sind. Er gelangt anhand der Verhaltensfolge nach elektrischer Reizung von Rattengehirnen zu sieben elementaren emotionalen Reaktionsmustern: Erwartung, Wut, Furcht, Wollust, Fürsorge, Panik und Freude [Pan98].

Nach LeDoux hat jede Basisemotion ihren phylogenetischen Ursprung in gewissen entwicklungsgeschichtlichen Erfordernissen im Hinblick auf Bewältigung bestimmter Anpassungsprobleme. Jede Basisemotion erfordert die Beteiligung spezifischer Subsysteme des Gehirns [LeD96a].

5. Unmittelbare und signifikante automatische körperliche Veränderungen: Basisemotionen werden von unmittelbaren körperlichen Veränderungen begleitet (wie Blutdruck- und Herzschlaganstieg, hormonelle Veränderungen, Schweißausbrüche etc.).

5. Emotionsdefinition

6. Basisemotionen treten in der Ontogenese sehr früh auf und verschwinden bei degenerativen Hirnerkrankungen sehr spät.
7. Neugeborene, Erwachsene, die unter Altersdemenz leiden, und sogar einige Säugetier-Arten zeigen Emotionsausdrücke, ohne dass sie über ein Selbstbild verfügen. Daraus leitet [Buc00] ab, dass kein Selbstbild erforderlich ist, um Basisemotionen zu empfinden. Er benennt eine Reihe von Emotionen, die in diesem Sinne kein solches Selbstbild benötigen, unter anderem solche, die allgemein zu den Basisemotionen gezählt werden, wie etwa Furcht, Ärger, Freude, Ekel oder Trauer. Um komplexe Emotionen empfinden zu können, müssen Subjekte im Gegensatz dazu über ein Selbstbild verfügen: "Emotions that are not basic, but are complex, are those requiring second-order intentionality. They depend on reference to a conception of the self and a social comparison of the present situation of the self with imagined alternatives of the self or others." [BZO96], S. 89).
8. Das letzte Kriterium besagt, dass keine höheren Kognitionen wie etwa Schlussfolgern, Planen, Entscheiden oder Werten zur Entstehung der Basisemotionen erforderlich sind (no thoughts necessary). Jedoch wird davon ausgegangen, dass kognitive Prozesse wie Reizverarbeitung oder Wahrnehmung an der Generierung von Basisemotionen beteiligt sind - eine Position, die in der Literatur relativ unstrittig ist (vgl. z.B. [CSC94], [Fri94]). Auch dieses letzte Kriterium ergibt sich aus Beobachtungen zu Emotionsphänomenen bei Säuglingen, Säugetieren und Demenzkranken.

Bedeutende Vertreter der Basisemotionenhypothese postulieren unterschiedliche Emotionstaxonomien, wobei keine Einigkeit darüber besteht, wie viel und welche Art Basisemotionen es gibt. Ekman ([Ekm92], S. 175) nennt neun Charakteristika von Basisemotionen: universelle Expression; Vorhandensein auch in nichtmenschlichen Primaten; distinktive Physiologie; einer Emotion vorausgehende Ereignismuster; Kohärenz emotionaler ANS-Reaktion und Expression; schneller Einsatz der Emotion; geringe Erregungsdauer; automatischer Einschätzungsprozess.

Sylvan Tomkins unterscheidet acht elementare Emotionen: Überraschung, Interesse, Freude, Wut, Furcht, Ekel, Scham, Angst.

Philip Johnson-Laird und Keith Oatley [OJL87] kamen anhand von linguistischen Untersuchungen zu einer ähnlichen Zusammenstellung.

Robert Plutchik bezog neben dem mimischen Ausdruck auch Handlungstendenzen und Gestik mit ein und fügte zur Ekmans Liste Akzeptanz, Erwartung und Erstaunen hinzu. Er ordnete sie kreisförmig an und vermutete, dass aus der Mischung zweier elementarer Emotionen abgeleitete Emotionen entstehen.

Tabelle 5.1 illustriert die von bedeutenden Vertretern der Basisemotionshypothese postulierten Emotionsskalen und die Kriterien für die Einschätzung einer Emotion als basal.

Auswertung für die informatische Modellierung

Die kategoriale Klassifikation von Emotionen in Basisemotionen und zusammengesetzte Emotionen sagt nichts über die Relation zwischen Emotionen und deren primäre Entste-

Author	Basic Emotion	Basis for Inclusion
Ekman	Anger, disgust, fear, joy, sadness, surprise	Universal facial expressions
Frijda	Desire, happiness, interest, surprise, wonder, sorrow	Forms of action readiness
Izard	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise	Hardwired link
James	Fear, grief, love, rage	Bodily involvement
Oatley, Johnson-Laird	Anger, disgust, anxiety, happiness, sadness	Do not require propositional content
Panksepp	Expectancy, fear, rage, panic, sexuality, care	Hardwired
Plutchik	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise	Relation to adaptive biological processes
Tomkins	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise	Density of neural firing
Weiner, Graham	Happiness, sadness	Attribution independent

Tabelle 5.1.: Basisemotionen (nach [OT90]).

hungsbedingungen aus. Daher liefert sie für die informatische Modellierung vorerst keine wesentlichen Erkenntnisse bezüglich der Entstehung und Wirkungsweise der emotionalen Mechanismen. In der psychologischen Literatur wurde mehrfach kritisiert, dass die Theorie der Basisemotionen keine strukturellen Aussagen über die Beschaffenheit bzw. Funktionalität von Emotionen trifft [OT90]. Ferner gibt es weder empirische Beweise für noch gegen die Existenz von Basisemotionen, es ist umstritten, wie viele und welche darunterfallen, und auch der Sinn und Zweck der Suche nach ihnen wurde angezweifelt: “The notion of basic emotions will not lead to a significant progress in the field“ ([OT90], S. 315).

Basisemotionen werden entweder als *biologisch verwurzelte* Verhaltensprogramme oder als *psychophysiologische* Reaktionsmuster [OT90] aufgefasst. Bei der ersten Annahme widmen sich die Forschungsarbeiten überwiegend der Diskussion verschiedener Skalen von Basisemotionen, der Bestimmung von Kriterien für die Identifizierung einer Emotion als basal bzw. der Suche nach empirischen Belegen. Bevorzugte Forschungsrichtung ist hierbei die Aufdeckung von Gemeinsamkeiten im Emotionsausdruck bei Menschen und Tieren oder bei Menschen verschiedener Kulturen. Hier finden sich kaum theoretische Definitionen bzw. Charakterisierungen, die für eine Formalisierung bzw. Operationalisierung in einem informatischen Modell geeignet wären. Bei der Betrachtung der Basisemotionen als physiologisch vorgegebene Reaktionsmuster bei spezifischen Stimuli liegt ihre Manifestation im Verhalten und Ausdruck im Zentrum der Forschungsinteresse.

5. Emotionsdefinition

Die Auslöser und die Reaktion werden als universell angesehen und höhere oder komplexere Emotionen werden als Mischung der einfachen Basisreaktionen betrachtet. Diese Sichtweise auf Basisemotionen liefert eher die Möglichkeit einer Modellierung, da hier die Ursachen und Auswirkungen einer Basisemotion thematisiert werden.

Diese lassen sich folgendermaßen operationalisieren: Nach [LeD96a] und [Pan98] werden elementare emotionale Reaktionen als automatische Antwort auf einen spezifischen Reiz ausgelöst. Die durch neuronale Verschaltung fest verankerten (semi)automatischen Reaktionssysteme generieren simultan einerseits elementare emotionale Reaktionen, andererseits unmittelbare körperliche Reaktionen wie Blutdruck- und Herzschlaganstieg, hormonelle Veränderungen, Schweißausbrüche etc.

Das Petrinetzmodell in Abb. 5.3 veranschaulicht diesen Sachverhalt: Ein spezifischer Reiz (Stelle **spezifisches Reiz**) löst automatisch körperliche Veränderungen und Basisemotionen (vgl. gleichnamigen Stellen) aus (Transition **auslösen**). Das Netz macht deutlich, dass die Begriffe “elementare emotionale Reaktion“ ([Pan98]) und “automatisches Affektprogramm“ ([LeD98]) mit “Basisemotion“ äquivalent verwendet werden.

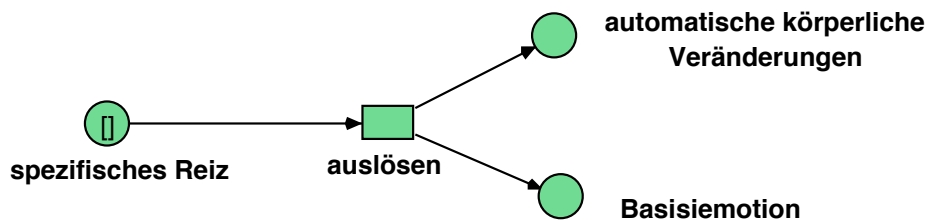


Abbildung 5.3.: Grafische Veranschaulichung der Definition einer Basisemotion.

Eine charakteristische Eigenschaft der Basisemotionen ist somit, dass sie automatisch durch einen spezifischen Stimulus ausgelöst werden, d.h. für ihre Generierung keine deliberativen Prozesse erforderlich sind.

Ein weiteres konstituierendes Merkmal von Basisemotionen ist, dass sie von physiologischen Veränderungen begleitet werden. In einigen Emotionsdefinitionen werden Emotionen mit diesen spezifischen Eigenschaften auch *elementare* oder *primäre Emotionen* genannt (z.B. bei [Laz91], [BDD00], [Slo04], [GM09b]). Jedoch zeigt die Betrachtung der Generierungsmechanismen dieser Emotionsklasse jeweils deutlich, dass sich die verschiedenen Begriffe auf das gleiche emotionale Phänomen beziehen (siehe auch Kapitel 6.1).

So hält Damasio ([BDD00], S. 131-139) die Wahrnehmung bestimmter Eigenschaften eines Stimulus für die unmittelbare Ursache der Generierung primärer, d.h. angeborener emotionaler Reaktionen. Zu diesen Eigenschaften zählen Größe (große Tiere), Bewegung (Reptilien), Geräusche (Knurren) oder bestimmte interne Körpersignale (Schmerz). Sekundäre Emotionen lassen sich dagegen als Ergebnis einer entwicklungstheoretischen Perspektive auf primäre Emotionen verstehen, die auf erlernten und erworbenen dispositionalen Repräsentationen basieren und nicht auf angeborenen Repräsentationen.

Wir benutzen im Folgenden den Begriff *Basisemotionen* als Oberbegriff für solche Emotionen, die durch *spezifische Eigenschaften der Umweltsituation automatisch, d.h. ohne Beteiligung höherer kognitiver Bewertungsprozesse*, ausgelöst werden.

5.2.2. Dimensionale Klassifikation

Bei der dimensionalen Emotionsklassifikation wird angenommen, dass Emotion als Resultat einer mehr oder minder starken Ausprägung in mehreren festgelegten Dimensionen beschrieben werden kann. Diese Beschreibungsdimensionen sollen für alle Emotionen gelten. Es besteht jedoch Uneinigkeit darüber, welche Dimensionen dies sind.

Nach Wundt (1910) lassen sich komplexe emotionale Gefühlszustände durch drei Dimensionen charakterisieren: Angenehmheit vs. Unangenehmheit, Anspannung vs. Entspannung, Erregung vs. Depression. Das Modell von Wundt ist rein introspektiv und besitzt kein empirisches Fundament, nichtsdestotrotz hat es einen starken Einfluss auf viele weitere dimensionale Modelle ausgeübt. Die Dimensionen von Wundt finden sich in vielen späteren Modellen wieder (z.B. bei Traxel und Heide (1961), Ertel (1965)). Osgood et al. (1957) definieren drei wertende Dimensionen für sprachlichen Emotionsausdruck und damit auch für Emotionen: Evaluation (angenehm - unangenehm), Erregung (beruhigend - erregend) und Potenz (stark-schwach). Plutchik (1962, 1982) und Russel (1980) vertreten die Ansicht, dass Emotionen durch die beiden Dimensionen Valenz und Aktivierung bestimmt werden und stellen sie in einem Kreis dar, der durch entsprechende Achsen definiert ist.

Auswertung für die informatische Modellierung

Im Allgemeinen basieren die dimensionalen Emotionsmodelle auf Introspektion und Analyse von Emotionswörtern. Sie eröffnen die Möglichkeit, Emotionen innerhalb von bestimmten Beschreibungsdimensionen einzuordnen, liefern aber keine genaue Beschreibung einer Emotion, z.B. als Funktion der gewählten Beschreibungsparameter. Über die Generierung und die funktionalen Auswirkungen einer Emotion wird in den dimensionalen Modellen ebenfalls keine Aussage getroffen, da diese Fragestellungen i.d.R. nicht in das Spektrum des Forschungsinteresses fallen. Zentner und Scherer (2000) kritisieren ferner, dass sich die meisten solcher Modelle auf eine der Emotionskomponenten (vorwiegend auf den subjektiven Gefühlszustand) beschränken. "Somit wird zwar eine Projektion der unterschiedlichen Emotionskomponenten in einen zweidimensionalen Raum möglich, ohne dass dies erlaubt, konkrete Vorhersagen über individualspezifische Emotionsauslösung und -differenzierung in Bezug auf bestimmte Ereignisse zu machen oder differentielle dynamische Reaktionsmuster vorherzusagen" ([ZS00], S. 153).

Sofern in einem informatischen Modell die Entstehung bzw. Funktionalität der emotionalen Mechanismen beschrieben werden soll, erscheint der Erklärungspotenzial der dimensionalen Emotionsmodelle für die Modellierung wenig hilfreich zu sein.

5. Emotionsdefinition

5.2.3. Emotion: Zustand vs. Prozess

Wie die bisherige Übersicht über die Sichtweisen auf Emotion gezeigt hat, können in der emotionspsychologischen Forschung zwei Betrachtungsperspektiven unterschieden werden: Einerseits wird Emotion als physischer, somatischer und mentaler *Zustand* betrachtet, andererseits wird die Entstehung von Emotion als ein kognitiver *Prozess* der Umweltbewertung (Einschätzungsprozess) beschrieben. Bei der ersten Betrachtungsperspektive wird Emotion als spezieller Zustand bestimmter mentaler und körperlicher Systeme aufgefasst, der für bestimmte Prozesse ursächlich ist und diese auslöst bzw. beeinflusst (z.B. Handlung, Interaktion, Informationsverarbeitung). Die Prozesshaftigkeit von Emotion lässt sich besonders in Bezug auf die Entstehung von Emotionen bzw. von emotionalen Zuständen vergegenwärtigen. Hier sind es vor allem die Einschätzungstheorien (siehe Kapitel 5.4), aber auch neurologische Emotionsmodelle, die Prozessannahmen treffen [Sch99a]. Dabei wird Emotion als ein Prozess der Informationsverarbeitung mit unterschiedlichem Grad (höherer) an kognitiver Beteiligung gesehen. Aktuelle Modelle - inspiriert von Ergebnissen aus den Neurowissenschaften - gehen dabei nicht mehr von einer sequenziellen Verarbeitung sondern von mehreren, parallel ablaufenden Informationsverarbeitungsprozessen aus ([LS87], [Sch93], [Rei00b]).

Die Dichotomie der Emotionsbetrachtung als Zustand bzw. Prozess spiegelt sich in der folgenden Emotionsdefinition wider, die beide Perspektiven auf Emotion zusammenbringt: “Emotion kann demnach als multidimensionaler *Zustand* beschrieben werden, der physiologische Erregung, Gefühle, kognitive *Prozesse* und Verhaltensweisen einschließt, die in Reaktion auf eine Situation auftreten, welche das Individuum als persönlich bedeutsam wahrgenommen hat“ ([Win06], S. 11).

5.3. Emotionspsychologische Arbeitsdefinition

Um die Vielfalt von Themen zu erfassen, die unter dem Begriff “Emotion“ zusammengefasst werden, wird in der Emotionspsychologie zurzeit i.d.R. eine Arbeitsdefinition verwendet, die zumindest eine grobe Einordnung des Begriffs ermöglichen soll. Viele Emotionspsychologen erwarten eine exakte Definition von Emotion erst als Ergebnis emotionspsychologischer Forschung und Theorieentwicklung und stellen zu Beginn ihrer Untersuchungen eine einigermaßen konsensfähige Arbeitsdefinition auf (z.B. [Fri86], [MSR01], [OJ96]). Eine solche Arbeitsdefinition hat durchaus provisorischen und vorläufigen Charakter und spiegelt den aktuellen Erkenntnisstand und den jeweiligen theoretischen Ansatz wider (vgl. [OEM00], S. 11).

Eine Arbeitsdefinition beinhaltet üblicherweise eine beispielhafte Auflistung der “typischen“ Emotionen und die Benennung einiger definierender Merkmale von Emotionen. Diese Merkmale grenzen Emotion zur sog. emotionalen Disposition ab: eine erhöhte Bereitschaft oder Neigung zum Auftreten von bestimmten emotionalen Zuständen, eine Einordnung nach den Dimensionen Qualität, Intensität und Dauer sowie eine Differenzierung in den *Erlebensaspekt*, den *physiologischen Aspekt* und den *Verhaltensaspekt* entsprechend des *Reaktionstrias-Konzepts* (s.u.).

Eine typische emotionstheoretische Arbeitsdefinition

Eine typische Arbeitsdefinition findet man z.B. bei Meyer et al. (2001):

“Emotionen sind zeitlich datierte konkrete einzelne Vorkommnisse (Einzelereignisse) von zum Beispiel Ärger, Angst, Enttäuschung, Eifersucht, Erleichterung, Stolz, Scham, Schuld, Überraschung, Traurigkeit, Freude, Mitleid, Neid sowie weiterer Arten von psychischen Zuständen, die den genannten genügend ähnlich sind. Diese Phänomene haben folgende Merkmale gemeinsam: Sie sind aktuelle psychische Zustände von Personen, haben eine bestimmte Qualität, Intensität und Dauer und sind in der Regel objektgerichtet. Personen, die sich in einem dieser Zustände befinden, haben normalerweise ein charakteristisches Erleben (Erlebensaspekt) und zeigen bestimmte physiologische Veränderungen (physiologischer Aspekt) und Verhaltensweisen (Verhaltensaspekt von Emotionen)“ ([MSR01], S. 3).

Reaktionstrias

Dieser Definitionsversuch macht vom Konzept der sog. *Reaktionstrias* Gebrauch. Das Konzept geht auf Eysenck (1965) und Lang (1968) zurück und wurde später von Lazarus, Averill und Opton (1970) aufgegriffen. Danach lassen sich bei einer Emotion ein *subjektiver Aspekt*, ein *physiologischer Aspekt* und ein *behavioristischer Aspekt* (bzw. *Verhaltensaspekt*) unterscheiden.

Unter dem *subjektiven Aspekt* (oft auch Erlebensaspekt genannt) wird das subjektive Erleben einer Emotion verstanden, d.h. mit der Emotion einhergehende affektive Erfahrungen bzw. Gefühlszustände, Kognitionen usw.

Zu dem *physiologischen Aspekt* rechnet man zentrale und periphere neurophysiologische Veränderungen.

Der *behavioristische Aspekt* umfasst emotionale (Verhaltens-)Reaktionen und wird wiederum oft in zwei Teilaspekte zerlegt. Zum einen unterscheidet man den *expressiven oder Ausdrucksaspekt*, der verschiedene Arten von meist unwillkürlichem Ausdrucksverhalten (sog. *expressives Verhalten*), wie Gestik, Mimik, Stimmschwankungen, Intonation umfasst. Zum anderen wird der *instrumentelle oder Handlungsaspekt*, der zielgerichtete Handlungen als Folge emotionaler Reaktionen umfasst, wie z.B. Angriffsverhalten bei Wut, häufig separat betrachtet.

In Abb. 5.4 wird das Konzept der Reaktionstrias grafisch veranschaulicht. Die Definition der Reaktionstrias lässt keine prozessorientierte bzw. strukturelle Beschreibung der einzelnen Emotionsaspekte zu, da in dieser Definition weder Aussagen über die relative zeitliche Abfolge der einzelnen Reaktionsaspekte (sequenziell, nebenläufig etc.) getroffen werden, noch Abhängigkeiten zwischen den einzelnen definierten Komponenten festgelegt sind. Die Definition sagt ferner nichts darüber aus, ob alle Teilaspekte einer Reaktionstrias vorhanden sein müssen, um von einer Emotion sprechen zu können. Daher ermöglicht diese Definition keine Spezifizierung in Form eines Petrinetzes. Die Abb. 5.4 soll lediglich als grafische Veranschaulichung der strukturellen Zusammensetzung einer Reaktionstrias

5. Emotionsdefinition

betrachtet werden und nicht formal interpretiert werden. Die Aufteilung in aktive und passive Netzelemente wird willkürlich vorgenommen, wobei die einzelnen Emotionsaspekte als Stellen und die zugehörigen Verhaltensweisen als Transitionen dargestellt werden. In der ursprünglichen Definition ist nicht spezifiziert, ob alle Reaktionsaspekte parallel stattfinden (müssen), wie die formale semantische Interpretation des Netzes in Abb. 5.4 implizieren würde.

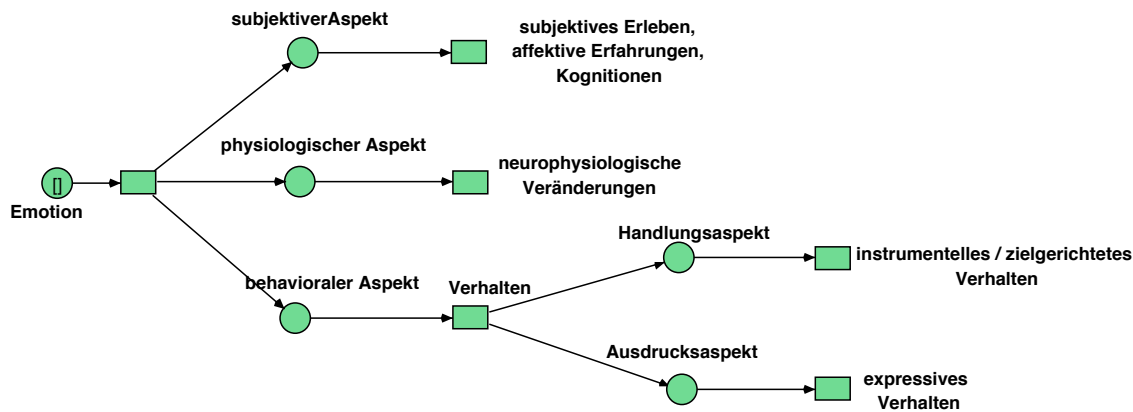


Abbildung 5.4.: Grafische Veranschaulichung der Reaktionstrias-Konzept.

Arbeitsdefinition von Kleinginna und Kleinginna (1981)

Kleinginna und Kleinginna (1981) haben in [KK81] über 100 Emotionsdefinitionen und Aussagen über emotionale Konzepte gesammelt, klassifiziert und analysiert. Die Autoren kamen zu dem Ergebnis, dass wenig Konsistenz zwischen den verschiedenen Definitionen besteht und dass viele Definitionen zu ungenau sind. Sich auf diesen Ergebnissen stützend, schlagen sie eine breite Arbeitsdefinition vor, die alle traditionell bedeutsamen Emotionsaspekte umfasst.

“Emotion is a complex set of interactions among subjective and objective factors, mediated by neural / hormonal systems, which can: (a) give rise to affective experiences such as feelings of arousal, pleasure / displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goal directed, and adaptive“ ([KK81], S. 355).³

³In der Übersetzung von Otto et al. (2000) lautet diese Arbeitsdefinition wie folgt: “Emotion ist ein komplexes Interaktionsgefüge subjektiver und objektiver Faktoren, das von neuronal / hormonalen Systemen vermittelt wird, die (a) affektive Erfahrungen, wie Gefühle der Erregung oder Lust / Unlust, bewirken können; (b) kognitive Prozesse, wie emotional relevante Wahrnehmungseffekte, Bewertungen, Klassifikationsprozesse, hervorrufen können; (c) ausgedehnte physiologische Anpassungen an die erregungsauslösenden Bedingungen in Gang setzen können; (d) zu Verhalten führen können, welches oft expressiv, zielgerichtet und adaptiv ist.“ ([OEM00], S. 15).

5.3. Emotionspsychologische Arbeitsdefinition

Diese Definition erweitert das Konzept der Reaktionstrias um eine neue Komponente, nämlich die *kognitiven Prozesse*, die durch eine Emotion ausgelöst werden, wie Wahrnehmungseffekte, Klassifikationsprozesse, Einschätzungsprozesse usw. Das Netz aus Abb. 5.4 kann um diese kognitive Komponente ergänzt werden. Ferner werden in der Definition von [KK81] die Auswirkungen einer Emotion (kognitive Prozesse, physiologische Anpassungen, Verhalten) deutlich als “Prozesse“ beschrieben. Dies rechtfertigt die Modellierung in Form von Transitionen. Das Modell der Emotionstrias kann diesbezüglich ergänzt und angepasst werden, wie Abb. 5.5 zeigt. Auf die Modellierung von einzelnen Aspekten eines Reaktionstrias in Form von Stellen (vgl. Abb. 5.4) wird nun verzichtet, da diese nur zum Zweck der Veranschaulichung und der konsistenten Begriffsverwendung eingeführt wurden und keine strukturelle Bedeutung besitzen.

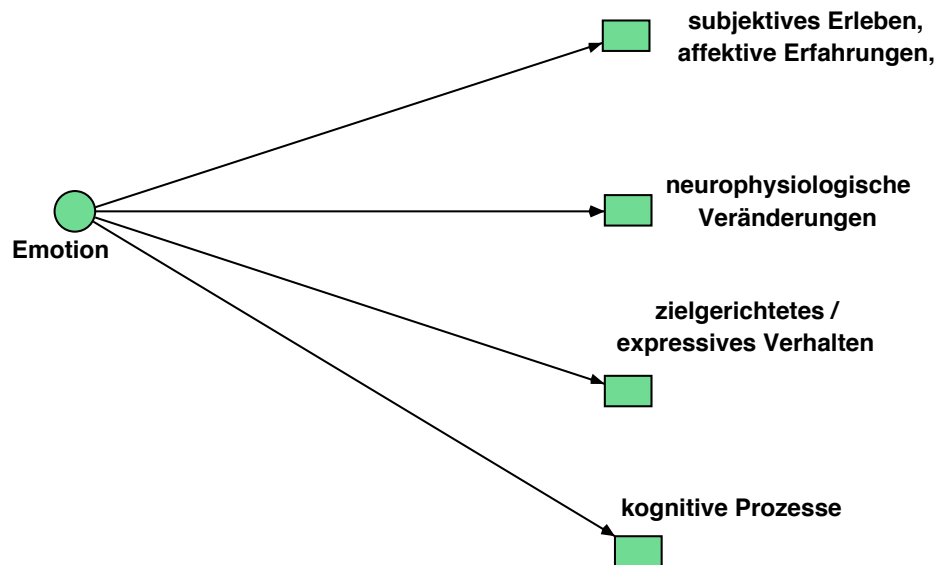


Abbildung 5.5.: Grafische Veranschaulichung der Emotionsdefinition von [KK81].

Die Definitionen in [Man83] bzw. in [PS96] orientieren sich ebenso eng am Reaktionstrias-Konzept. Von diesen Autoren werden Emotionen als subjektives Erleben definiert, “das emotionsspezifische Kognitionen sowie die Wahrnehmung von Ausdruck und periphere physiologische Veränderungen bedingt, oder aus diesen resultiert“ (zit. nach [Win06], S. 11). Diese Beschreibung ist mehrdeutig in Bezug auf die Rolle der Reaktionstrias: Es wird nicht deutlich, ob sie von einer Emotion “bedingt“ ist oder ob die Emotion als Resultat der Emotionstrias entsteht. Die Beschreibung ist an dieser Stelle zyklisch und ermöglicht keine kausale Modellierung.

Auswertung für die informatische Modellierung

Angesichts einer fehlenden allgemeinen Emotionsdefinition ist die Spezifizierung des Forschungsgegenstandes in einer vorläufigen *Arbeitsdefinition* die zurzeit in der emotionspsy-

5. Emotionsdefinition

chologischen Forschung am weitesten verbreitete Definitionsmethode. Die meisten Arbeitsdefinitionen sind jedoch wenig präzise und erscheinen für die informatische Modellierung daher wenig geeignet.

So werden die Interaktion und der relative zeitliche Ablauf der einzelnen Komponenten in den vorgestellten Definitionen nicht näher spezifiziert. Die wechselseitigen kausalen oder strukturellen Abhängigkeiten werden nicht beschrieben und die Prozesse der Emotionsentstehung (Emotionsgenerierung) werden gar nicht betrachtet. Auch die Auswirkung der Emotionen auf das Gesamtverhalten, d.h. die *funktionale Rolle* der Emotionen, wird nur allgemein und nicht explizit definiert. Eine solche Emotionsdefinition ist nicht direkt operationalisierbar und bedarf zusätzlicher Interpretation bzw. der Einführung ergänzender Konzepte, um in ein informatisches Modell umgewandelt zu werden.

Ferner wird keine Aussage darüber gemacht, ob alle definierten Komponenten vorhanden sein müssen, um von einer Emotion sprechen zu können. Z. B. ist unklar, welche Rolle der subjektive Aspekt des emotionalen Erlebens spielt und ob er bei einer Emotion unbedingt beteiligt sein muss.

Einige Autoren definieren Emotionen durch das Anführen von Beispielen (vgl. exemplarisch [MSR01]). Dies ermöglicht keine allgemeine Spezifikation des Konzeptes und gibt keine Auskunft über die funktionale Rolle der Emotionen in einem kognitiven System. [MSR01] räumt auch selber ein, dass "die Abgrenzung von Emotion durch die Methode des Angebens von Beispielen unscharf bleibt" ([MSR01], S.3).

Die meisten Emotionstheorien behandeln nur eine der definierten Komponenten. So wird der Verhaltensaspekt nur in behavioristischen Theorien, seltener in Einschätzungstheorien miteingefasst (z.B. [Fri86]). Der physiologische Emotionsaspekt wird von einigen Autoren bewusst vernachlässigt (vgl. [Slo03]). Eine eingehende Analyse in [ZS00] ergab, dass die meisten aktuellen Emotionstheorien nur Teilaspekte der oben dargestellten Definition betrachten und andere wesentliche Inhalte nicht zu erklären versuchen. In diesem Sinne kann die vorgestellte Definition dazu dienen, das Forschungsfeld zu systematisieren.

Das Konzept der Emotionstrias, das sehr oft in die Arbeitsdefinitionen mit aufgenommen wird, wird in der Literatur wiederum hinsichtlich mehrerer Gesichtspunkte kritisiert. Insbesondere werde bei der Emotionstrias nicht spezifiziert, ob alle drei Komponenten als notwendige definierende Merkmale einer Emotion verstanden werden. Außerdem existiere nur eine mäßige Korrelation zwischen den messbaren Indikatoren der einzelnen Komponenten (Ricco und Silvestri 1973). Ferner komme es zu vielfältigen Verzerrungen zwischen den Komponenten und ihren messbaren Output-Parametern. Das allgemeine Fazit lautet, dass die Emotionstrias nur bedingt dazu geeignet sei, Emotionen zu definieren, und dass sie auch empirisch nicht ausreichend belegt sei [Rot99].

Die kognitive Komponente wird oft von den eigentlichen emotionalen Verhaltensbereichen (physiologisch, erlebnismäßig, expressiv) abgehoben und als wesentlich für die Entstehung und Erklärung von Emotionen angesehen (z.B. [MSR01], [Sch99a]). Diese Betonung des Kognitiven bringt mit sich, dass die Verhaltenskomponente als Folgeerscheinung oft aus dem emotionalen Geschehen ausgegrenzt wird. Im folgenden Abschnitt werden Emotionsdefinitionen vorgestellt, die die kognitive Bewertung (*appraisal*) als einen für die Entstehung und Funktionalität einer Emotion zentralen Prozess ansehen.

5.4. Emotion als Anpassungsmechanismus: Evolutionspsychologische Definitionen

Als ein wesentlicher Nachteil einer klassischen emotionspsychologischen Emotionsdefinition wurde die fehlende Spezifizierung der Emotionsentstehungsprozesse (Emotionsgenerierung) und der Auswirkungen der Emotionen auf das Gesamtverhalten, d.h. die *funktionale Rolle* der Emotionen, identifiziert.

Diese Prozesse, die oft unter dem Begriff *Emotionsprozess* [Fri86] zusammengefasst werden, bilden den zentralen Untersuchungsgegenstand der Einschätzungstheorien.

Oft werden evolutionspsychologische Emotionstheorien als “Funktionsmodelle der Emotion“ [McD69] bezeichnet, da sie Emotion als Reaktion auf Ereignisse im Kontext der Zielverfolgung betrachten und insbesondere die Implikationen von Emotionen für Motivation und Handlungssteuerung zum Forschungsgegenstand machen. Emotion wird in diesem Sinne definiert als “eine ererbte oder angeborene psychophysische Disposition, welche [...] befähigt, bestimmte Gegenstände wahrzunehmen und ihnen Aufmerksamkeit zu schenken, (dadurch) eine emotionale Erregung von ganz bestimmter Qualität zu erleben und daraufhin in einer bestimmten Weise zu handeln oder wenigstens den Impuls zu solch einer Handlung zu erleben“ ([McD69], S. 24).

Allgemein wird Emotion von vielen Anhängern des evolutionspsychologischen Standpunktes als ein flexibler Anpassungsmechanismus zur Aufrechterhaltung des optimalen Funktionierens des Organismus gesehen, der auf einer höheren Evolutionsstufe als andere, weniger flexible adaptive Mechanismen, wie Reflexe und physiologische Triebe, anzusiedeln ist (z.B. [CSC94], [Sch84], [SE85], [Laz91], [LAO70], [Rol05], [Ekm92], [Fri07]).

Die Flexibilität von Emotionen wird durch die isolierte Betrachtung der Verhaltensreaktion und des Auslöseereignisses erreicht: An die Stelle von unflexiblen Stimulus-Reaktion-Mustern bei Reflexen tritt im Falle von Emotion ein kognitiver *Einschätzungs- bzw. Bewertungsprozess*, der die Relevanz des Stimulusereignisses für das individuelle Wohlergehen bewertet. Dieser Bewertungsprozess wird von der Generierung von Reaktionen bzw. Reaktionstendenzen gefolgt, die zum Ziel haben, die Individuum-Umwelt-Beziehung zu verändern ([Fri86], [Fri07]).

Die Klassifikation in Basisemotionen und zusammengesetzte Emotionen spezifiziert keine Relation zwischen Emotionen und deren primären Entstehungsbedingungen. Im Gegensatz dazu wird bei den *kognitiven Einschätzungstheorien* genau diese Relation als zentral angesehen. Einschätzungstheorien sehen Emotion als Ergebnis der kognitiven Einschätzung der Stimulusrelevanz für das individuelle Wohlergehen.

Emotionsdefinition von Lazarus

Um in einer komplexen Umwelt die für ein spezifisches Anpassungsproblem jeweils entscheidende “Kernaussage“ zu entdecken, muss die Umweltsituation fortlaufend hinsichtlich ihrer Bedeutung für das eigene Wohlergehen bewertet werden. Ein Auslöser von Emotion ist nach einschätzungstheoretischer Sichtweise also nicht ein konkretes Ereignis, sondern die

5. Emotionsdefinition

abstrakte Bedeutung, die diesem Ereignis durch den Einschätzungsprozess zugeschrieben wird.

Die Arbeitsdefinition von Lazarus besagt, dass Emotionen organisierte psychophysiologische Reaktionen auf Veränderung der Beziehungen (der Person) zu ihrer Umwelt darstellen, wobei sich der Begriff "Veränderung" konkret auf die Änderung der Überzeugungen hinsichtlich der Bedeutung der Person-Umwelt-Beziehungen für das persönliche Wohlergehen bezieht.

"Die Qualität (z.B. Ärger vs. Furcht) und die Intensität (die Stärke der Mobilisierung oder der motorisch-physiologischen Veränderung) der emotionalen Reaktion hängt ab von subjektiven Bewertungen - ich nenne sie kognitive Bewertungen - dieses Wissens in Bezug darauf, wie es auf kurze und lange Sicht um unsere Ziele steht, und von der auf diese Beziehung gerichteten Handlungstendenz. [...] Emotionen sind im Grunde organisierte kognitiv-motivational-relationale Konfigurationen, deren Status sich mit Änderungen der Person-Umwelt-Beziehung, wie sie wahrgenommen und bewertet wird, ändert." ([Laz91], S. 38).

Die Sichtweise von Lazarus wird in einem Petrinetz in Abb. 5.6 zusammengefasst. Eine emotionale Reaktion (Stelle **emotionale Reaktion**) wird durch die Änderung der Überzeugungen hinsichtlich der Bedeutung der Person-Umwelt-Beziehungen ausgelöst, die in Bezug darauf bewertet werden (Transition **kognitiven Bewertungen**), welche Auswirkungen sie für das persönliche Wohlergehen bzw. die Ziele eines Organismus (Stelle **Ziele**) haben.

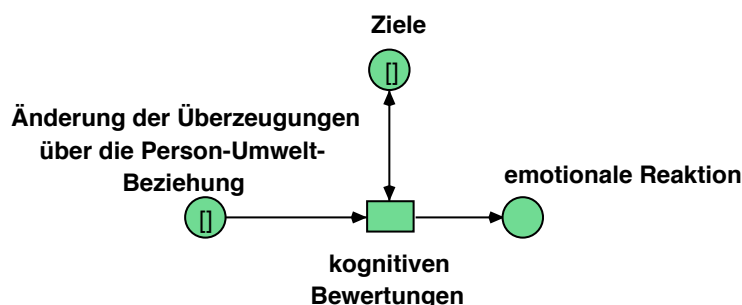


Abbildung 5.6.: Grafische Veranschaulichung der Emotionsdefinition von Lazarus (nach [Laz91]).

Emotionsdefinition von Scherer

Scherer ([Sch93], S. 4) schlägt folgende Arbeitsdefinition vor:

"Emotion ist eine Episode zeitlicher Synchronisation aller bedeutender Subsysteme des Organismus, die fünf Komponenten bilden (Kognition, physiologische Regulation, Motivation, motorischer Ausdruck [motor expression] und Monitoring / Gefühl), und die eine Antwort auf die Bewertung eines externen oder

5.4. Emotion als Anpassungsmechanismus: Evolutionspsychologische Definitionen

internalen Reizereignisses als bedeutsam für die zentralen Bedürfnisse und Ziele des Organismus darstellt.“ (zit. nach [OEM00], S. 16).

Das Netz in Abb. 5.7 veranschaulicht diese Emotionsdefinition. Ein “externales oder internes Reizereignis“ (Stelle **externales / internes Reizereignis**) wird in Bezug auf “zentrale Bedürfnisse und Ziele des Organismus“ (Stelle **Relevanz für die zentralen Bedürfnisse und Ziele**) bewertet (Transition **Bewertung**). Dadurch wird eine Emotion (Stelle **emotionale Episode**) hervorgerufen, die mit einer “Episode zeitlicher Synchronisation aller bedeutender Subsysteme des Organismus“ einhergeht (Transition **Synchronisation der Subsysteme**).

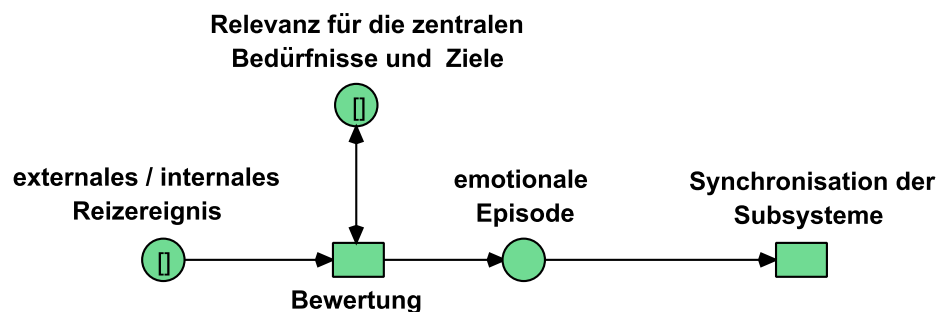


Abbildung 5.7.: Grafische Veranschaulichung der Emotionsdefinition von Scherer (nach [Sch93]).

Emotionsdefinition von Frijda

Oatley und Jenkins haben Frijdas Definitionselemente zu der folgenden Arbeitsdefinition zusammengestellt ([OJ96], S. 96):

“(1) Eine Emotion wird üblicherweise dadurch verursacht, daß eine Person - bewußt oder unbewußt - ein Ereignis als bedeutsam für ein wichtiges Anliegen (ein Ziel) bewertet [...] (2) Der Kern einer Emotion sind Handlungsbereitschaft (readiness to act) und das Nahelegen (prompting) von Handlungsplänen; eine Emotion gibt einer oder wenigen Handlungen Vorrang, denen sie Dringlichkeit verleiht. So kann sie andere mentale Prozesse oder Handlungen unterbinden oder mit ihnen konkurrieren [...] (3) Eine Emotion wird gewöhnlicherweise als ein bestimmter mentaler Zustand erlebt, der manchmal von körperlichen Veränderungen, Ausdruckserscheinungen und Handlungen begleitet oder gefolgt wird.“ (zit. nach [OEM00], S. 16).

Die Emotionsdefinition von Frijda in der Formulierung von [OJ96] wird in Abb. 5.8 als ein Petrinetz-Modell dargestellt. Ein Ereignis (Stelle **Ereignis**) wird in Bezug auf ein wichtiges Anliegen (Stelle **Anliegen (Ziel)**) bewertet (Transition **Bewertung**). Dadurch wird eine Emotion hervorgerufen, die in der Formulierung von [OJ96] “gewöhnlicherweise als ein bestimmter mentaler Zustand erlebt“ wird. Um dieser Formulierung besser Rechnung zu tragen und aufgrund der grundlegenden Dichotomie in der Betrachtung des Emotionsbegriffs

5. Emotionsdefinition

(zustandsorientiert vs. prozessorientiert, vgl. Kapitel 5.2.3), bezeichnen wir die generierte Emotion als “emotionalen Zustand“ (Stelle **emotionaler Zustand**). Bedingt durch diesen emotionalen Zustand, werden Handlungsbereitschaften verändert bzw. Handlungspläne angepasst (Transition **Veränderung der Handlungsbereitschaft**).

Aus Punkt (2) der Definition wird deutlich, dass Handlungen durch die Veränderung der Handlungsbereitschaft ausgelöst oder unterbunden werden können. Punkt (3) spezifiziert ferner die Auswirkungen einer Emotion (körperliche Veränderungen, Ausdruckserscheinungen und Handlungen), die sinngemäß der Reaktionstrias (vgl. Kapitel 5.3) entsprechen.

Um die Netzmodelle verschiedener Emotionsdefinitionen miteinander vergleichen zu können, muss auf eine konsistente Verwendung der Terminologie geachtet werden. Deshalb wählen wir an dieser Stelle die gleichen Bezeichnungen für die Auswirkungen der Emotion, die bei der Definition der Reaktionstrias benutzt wurden. Die “Ausdruckserscheinungen und Handlungen“, die nach der Definition von [OJ96] als Resultat einer veränderten Handlungsbereitschaft einsetzen (können), entsprechen offensichtlich den beiden Verhaltensaspekten, die in der Definition von der Reaktionstrias unterschieden wurden, nämlich dem “expressiven oder Ausdrucksaspekt“ (sog. *expressives Verhalten*) und dem *instrumentellen oder Handlungsaspekt*, d.h. zielgerichteten Handlungen. Wir modellieren diese beiden Verhaltensaspekte durch die Stellen **zielgerichtetes Verhalten** und **expressives Verhalten**, die beim Schalten der Transition **Veränderung der Handlungsbereitschaft** markiert werden.

Die “körperlichen Veränderungen“ [OJ96] werden zur Erhaltung der Terminologiekonsistenz ebenfalls als **physiologische Veränderungen** bezeichnet. Der in der Definition von [OJ96] auch erwähnte Erlebensaspekt einer Emotion wird entsprechend der Begriffsverwendung in der Definition der Reaktionstrias mit der Transition **subjektives Erleben** modelliert.

Im Punkt (3) wird bezüglich des chronologischen Auftretens von Emotionsauswirkungen eine mehrdeutige Formulierung verwendet: Demnach wird eine Emotion “manchmal“ von ihnen “begleitet oder gefolgt“. Diese Unbestimmtheit kann mit dem Petrinetz-Formalismus nicht formuliert werden. Das Netz in Abb. 5.8 stellt den Fall dar, dass eine Emotion von der Reaktionstrias (körperliche Veränderungen, expressives Verhalten und Handlungen) *gefolgt* wird. Dadurch soll der kausale Zusammenhang zwischen Emotion und Reaktionstrias verdeutlicht werden: Emotion wird als Auslöser für die Reaktionstrias gesehen, die ihrerseits als Reaktion des Organismus auf eine Emotion (emotionale Reaktionen) verstanden werden muss.

Zusammenfassung der Modelle

Wenn man die Netzmodelle der drei vorgestellten einschätzungstheoretischen Definitionen vergleicht, fällt die strukturelle Ähnlichkeit der Modelle - zumindest in Bezug auf die Generierung einer Emotion - sofort auf. In jedem Modell wird Emotion als Ergebnis der (kognitiven) Bewertung eines (internen oder externen) Ereignisses in Bezug auf Bedürfnisse und Ziele (bzw. Anliegen) eines Organismus gesehen. Lazarus spricht zwar von der “Änderung der Überzeugungen über die Person-Umwelt-Beziehung“ als einer Auslösebedingung für den Bewertungsprozess, eine solche Änderung der Überzeugungen setzt aber das Hinzukommen von neuen Fakten über die Umwelt oder das Selbst voraus. Diese Fakten können entweder durch die Wahrnehmung von neuen Informationen aus der Umwelt (externe Ereignisse) oder

5.4. Emotion als Anpassungsmechanismus: Evolutionspsychologische Definitionen

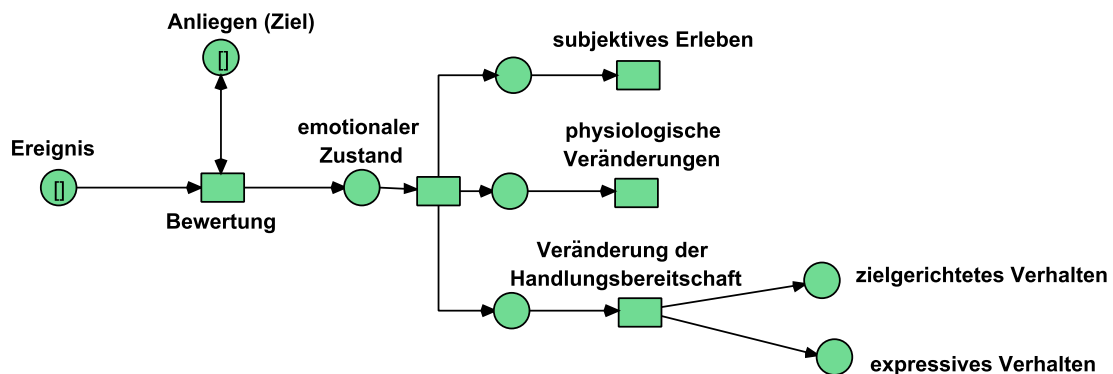


Abbildung 5.8.: Grafische Veranschaulichung der Emotionsdefinition von Frijda (nach [OJ96]).

durch Schlussfolgern angesichts der aktuellen Wissensbasis (interne Ereignisse) gewonnen werden. Folglich setzt auch die Definition von Lazarus die Existenz eines (internen oder externen) Ereignisses bzw. eines Stimulus als Ursache und Auslösebedingung für den Bewertungsprozess voraus.

Damit zeigt sich, dass die drei betrachteten Definitionen die gleichen konstituierenden Elemente für eine Emotion postulieren. Das Netz in Abb. 5.9 zeigt eine allgemeine Struktur der Emotionsgenerierung, wie sie in den einschätzungstheoretischen Definitionen beschrieben wird. Um allen drei Definitionen gerecht zu werden, bezeichnen wir das (interne oder externe) Auslöseereignis für eine Emotion allgemein als Stimulus (Stelle **Stimulus**). Damit wird jede Art von neuen Erkenntnissen über die externe und interne Situation erfasst, die z.B. durch Wahrnehmung neuer Fakten oder durch Schlussfolgerungen aufgrund des aktuellen Wissensstandes gewonnen werden. Die neue Erkenntnis bringt möglicherweise eine Veränderung der aktuellen Person-Umwelt-Beziehung mit sich (Lazarus) und muss deshalb in Bezug auf die aktuellen Ziele (Frijda, Lazarus) und Bedürfnisse (Scherer) (Stelle **Ziele / Bedürfnisse**) bewertet werden (Transition **Bewertung**). Als Ergebnis der kognitiven Bewertungsprozesse entsteht eine Emotion (Stelle **Emotion**).

Auf der allgemeinen Ebene schreiben alle Definitionen der Emotion dieselbe Funktion zu, nämlich die Unterstützung der Anpassung an die (ggf. veränderten) Umweltbedingungen durch die Steuerung von Verhalten. Die konkreten Mechanismen der Verhaltenssteuerung unterscheiden sich jedoch in den verschiedenen Definitionen: So sehen Lazarus und Frijda die Generierung einer Handlungstendenz (Lazarus) bzw. die Veränderung der Handlungsbereitschaft (Frijda) als unmittelbares Ergebnis und Kernelement einer Emotion. Scherer spricht dagegen von einer Episode zeitlicher Synchronisation aller bedeutender Subsysteme des Organismus; der Bezug zum Verhalten wird in seiner Emotionsdefinition vorerst nicht näher spezifiziert. Die Funktion der Emotion in Bezug auf das Verhalten kann am besten im Kontext der jeweiligen Einschätzungstheorie erklärt werden. Wir werden uns mit der Beschreibung und Modellierung der funktionalen Rolle der Emotion in einem Organismus weiter unten im Zusammenhang mit den Einschätzungstheorien von Scherer, Lazarus und Frijda beschäftigen (siehe Kapitel 6).

5. Emotionsdefinition

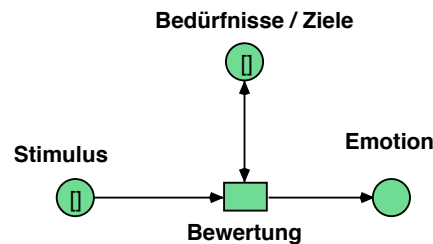


Abbildung 5.9.: Emotionsgenerierung: Zusammenfassung der einschätzungstheoretischen Definitionen.

Auswertung für die informatische Modellierung Einschätzungstheorien erklären Emotion als Ergebnis der kognitiven Bewertung der Umwelt mit Bezug auf das individuelle Wohlergehen (Ziele, Pläne, Überzeugungen) eines Individuums. Die meisten einschätzungstheoretischen Emotionsdefinitionen treffen eine Aussage über die Entstehung und Auswirkungen einer Emotion. Daher sind sie vom Grundsatz her operationalisierbar und für die Umsetzung in informatischen Emotionsmodellen gut geeignet. Diese Tatsache spiegelt sich auch in der Anzahl der bisherigen Umsetzungen wider: Einschätzungstheorien werden von allen sozialwissenschaftlichen Emotionstheorien am häufigsten implementiert.

Außer der guten Operationalisierbarkeit liegen weitere Vorteile der Einschätzungstheorien in Bezug auf informatische Modellierung in der funktional orientierten Betrachtung der Emotion sowie in der hohen Kompatibilität mit der BDI-Agentenarchitektur. Die Modellierung und der Vergleich der Emotionsdefinitionen verschiedener Forscher hat deren strukturelle Ähnlichkeit hinsichtlich der Emotionsgenerierung erkennen lassen: Alle betrachteten Definitionen sehen die Bewertung der Stimulusrelevanz für die Ziele / Anliegen des Akteurs / Agenten als eine notwendige Bedingung der Emotionsgenerierung (vgl. Einschätzung der Anliegerrelevanz bei Frijda, der Zielrelevanz bei Lazarus bzw. den *goal significance check* bei Scherer). Wird das Ereignis für die Ziele des Agenten als nicht relevant bewertet, entsteht keine Emotion: “Wir reagieren bei unwichtigen Dingen nicht mit einer Emotion, sondern bei Werten und Zielen, zu denen wir eine starke Verpflichtung eingegangen sind ([Laz91], S. 819f; vgl. auch [Fri07]). Das Konzept der Ziele bzw. Anliegen in den betrachteten einschätzungstheoretischen Definitionen weist eine bemerkenswerte Ähnlichkeit mit den Desires eines BDI-Agenten auf. Damit lässt sich die Möglichkeit einer einfachen Einbindung des Emotionsgenerierungsprozesses in die BDI-Architektur erkennen, die weiter unten untersucht wird (siehe Kapitel 6.1.2).

5.5. Eine neurobiologische Emotionsdefinition

Die aus den Neurowissenschaften stammende Emotionstaxonomie von Damasio ist insofern besonders interessant, als sie den Emotionsprozess sehr differenziert behandelt, andere Emotionstaxonomien (z.B. [Laz91]) teilweise einschließt und auf experimentell belegten

5.5. Eine neurobiologische Emotionsdefinition

neurobiologischen Forschungsergebnissen aufbaut. Damasio bezieht *soziale* Emotionen in seine Definition mit ein. Damit kommt er dem Wunsch der Informatik, integrative soziologisch-psychologische Modelle zu bauen, um z.B. das Mikro- und Makro-Ebene in einem Modell betrachten und untersuchen zu können, entgegen (s. Kapitel 2.2). Ferner ist Damasios Sichtweise mit den oben dargestellten Annahmen der Evolutionspsychologen weitgehend kompatibel und lässt evtl. eine Integration zu. Auf Damasios Emotionstaxonomie wird aus diesen Gründen genauer eingegangen.

Damasio postuliert die regulatorische Funktion der Emotionen: Wenn ein Organismus bestimmte Veränderungen in seiner Umwelt feststellt, werden durch Emotionen Reaktionen ausgelöst, die optimale Selbsterhaltungsbedingungen und effizientes Funktionieren herbeiführen sollen. Sowohl sichtbaren als auch unsichtbaren emotionalen Reaktionen - zu solchen werden nach Damasio alle internen emotionalen Zustände von einfachen Hungergefühlen bis zu komplizierten Emotionen wie Furcht, Ekel, Glück, Leid, Sympathie oder Scham gezählt - sowie die untergeordneten Prozesse, die für die Entstehung der Emotionen sorgen, sind nach Damasio ein Teil des grundlegenden Anpassungsmechanismus der "life regulation" [Dam94]. Demnach stimmt Damasios Auffassung der Emotionsfunktion mit der Sichtweise der Anhänger eines evolutionspsychologischen Standpunktes überein, die Emotion als einen flexiblen Anpassungsmechanismus zur Aufrechterhaltung des optimalen Funktionierens eines Organismus (vgl. z.B. [Laz91]) sehen.

Damasio untersucht die zentralen und peripheren neurobiologischen Mechanismen, die für die Auslösung und Ausführung emotionaler Reaktionen verantwortlich sind. Emotionen sind nach Damasio das Ergebnis einfacher Reaktionen, die sich im Laufe der Evolution als die sichersten für das Überleben erwiesen haben. Alle lebenden Organismen sind mit einem Reaktionsapparat (sog. Homeostase) ausgestattet, der automatisch, ohne Beteiligung der Kognition, die wichtigsten Probleme der Selbsterhaltung (Energiequellen finden und verwerten, biochemisches Gleichgewicht aufrechterhalten, sich vor Verletzungen schützen usw.) bewältigt. Dieser Lebenserhaltungsapparat setzt sich zusammen aus folgenden, hierarchisch geordneten Mechanismen:

- metabolischen Prozessen (endokrine und hormonelle biochemische Regulation),
- grundlegenden Reflexen (z.B. Erschrecken),
- mit Lust oder Schmerz assoziierten Verhaltensweisen (Zusammenzucken, verzerrter Gesichtsausdruck, Schützen des verletzten Körperteils beim Schmerz),
- einer Menge von Trieben und Motivationen (Hunger, Durst, Neugier, Spiel, Sex) und schließlich, an der Spitze der automatischen Lebensregulation,
- den eigentlichen Emotionen (emotion-proper) - von Freude, Trauer und Angst bis Stolz, Scham und Sympathie.

Nach Damasios *Nesting Principle* liegen evolutionär ältere und einfachere Regulationsmechanismen evolutionär jüngeren und komplexeren zugrunde. In ihrem Zusammenspiel dienen diese Mechanismen der inneren Regulation und der Selbsterhaltung des Organismus, wobei sie automatisch und spontan eine vorprogrammierte stereotype

5. Emotionsdefinition

Reaktion auslösen. Auch die “eentlichen“ Emotionen (*emotion-proper*) wie Ekel, Furcht, Glück, Leid, Sympathie und Scham dienen direkt oder indirekt der Selbsterhaltung, sei es durch Gefahrenvermeidung, Chancenausnutzung oder Förderung sozialer Beziehungen.

“Without exception, all of these phenomena are related to adaptive adjustment in body state and eventually lead to the changes in the brain mapping of body states, which form the basis for feelings. The nesting of the simple within the complex ensures that the regulatory purpose remains present in the higher echelons of the chain. While the purpose remains constant, complexity varies.“([Dam03], S. 43).

Aus dieser Perspektive gibt Damasio die folgende allgemeine Definition der eigentlichen Emotionen (*emotion proper*), die Gegenstand der meisten Emotionstheorien sind:

Eine Emotion ist eine komplexe Sammlung von automatischen, chemischen und neuronalen Reaktionen, die im Gehirn durch die Gegenwart eines wahrgenommenen oder mental vorgestellten emotional repräsentativen Stimulus (*emotionally competent stimulus - ECS*) ausgelöst wird. Das Gehirn ist genetisch vorprogrammiert, bestimmten ECSs mit entsprechenden Reaktionen zu begegnen, Reaktionen auf andere ECSs sind erlernt. Das unmittelbare Ergebnis dieser Reaktionen ist eine temporäre Veränderung der körperlichen Zustände und, daraus folgend, der Abbildungsmuster dieser Zustände in den Gehirnstrukturen (body-mapping). Im Endeffekt wird der Organismus in einen Zustand versetzt, der für sein Überleben und Wohlergehen förderlich ist - den *emotionalen Zustand*. ([Dam03], S. 53, Hervorh. im Orig.).

Diese Definition liefert eine detailliertere Beschreibung der Entstehung einer Emotion im Hinblick auf die zugrundeliegenden neurobiologischen Prozesse. Sie lässt sich netztechnisch folgendermaßen formalisieren (siehe Abb. 5.10): Die Präsenz eines emotional repräsentativen Stimulus, des sog. ECS (Stelle ECS), löst automatische, chemische und neuronale Reaktionen aus (Transition *auslösen*), die ihrerseits zu einer temporären Veränderung der körperlichen Zustände führen (Transition *Veränderung der körperlichen Zustände*). Dadurch wird das System in einen emotionalen Zustand versetzt (Stelle *emotionaler Zustand*).

Damasio unterteilt die eigentlichen Emotionen (*emotion-proper*) in drei Klassen: *Hintergrundemotionen*, *primäre Emotionen* und *soziale Emotionen*.

Eine *Hintergrundemotion* sieht er als ein unvorhersagbares Ergebnis des Zusammenspiels mehrerer niedriger regulatorischer Prozesse (metabolische Prozesse, Reflexe, momentan aktivierte Triebe, Umweltsituation usw.). Die jeweils aktuelle Kombination aller dieser Prozesse macht unser momentanes Allgemeinbefinden (“state of being“) aus, d.h. sie bestimmt, ob wir uns insgesamt wohl fühlen oder nicht.

Die *primären Emotionen* sind bei Damasio die traditionell von den Emotionstheoretikern angenommenen Basisemotionen (vgl. Kapitel 5.2.1) wie Furcht, Ärger, Ekel,

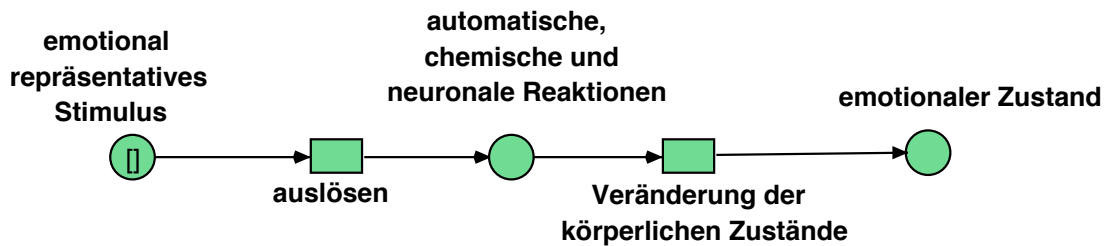


Abbildung 5.10.: Emotionsdefinition nach [Dam03].

Überraschung, Traurigkeit und Glück. Sie sind interkulturell nachweisbar und können auch bei bestimmten Tierarten beobachtet werden.

Zu den *sozialen Emotionen* gehören Sympathie, Verlegenheit, Scham, Schuld, Stolz, Neid, Eifersucht, Dankbarkeit, Bewunderung, Empörung und Verachtung. Auch hier lässt sich das *Nesting Principle* anwenden: Die ganze Palette der regulatorischen Mechanismen kann zusammen mit den Elementen der primären Emotionen als Teilkomponenten der sozialen Emotionen identifiziert werden. Damasio hält soziale Emotionen ebenso wie alle niedrigeren Emotionsformen für angeborene automatische Anpassungsmechanismen.

Auswertung für die informatische Modellierung Wie die Anhänger des evolutionspsychologischen Standpunktes hält Damasio Emotionen für angeborene automatische Anpassungsmechanismen. Da er jedoch Emotionen aus der neurowissenschaftlichen Perspektive betrachtet, liefert seine Definition eine detailliertere Beschreibung des Entstehungsmechanismus einer Emotion im Hinblick auf die zugrundeliegenden neurobiologischen Prozesse. Insofern kann sie als eine Verfeinerung von einschätzungstheoretischen Definitionen mit Bezug auf bestimmte Schritte des Emotionsgenerierungsprozesses gesehen werden.

Ein weiterer Vorteil einer neurobiologischen Definition ist die experimentell nachgewiesene Validität der postulierten Mechanismen der Emotionsgenerierung. Nachteilig ist dagegen im Hinblick auf die informatische Modellierung der fehlende Bezug zu den wesentlichen Konzepten der Agentenorientierung bzw. der Agentenarchitekturen: So wird z.B. die Beziehung zwischen emotionalen Prozessen und kognitiven, wie Wahrnehmung, Informationsverarbeitung, Planen etc. nicht beschrieben.

Im Hinblick auf die Abbildbarkeit in einem technischen System erscheint die Emotionsdefinition von Damasio besonders zur Umsetzung in informatischen Anwendungen geeignet, einerseits aufgrund ihrer Differenziertheit und Universalität, andererseits aufgrund ihrer guten Operationalisierbarkeit. So hängen die Hintergrundemotionen von der aktuellen Kombination von metabolischen Prozessen, Reflexen, aktuell relevanten Trieben, der Umweltsituation usw. ab und machen das momentane Allgemeinbefinden (“state of being“) des (natürlichen) Agenten aus. Die primären Emotionen werden von den Hintergrundemotionen bestimmt sowie durch die Kombination der gleichen niedrigeren Regulationsprozesse beschrieben, die für die Hintergrundemotionen verantwortlich sind. Entsprechend werden

5. Emotionsdefinition

soziale Emotionen als Funktion der ganzen Palette der regulatorischen Mechanismen zusammen mit den Elementen der primären Emotionen beschrieben. Damit ist Damasio's Begriffsbestimmung nicht nur sehr präzise, sondern sie erlaubt auch den Aufbau einer funktionalen Zusammenhangbeziehung zur Emotionsdefinition. Durch niedrigere Regulationsmechanismen (metabolische Prozesse, Reflexe, mit Schmerz und Lust verbundene Verhaltensweisen und Motivationen) werden drei Emotionsklassen auf drei hierarchisch aufeinander aufbauenden Ebenen definiert (vgl. Damasio's *Nestling Principle*) und sind weitgehend für eine Implementation etwa in einer BDI-Agentenarchitektur geeignet.

Im nächsten Abschnitt wird die Möglichkeit einer Integration von Damasio's neurobiologischer Definition in die evolutionspsychologischen Definitionen der Einschätzungstheoretiker beschrieben und diskutiert.

5.6. Zusammenfassung und (Neu-)Definition des Emotionsbegriff für technische Systeme

Die Analyse der Emotionsdefinitionen verschiedener Emotionstheoretiker hat gezeigt, dass die Definition des Forschungsgegenstandes ein grundlegendes Problem der zeitgenössischen Emotionsforschung darstellt. Bezüglich der Abgrenzung der Emotion zu verwandten Phänomenen, der Identifikation verschiedener Emotionsklassen und -taxonomien sowie der postulierten Funktionen liegen die betrachteten Emotionsdefinitionen teilweise sehr weit auseinander.

Übersicht über die betrachteten Emotionsdefinitionen

Unter dem Begriff *Emotion* werden sehr heterogene Phänomene zusammengefasst, von psychophysiologischen Reaktionen bis zu kognitiven Bewertungsprozessen, wobei bei manchen Forschern auch Gefühle, Stimmungen, Verhaltensdispositionen, Wünsche oder Motive unter den Begriff "Emotionen" fallen (z.B. [OEM00], [MSR01]). Zurzeit besteht kein Konsens bezüglich einer qualitativen Abgrenzung der Emotion zu anderen, verwandten Phänomenen: "It [Emotion -Anm. d. Verf.] has dozens of definitions in the psychological literature, because different authors attend to different subsets of emotional phenomena" ([SS02], S. 2).

Die Abgrenzung der Emotionen zu *Reflexen* und *physiologischen Trieben* einerseits, zu *Stimmung*, *Affekt* und *Gefühl* andererseits, ist gleichermaßen problematisch.

In der emotionspsychologischen Forschung wird üblicherweise keine strukturelle bzw. qualitative Unterscheidung zwischen den Begriffen *Affekt*, *Stimmung* und *Emotion* gemacht. Die vorgeschlagenen Unterscheidungskriterien sind meist quantitativer Art.

Reflexe und *physiologische Triebe* werden aus der evolutionspsychologischen Perspektive als Vorstufen der Emotion gesehen. Alle drei Mechanismen dienen der optimalen Anpassung an sich verändernde Umweltbedingungen und unterscheiden sich durch zunehmende Flexibilisierung der Verhaltensreaktionen auf spezifische Reize: Reflexe stellen starre Reiz-Reaktions-Muster dar, während Emotionen nur grobe Handlungsimpulse (*Handlungstendenzen*) vorgeben, ohne das Verhalten wirklich einzuschränken.

5.6. Zusammenfassung und (Neu-)Definition des Emotionsbegriff für technische Systeme

Das *Gefühl* wird von den meisten Autoren als Teilkomponente einer Emotion angesehen, entsprechend dem *Reaktionstrias-Konzept*.

Als *elementare Affektzustände* werden solche Emotionen bezeichnet, die im Zusammenhang mit biologischen Grundbedürfnissen und ihrer Befriedigung auftreten und mit physiologischen Veränderungen einhergehen. Bezüglich der Abgrenzung der Emotionen, insbesondere der Basisemotionen, von elementaren Affektzuständen besteht in der aktuellen psychologischen und neurophysiologischen Forschung kein Konsens. Bei vielen Autoren (Ciompi, Ekman, Panksepp, Tomkins, James, James-Lange) werden elementare Affektzustände wie Aggression, Wut, Schmerz und Lust zu den Emotionen gezählt. Die Abgrenzung zu den im Zusammenhang mit physiologischen Bedürfnissen auftretenden Wahrnehmungen und Erlebnissen (Hunger, Durst, Müdigkeit) ist bei diesen Definitionen konzeptionell unklar. Für die Informatik liegt der zentrale Nachteil solcher Definitionen in der mangelnden Beleuchtung der Entstehungsmechanismen der elementaren Affektzustände sowie ihrer Funktion bei Wechselwirkungen mit höheren kognitiven Mechanismen des intelligenten Systems, wie z.B. der Einfluss auf Perzeption, Denkprozesse, Aufmerksamkeit, Motivation und Verhalten.

Die kategoriale Klassifikation von Emotionen in *Basisemotionen* und *zusammengesetzte* Emotionen spezifiziert ebenfalls keine Relation zwischen Emotionen und deren primären Entstehungsbedingungen. Sie liefert für die informatische Modellierung vorerst keine wesentlichen Erkenntnisse bezüglich der Entstehung und Wirkungsweise der emotionalen Mechanismen. Es gibt weder empirische Beweise für noch gegen die Existenz von Basisemotionen, es ist umstritten, wie viele und welche das sind und auch der Sinn und Zweck der Suche nach ihnen wurde angezweifelt (z. B. [OT90], S. 315). Ebenso fallen bei den *dimensionalen Modellen* der Emotionen die Fragen nach der Generierung und den funktionalen Auswirkungen einer Emotion nicht in das Spektrum des Forschungsinteresses. Sofern in dem informatischen Modell die Entstehung bzw. Funktionalität der emotionalen Mechanismen beschrieben werden soll, erscheinen die beiden Kategorisierungsmöglichkeiten für die Modellierung nicht sehr hilfreich zu sein..

In den meisten *emotionspsychologischen Arbeitsdefinitionen* werden zwar viele emotionstheoretische Konzepte erfasst, die Interaktion und der relative zeitlicher Ablauf der einzelnen Komponenten wird jedoch i.d.R. nicht näher spezifiziert. Auch die wechselseitigen kausalen oder strukturellen Abhängigkeiten sowie die Prozesse der Emotionsentstehung werden meist nicht betrachtet. Zudem werden die Auswirkung der Emotionen auf das Gesamtverhalten, d.h. die *funktionale Rolle* der Emotionen, nicht explizit definiert. Eine solche Emotionsdefinition ist nicht direkt operationalisierbar und bedarf zusätzlicher Interpretationen bzw. der Einführung ergänzender Konzepte, um in ein informatisches Modell umgewandelt zu werden.

Das Konzept der *Reaktionstrias*, das sehr oft in die Arbeitsdefinition mit aufgenommen wird, wird in der Literatur ebenfalls kontrovers diskutiert. Es ist nur bedingt als Emotionsdefinition geeignet und empirisch nicht ausreichend belegt [Rot99]. Eine eingehende Analyse in [ZS00] ergab, dass die meisten aktuellen Emotionstheorien nur Teilaspekte des Emotionsprozesses betrachten und / oder wesentliche Inhalte nicht erklären. So wird der Verhaltensaspekt nur in behavioristischen Theorien, seltener in Einschätzungstheorien mit-

5. Emotionsdefinition

erfasst (z.B. [Fri86], [Fri07]). Der physiologische Emotionsaspekt wird von einigen Autoren bewusst vernachlässigt [Slo03].

Angeichts der ambivalenten Begriffsdiskussion in der Emotionsforschung erscheinen insbesondere solche Emotionsdefinitionen in Bezug auf Formalisierbarkeit und Operationalisierbarkeit zur Umsetzung in einem informatischen System geeignet, die Emotionen nicht konzeptuell, sondern *funktional* beschreiben, indem sie die Auswirkungen der emotionalen Prozesse auf das Verhalten des Gesamtsystems spezifizieren.

Einschätzungstheorien erklären Emotionen als Ergebnis der kognitiven Bewertung der Umwelt in Bezug auf das individuelle Wohlergehen (Ziele, Pläne, Überzeugungen) eines Individuums. Die meisten einschätzungstheoretischen Emotionsdefinitionen treffen eine Aussage über die Entstehung und Auswirkungen einer Emotion. Sie entsprechen daher am ehesten dem Wunsch der Informatik nach einer funktionalen Beschreibung der emotionalen Mechanismen und nach einer Definition, die Emotionen “in terms of *their functional role*“ ([Sch02], S. 3; Hervorh. i.O.) beschreibt. Weitere Vorteile der Einschätzungstheorien in Bezug auf informatische Modellierung liegen in ihrer hohen Kompatibilität mit den Konzepten der BDI-Agentenarchitektur. Ziele bzw. Anliegen stellen in allen betrachteten einschätzungstheoretischen Definitionen eine wichtige Bewertungsdimension dar und weisen eine deutliche Ähnlichkeit mit dem Konzept der Desires eines BDI-Agenten auf. Einschätzungstheorien wurden bisher am häufigsten in informatischen Modellen umgesetzt.

Damasio betrachtet Emotionsprozesse aus der neurowissenschaftlichen Perspektive. Er hält Emotionen für angeborene automatische Anpassungsmechanismen und sieht ihren phylogenetischen Ursprung in primitiven regulatorischen Mechanismen (metabolischen Prozessen, Reflexen, mit Schmerz und Lust verbundenen Verhaltensweisen usw.). Seine Sichtweise entspricht diesbezüglich weitgehend dem evolutionspsychologischen Standpunkt, den er durch seine neurobiologischen Untersuchungsergebnisse zusätzlich stützen kann.

Damasio unterscheidet explizit metabolische Prozesse, Reflexe, mit Schmerz und Lust verbundene Verhaltensweisen, Motivationen und Emotionen. Eine weitere Unterteilung der Emotionen in Hintergrundemotionen, primäre Emotionen und soziale Emotionen sorgt für die terminologische Erfassung und Differenzierung vieler verschiedenartiger emotionaler Phänomene. Damasio's Emotionsdefinition liefert eine sehr differenzierte Beschreibung der neurobiologischen Prozesse bei der Emotionsgenerierung und kann als Verfeinerung der einschätzungstheoretischen Definitionen in Bezug auf bestimmte Schritte im Emotionsgenerierungsprozess gesehen werden.

Auswertung und Integration der Netzmodelle

Die Modellierung der betrachteten Emotionsdefinitionen mit Petrinetzen lässt gewisse strukturelle Ähnlichkeiten und grundsätzliche Anschlussmöglichkeiten verschiedener Definitionen erkennen.

So weisen die Netzmodelle der Basisemotionen-Definition und der einschätzungstheoretischen Emotionsdefinitionen die gleiche Grundstruktur auf.

Basisemotionen wurden als elementare emotionale Reaktionen aufgefasst, die automatisch durch einen spezifischen Stimulus ausgelöst werden. Sie werden durch angeborene neuronale Mechanismen (sog. angeborene Affektprogramme bzw. automatische Reaktionssysteme)

5.6. Zusammenfassung und (Neu-)Definition des Emotionsbegriff für technische Systeme

erzeugt. Das Auftreten eines spezifischen Stimulus ist somit notwendige und hinreichende Bedingung für die Entstehung einer Basisemotion (vgl. Kapitel 5.3).

Nach der einschätzungstheoretischen Definition wird eine Emotion als Ergebnis der (kognitiven) Bewertung eines (internen oder externen) Ereignisses in Bezug auf Bedürfnisse, Ziele (bzw. Anliegen) und Überzeugungen eines Organismus angesehen. Die Netzmodelle der betrachteten Emotionsdefinitionen von Lazarus, Frijda und Scherer konnten zu einem allgemeinen Netz zusammengefasst werden, das eine generelle Struktur der Emotionsgenerierung, wie sie in den einschätzungstheoretischen Definitionen beschrieben wird, zeigt (vgl. Kapitel 5.9).

Ein gemeinsames strukturelles Merkmal der Emotionsgenerierung ist bei beiden Netzmodellen das Auftreten eines internen oder externen Stimulus, das als Auslösebedingung für den Emotionsprozess fungiert und entweder zur automatischen Emotionsgenerierung oder zur Auslösung des Einschätzungs- bzw. Bewertungsprozesses führt.

In beiden Fällen stellen Emotionen eine *Reaktion* auf ein Ereignis dar. Wir werden in den Kapiteln 8 und 9 Emotionen, die durch einen *externen* Stimulus verursacht werden, als *reaktive Emotionen* und den zugehörigen Generierungsprozess als *reaktive Emotionsgenerierung* bezeichnen. Es lassen sich zwei strukturell verschiedene Arten der reaktiven Emotionsgenerierung unterscheiden: Einerseits unmittelbare automatische Reaktionen auf bestimmte (kritische) Ereignisse in der Umgebung, die Basisemotionen (in der Literatur oft auch als primäre oder elementare Emotionen bezeichnet) auslösen; andererseits die komplexen deliberativen Prozesse der Einschätzung der Umweltereignisse in Bezug auf Ziele, Bedürfnisse und Überzeugungen (vgl. die BDIs eines Agenten), die zur Generierung von Emotionen im Sinne der Einschätzungstheorien führen.

Ausgehend von den Netzmodellen in Abb. 5.3 und 5.9, kann man diese beiden Arten der Emotionsgenerierung in einem Netzmodell zusammenfassen. Abb. 5.11 zeigt ein kantenkonstantes gefärbtes Petrinetz, das mittels (Platz-)Faltung der Netzmodelle der Basisemotionen (Abb. 5.3) und der einschätzungstheoretischen Emotionsdefinition (Abb. 5.9) erzeugt wurde. Der Stelle **Stimulus** werden zwei unterschiedliche Farben zugewiesen, die zwei verschiedenen Arten von Ereignissen repräsentieren. Farbe a korrespondiert mit den spezifischen Reizen, die durch automatische Auslösung der Basisemotionen (Transition **auslösen**) beantwortet werden. Farbe b bezieht sich auf die Stimuli, die kognitive Bewertung erforderlich machen (Transition **Einschätzung**), was ggf. zur Generierung von Emotionen im Sinne der Einschätzungstheorien führt.

Das Netz aus Abb. 5.11 kann ferner um das Modell der Reaktionstrias ergänzt werden, welche nach [KK81] die Auswirkungen einer Emotion beschreibt (vgl. Abb. 5.5). Die beiden Netzmodelle werden in Abb. 5.12 zusammengefasst. Bei dem resultierenden Modell fällt auf, dass die Basisemotionen von automatischen körperlichen Veränderungen (vgl. gleichnamige Stelle in Abb. 5.11) begleitet werden (die entsprechenden Stellen im Netz 5.11 können nur gleichzeitig markiert werden), während die durch Einschätzungsprozesse zustandekommenden Emotionen (Stelle **eingeschätzte Emotion**) von neurophysiologischen Veränderungen (vgl. die gleichnamige Transition in Abb. 5.11) begleitet werden. Es ist anzunehmen, dass durch die beiden Transitionen der gleiche Prozess modelliert wird.

Bei der Definition von Basisemotionen wurde die zeitliche Abfolge bzw. der kausaler

5. Emotionsdefinition

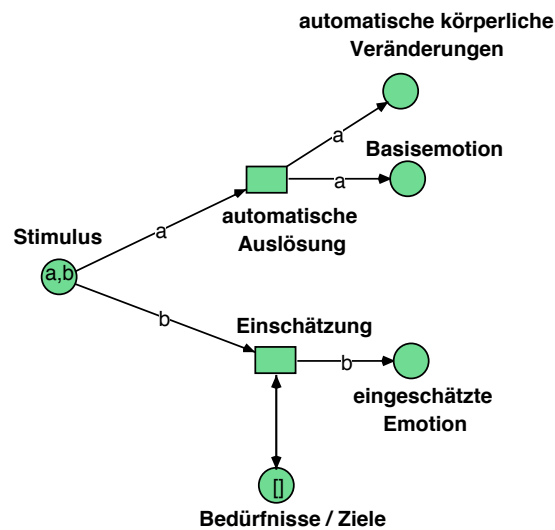


Abbildung 5.11.: Reaktive Emotionsgenerierung. Zusammenfassung der Modelle aus Abb. 5.3 und 5.9.

Zusammenhang zwischen der Emotionsentstehung und den körperlichen Reaktionen nicht genau festgelegt. Dem Modell wurde eine informelle Beschreibung von [Buc00] zugrundegelegt (vgl. Kapitel 8.6). Es liegt die Vermutung nahe, dass auch im Falle von Basisemotionen die gleiche zeitliche Reihenfolge der Emotionsentstehung und Reaktionstrias angenommen werden kann wie bei den “eingeschätzten Emotionen“. Dafür spricht zum einen die bessere theoretische Fundierung der Emotionsdefinition nach [KK81], auf der das Reaktionstrias-Modell gründet, zum anderen die Tatsache, dass viele Basisemotionen in die allgemeine Arbeitsdefinition von Emotionen mit aufgenommen wurden (vgl. Abb. 5.3) und auch von den einschätzungstheoretischen Definitionen erfasst werden (vgl. Abb. 5.4). Man kann also davon sprechen, dass der Begriff “Basisemotionen“ eine spezifische Teilmenge von Emotionen beschreibt, wie sie in den einschätzungstheoretischen und emotionspsychologischen Definitionen spezifiziert wird.

Unter Berücksichtigung dieser Erkenntnisse kann das Modell aus Abb. 5.11 folgendermaßen verallgemeinert werden (vgl. Abb. 5.12): Ein externer oder interner Stimulus (Stelle **Stimulus**) löst den Emotionsgenerierungsprozess (Transition **Emotionsgenerierung**) aus. Dieser kann sowohl in einer automatischen Erzeugung von Basisemotionen aufgrund von charakteristischen Stimuli bestehen, als auch die (komplexeren) kognitiven Prozesse der Stimuluseinschätzung in Abhängigkeit von den individuellen Bedürfnissen und Zielen (Stelle **Bedürfnisse / Ziele**) einschließen. Durch den Emotionsgenerierungsprozess wird der Organismus in einen emotionalen Zustand versetzt (vgl. gleichnamige Stelle), in dem eine Reihe von Veränderungen im Ablauf der neurophysiologischen und kognitiven Prozesse, im (zielgerichtetem und expressivem) Verhalten und im subjektivem Erleben, einsetzt. Dadurch wird der Organismus in die Lage versetzt, sich der aktuellen Umweltsituation optimal anpassen zu können (vgl. z. B. [Laz91], [Dam03]).

5.6. Zusammenfassung und (Neu-)Definition des Emotionsbegriff für technische Systeme

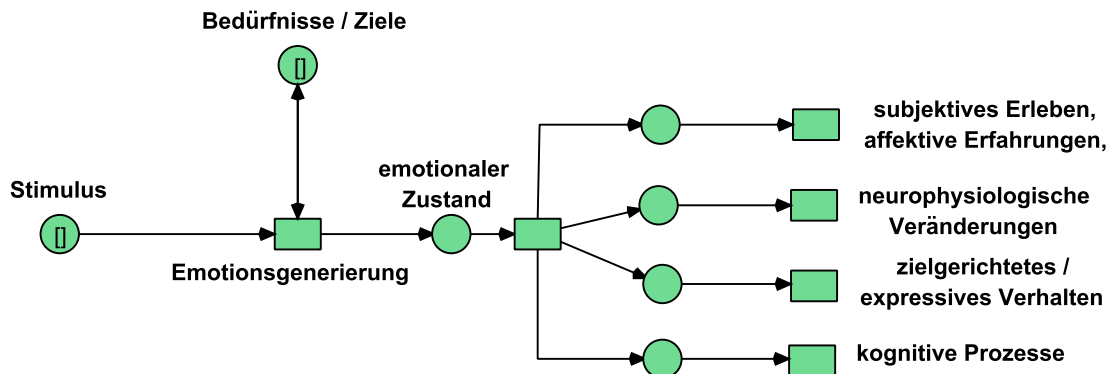


Abbildung 5.12.: Allgemeines Modell der Emotionsgenerierung. Integration der Modelle aus Abb. 5.11 und 5.5

Definition des Emotionsbegriffs für technische Systeme

Speziell für die informatische Zielsetzung der technischen Systemoptimierung, bei der es primär um die Nachbildung der natürlichen Funktionalität von Emotionen in künstlichen intelligenten Systemen geht, wäre es möglich, eine engere Emotionsdefinition zu geben, als beispielsweise in der psychologischen und neurowissenschaftlichen Forschung benötigt.

Während bei den sozial- und neurowissenschaftlichen Theorien der wissenschaftstheoretische Erklärungsanspruch im Vordergrund steht, emotionale Prozesse, wie sie in natürlichen Organismen und Menschen ablaufen, in ihrer gesamten Komplexität zu beschreiben, reicht es bei der Nachbildung der Funktionalität von Emotionen in einem *KI-Emotionsmodell*⁴ aus, lediglich die einzelnen, für die jeweilige technische Anwendung relevanten Aspekte der Emotionalität auszusondern und sie formal zu beschreiben bzw. nachzumodellieren.

Die Abgrenzung der Emotionen zu Reflexen, Trieben, Motivationen usw. spielt hierbei nicht die konzeptionell entscheidende Rolle, sondern ist allein wegen der funktionalen Umsetzung und eventuellen wechselseitigen Abhängigkeiten interessant. Ferner sind etwa das subjektive Erleben und affektive Erfahrungen, die mit einer Emotion zusammenhängen, für die meisten informatischen Fragestellungen nicht relevant. Ebenso sind die neurophysiologischen Veränderungen i.d.R. kein primärer Gegenstand der informatischen Modellierung, es sei denn, das Forschungsinteresse ist speziell auf die Modellierung der zugrunde liegenden Mechanismen gerichtet.

Daher kann das Modell der Emotionsgenerierung aus Abb. 5.12 für den speziellen Fall der Emotionsmodellierung in technischen Systemen vereinfacht werden, da die neurophysiologischen Veränderungen (Transition *neurophysiologische Veränderungen*) und das subjektive Emotionserleben (Transition *subjektives Erleben, affektive Erfahrungen*) in diesen Systemen

⁴Vgl. Kapitel 1.5 zur Terminologie.

5. Emotionsdefinition

meist nicht berücksichtigt werden müssen und aus dem allgemeinen Modell eliminiert werden können.

Es ergibt sich folgendes generelle Emotionsmodell, das die Bedeutung des Begriffs “Emotion“ im Kontext der Emotionsmodellierung in technischen Systemen, insbesondere in KI-Systemen, spezifiziert (Abb. 5.13): Ein externer oder interner Stimulus (Stelle *externe / interne Stimuli*) löst den Emotionsgenerierungsprozess (Transition *Emotionsgenerierung*) aus. Bei diesem Prozess wird der Auslösestimulus im Hinblick auf Ziele, Überzeugungen und Standards (Stelle *BDI*) eines Agenten ausgewertet. Durch den Emotionsgenerierungsprozess wird der Agent in einen emotionalen Zustand versetzt (Stelle *emotionaler Zustand*), in dem eine Reihe von Veränderungen im Ablauf der internen Verarbeitungsprozesse (Transition *kognitive Prozesse*), insbesondere der Wahrnehmung, Aufmerksamkeitsverteilung, Informationsverarbeitung, Planung, Aktionswahl usw., einsetzt bzw. diese Prozesse initiiert werden. Daneben kann sich der emotionale Zustand auf das zielgerichtete und expressive Verhalten (Transition *zielgerichtetes / expressives Verhalten*) auswirken und sich z.B. in verbaler und nonverbaler Emotionsrepräsentation bzw. -kommunikation manifestieren. Im emotionalen Zustand wird das System in die Lage versetzt, auf seine (ggf. veränderten) Umweltbedingungen optimal zu reagieren.

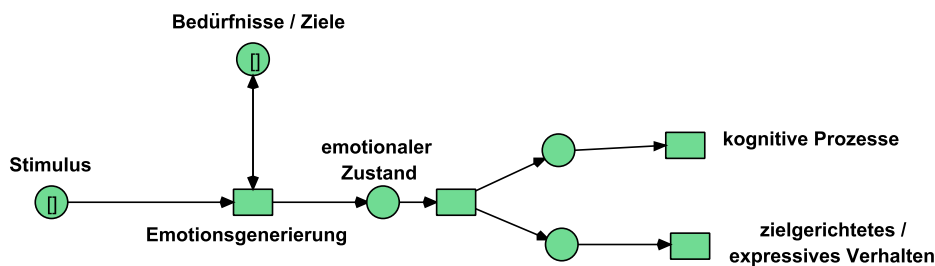


Abbildung 5.13.: Allgemeines Emotionsmodell für technische Systeme.

Aus diesem Modell lässt sich eine Emotionsdefinition ableiten, die wir zur Spezifikation des Emotionsbegriffs im Kontext der Modellierung der Emotionen in Agenten- und KI-Systemen für die weiteren Teile dieser Arbeit als Grundlage nehmen.

In dieser Arbeit werden wir Emotionen als komplexe natürliche Mechanismen verstehen, die in einem natürlichen intelligenten System als Resultat der Einschätzung seiner Umweltsituation entstehen und das System in einen Zustand versetzen, in dem es auf seine Umwelt optimal reagieren kann (*emotionaler Zustand*). In einem emotionalen Zustand verändert sich der Ablauf vieler Systemprozesse wie Wahrnehmung, Aufmerksamkeitsverteilung, Entscheidungsprozesse, Informationsverarbeitung, Planung, Aktionswahl u.a. Die Funktionalität der Emotionen hinsichtlich der Veränderung des Systemverhaltens lässt sich nach der hier vertretenen Hypothese in ihren einzelnen Aspekten in technischen Artefakten nachbilden und trägt zu ihrer Optimierung bei.

5.6. Zusammenfassung und (Neu-)Definition des Emotionsbegriff für technische Systeme

In [TWB07] wird die folgende Definition eines KI-Emotionsmodells gegeben: “Computational emotion model should include two parts:

- mechanisms eliciting emotions from external and internal stimuli, including potentially the agent’s own goals, beliefs and standards;
- emotion representations keeping track of the emotional states and their changes over time. “ ([TWB07], S. 1537).

In Abb. 5.14 wird die Definition von [TWB07] mit einem Netz grafisch veranschaulicht. Der Einschätzungsmechanismus und die Emotionsrepräsentation werden als zwei notwendige Bestandteile eines Emotionsmodells identifiziert. Wie in den meisten einschätzungstheoretischen Definitionen wird eine Emotion als Ergebnis der Einschätzung (Transition Emotionsgenerierung) der (internen oder externen) Ereignisse (Stelle externe / interne Stimuli) im Hinblick auf Ziele, Überzeugungen und Standards (Stelle BDI) eines Agenten gesehen. Die gesonderte Repräsentation des emotionalen Zustandes (Stelle **Emotionsrepräsentation, emotionaler Zustand**) ermöglicht die zusätzliche Berücksichtigung der aktuellen emotionalen Verfassung bei der Einschätzung der Stimuli. Damit kann die unterschiedliche Einschätzung des gleichen Stimulusereignisses aufgrund von verschiedenen emotionalen Zuständen repräsentiert werden. Der aktuelle emotionale Zustand steuert den Emotionsausdruck (Transition Emotionsausdruck), der in dem Modell von [TWB07] als Verhalten (“behaviour activations”) aufgefasst wird (vgl. [TWB07], S. 1538).

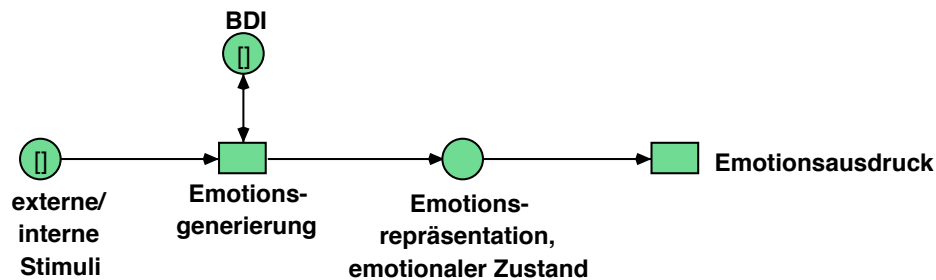


Abbildung 5.14.: Grafische Veranschaulichung der Definition eines KI-Emotionsmodells nach [TWB07].

Wie man sieht, deckt sich diese Definition des KI-Emotionsmodells weitestgehend mit der Darstellung in Kapitel 5.13. Die Einschätzungsprozesse generieren aus den externen und internen Stimuli einen emotionalen Zustand, der sich auf das Agentenverhalten auswirkt.

Allerdings wird in [TWB07] unter “Verhalten“ vor allem der Emotionsausdruck verstanden (“Behaviour activations: happy, angry [...] correspond most directly to emotion-related behaviours such as smiling“ [TWB07], S. 1538). Nach dem Reaktionstrias-Konzept (vgl. Kapitel 5.5) ist jedoch der Aspekt des Ausdrucks, insbesondere das expressive Verhalten, lediglich eine (Teil-)Komponente des Einflusses von Emotionen auf das Verhalten. Der Einfluss der Emotionen auf die weiteren Komponenten, insbesondere auf das zielgerichtete Verhalten - d.h. die Aktionen eines Agenten - und die von [KK81] spezifizierten Auswirkungen auf

5. *Emotionsdefinition*

kognitive Prozesse (wie Wahrnehmung, Bewertungen, Klassifikationsprozesse usw.) werden in der Emotionsmodell-Definition von [TWB07] nicht berücksichtigt. Dabei sind solche Effekte der Emotionen für die Modellierung in KI-Systemen besonders interessant, weil sie zur Optimierung der Verarbeitungsprozesse und zur Steigerung der Systemeffizienz beitragen könnten ([Can03], [Sch02]).

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Um Emotionen theoretisch fundiert modellieren zu können, benötigt die Informatik eine interdisziplinäre integrative Emotionstheorie, die sowohl psychologische als auch soziologische und neurobiologische Faktoren berücksichtigt. Will man Emotion zur technischen Systemoptimierung einsetzen, braucht man operationalisierbare psychologische und neurowissenschaftliche Beschreibungen, die funktionale Wechselwirkungen zwischen Emotionen und anderen internen Komponenten eines Agenten wie Kognition, Perzeption und Verhalten spezifizieren. Im Kontext der Implementierung der Emotionen in Multiagentensystemen ist die Auseinandersetzung mit soziologischen Grundlagen von besonderer Bedeutung. Es werden soziologische Theorien benötigt, die Zusammenhänge zwischen Emotionen und Strukturation bzw. Kontrollprozessen in (Agenten-)Gesellschaften beschreiben (vgl. Kapitel 2).

In diesem Kapitel werden die Emotionstheorien verschiedener Disziplinen betrachtet mit dem Ziel, ein möglichst umfassendes Bild des aktuellen Standes der Emotionsforschung aufzuzeigen. Da die zentrale Zielsetzung dieser Theorienbetrachtung in der Auswahl und Bewertung der Emotionstheorien im Hinblick auf ihre Eignung für die Emotionsmodellierung in der Informatik liegt, werden zunächst die Bewertungskriterien definiert, die eine entsprechende Beurteilung ermöglichen. Im Anschluss an die Theorienbeschreibung und -modellierung wird immer eine Auswertung der betrachteten Theorie nach diesen Bewertungskriterien vorgenommen.

Bewertungskriterien Die Übersicht der Emotionstheorien soll die Entstehung und Funktion der Emotion bei einem Akteur, einer Gruppe bzw. einem natürlichen Organismus in der für einen Informatiker leicht zugänglichen Beschreibungssprache der Petrinetz-Modelle darstellen und dadurch den Transfer der modellierten sozialwissenschaftlichen Konzepte in die Informatik-Domäne erleichtern.

Der besondere Schwerpunkt wird, entsprechend des Gesamtziels der Arbeit (vgl. Kapitel 1.3), auf die Aufbereitung der emotionstheoretischen Grundlage für die Emotionsmodellierung in Agentensystemen gelegt. Deshalb werden solche sozialwissenschaftlichen Theorien untersucht, die die Erfassung funktionaler und sozialer Emotionskomponenten ermöglichen bzw. Ansatzpunkte für die Integration akteursexterner soziologischer und akteursinterner psychologischer Sichtweisen im Hinblick auf die Strukturier- und Kontrollierbarkeit künstlicher Agentengesellschaften bieten. Damit wird die bevorzugte Betrachtung soziologischer

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

und neurowissenschaftlicher Ansätze begründet, welche schwerpunktmäßig die Problematik der Gesellschaftsstrukturierung und die Normetablierung durch Emotion betrachten.

Für die Beurteilung der sozialwissenschaftlichen Emotionstheorien werden aus diesen Zielsetzungen folgende Bewertungskriterien abgeleitet:

1. **Integrierbarkeit von verschiedenen Emotionstheorien:** Die Anschlussfähigkeit der Emotionstheorien an andere Emotionstheorien aufgrund ihrer konzeptuellen Emotionsdefinition soll geprüft werden, um dem Bedarf der Informatik nach einer umfassenden theoretischen Fundierung der Emotionalität und einer generalisierenden Emotionstheorie gerecht zu werden, die für verschiedenartige Anwendungen, vor allem im Bereich der Agententechnologien einsetzbar wäre.
2. **Exaktheit, Detailliertheit und Eindeutigkeit der Emotionsdefinition:** Nicht nur für die Integration verschiedener emotionstheoretischer Ansätze, sondern auch für das allgemeine Verständnis und die Positionierung einer Emotionstheorie sowie für die Emotionsspezifikation in einem ComputermodeLL spielt die präzise Emotionsdefinition eine grundlegende Rolle.
3. **Die Formalisierbarkeit und Operationalisierbarkeit der Emotionsdefinition bzw. -theorie** muss beurteilt werden, um die Umsetzbarkeit der Theorien in einem informatischen Modell zu überprüfen.
4. **Die empirische Überprüfung einer Emotionstheorie** ist Voraussetzung für die Korrektheit der auf dieser gründenden Emotionsmodelle.
5. **Erklärungspotenzial der Emotionstheorien:** Das Erklärungspotenzial der Emotionstheorie soll bewertet werden, um auf die Anwendungsbereiche der Theorie schließen zu können. Für die Informatik sind die funktionalen Zusammenhänge zwischen Emotionen und anderen Mechanismen des intelligenten Systems wichtig. Ebenso ist die Berücksichtigung der sozialen Faktoren der Emotionsgenese und Emotionsexpression in einer Emotionstheorie für viele informatische Anwendungen wünschenswert. Es soll daher beurteilt werden, inwiefern die Emotionstheorien diese und andere Zusammenhänge zu beleuchten und formal darzustellen vermögen.

Da die zentrale Zielsetzung der Theorienbetrachtung in der Auswahl und Bewertung der Emotionstheorien im Hinblick auf ihre Eignung für die Emotionsmodellierung liegt, kann der Anspruch dieses Kapitels nicht darin bestehen, jeden theoretischen Ansatz in seiner gesamten Komplexität zu modellieren. Vielmehr geht es darum, für die informatische Modellierung besonders interessante Ansätze zu identifizieren und ihre Kernelemente unter Anwendung des sozionischen Modellierungsansatzes als Petrinetz-Modelle zu repräsentieren. Die Petrinetz-Modelle dienen hierbei lediglich der graphischen Veranschaulichung einzelner Aspekte der betrachteten Theorien und sollen nicht als semantisch korrekte Theorienmodelle bzw. gar als "Emotionsmodelle" interpretiert werden.

6.1. Psychologische Theorien der Emotion

In der Psychologie bilden Emotionen zwar seit langer Zeit einen wichtigen Forschungsgegenstand, eine allgemeine psychologische Theorie der Emotion gibt es bis heute jedoch nicht. Die meisten psychologischen Emotionstheorien konzentrieren sich auf einen oder mehrere *Teilaspekte des Emotionsphänomens* (im Sinne unseres Referenzmodells aus Abb. 5 bzw. der Emotionsdefinition in Kapitel 5.12), während sie andere wesentliche Komponenten nicht zu erklären versuchen.

Diese Tendenz zur Zersplitterung des Emotionsvorganges in einzelne Aspekte ist sowohl für die historischen Ansätze, als auch für die neueren Emotionstheorien charakteristisch wie die Analyse des Geltungsbereichs der aktuellen psychologischen Emotionstheorien in [ZS00] gezeigt hat. In vielen Fällen wird “lediglich eine der Konstituenten des Emotionskonstrukts in den Mittelpunkt gestellt“ ([ZS00], S. 151). “Eine umfassende theoretische Erklärung des Emotionsphänomens [...] erfordert jedoch *integrative Theorien*, die der Multikomponentalität des Phänomens Rechnung tragen und die insbesondere konkrete Vorhersagen über Auslösung und Differenzierung von Emotionsprozessen als auch über die hierbei auftretenden Reaktionsmuster in den verschiedenen [System-]Komponenten machen können.“ ([ZS00], S. 157).

Wenn man das Emotionsphänomen definiert als einen “multikomponentiellen Prozess, der sich primär durch Anpassungsreaktionen auf Ereignisse und Objekte auszeichnet, die ein Organismus als wichtig für sein Wohlergehen einschätzt“ ([ZS00], S. 151; vgl. auch Kapitel 5.12) und zu den Komponenten dieses Prozesses die emotionale Reaktionstrias (vgl. Kapitel 5.4), emotionsdifferenzierende Kognitionen (Einschätzungs- bzw. Bewertungsprozess; vgl. Kapitel 5.4) und die resultierenden Verhaltensdispositionen oder Handlungstendenzen rechnet, so ist die Aufgabe einer dieser Phänomensdefinition angemessenen Theorie die Vorhersage, “unter welchen Bedingungen ein solcher multikomponentieller Prozess einsetzt und wie er verläuft.“ (ebd.).

Das Ziel dieses Abschnittes besteht daher in der Untersuchung der psychologischen Ansätze und der Analyse ihres jeweiligen Beitrages zur Erklärung einzelner Aspekte des emotionalen Prozesses.

Als Referenzmodell für die Auswertung der psychologischen Theorien der Emotion findet einerseits die im Abschnitt 5 spezifizierte funktionale Beschreibung des Emotionsprozesses (vgl. Abb. 5.2) Anwendung, andererseits wird die im Abschnitt 5.12 hergeleitete Emotionsdefinition hinzugezogen.

Ferner benutzen wir die in [ZS00] aufgestellten Kriterien für eine *integrative Theorie*, um die Vollständigkeit und den Erklärungsanspruch der emotionspsychologischen Ansätze zu beurteilen.

Zuerst erfolgt die Klassifizierung der psychologischen Emotionstheorien anhand ihrer jeweiligen zentralen Fragestellung. Anschließend betrachten wir die klassischen Emotionstheorien der Psychologie, insbesondere die Theorien von James-Lange (1894), Cannon-Bard (1929) und Schachter und Singer (1964), und modellieren ihre grundlegenden Aussagen entsprechend dem sozionischen Modellierungsansatz unter Verwendung der vermittelnden Netze (siehe Kapitel 4.4.2). Im nächsten Abschnitt wird eine wichtige Unterklasse der

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

kognitiven Emotionstheorien, die Einschätzungstheorien, vorgestellt. Die Theorien der wichtigsten Vertreter dieser Klasse - Ortony, Clore und Collins (1988), Scherer (1987/1993), Lazarus (1991) und Frijda(1986) - werden ausführlich beschrieben und modelliert. Eine Gesamtauswertung der betrachteten psychologischen Theorien aus Sicht der informatischen Modellierung erfolgt am Ende des Kapitels.

Klassifizierung der psychologischen Emotionstheorien

Anhand der zentralen Fragestellung lassen sich in der psychologischen Emotionsforschung evolutionspsychologische, behavioristische bzw. lerntheoretische und kognitive Emotionstheorien unterscheiden (vgl. [Rei00b], [Rei00a]).

- Für **evolutionspsychologische Emotionstheorien** ist die Frage nach der phylogenetischen Entwicklung und der biologischen Funktion von Emotionen zentral.
- **Behavioristische / lerntheoretische Emotionstheorien** gehen von der Annahme aus, dass emotionale Dispositionen nicht vererbt, sondern zu einem erheblichen Teil erlernt sind. Emotion wird von Behavioristen als ein intersubjektiv beobachtbares Reaktionsmuster angesehen, das durch bestimmte Umweltgegebenheiten verlässlich ausgelöst wird. Den evolutionspsychologischen und lerntheoretischen Emotionstheorien gemeinsam ist die zentrale Frage nach der Herkunft der grundlegenden Dispositionen bzw. Mechanismen der Emotionen sowie nach dem Rollenverhältnis von Anlage und Umwelt (Evolution und Sozialisation).
- **Kognitive Emotionstheorien** führen die Intensität und Qualität der Emotion gegenüber einem Objekt auf eine bestimmte Art von Kognitionen hinsichtlich dieses Objektes – die (kognitiven) Einschätzungen (*appraisal*) – zurück. Die Theorie von Schachter und Singer ist ein klassisches Beispiel einer kognitiven Emotionstheorie. Im Unterschied zu evolutionspsychologischen und lerntheoretischen Emotionstheorien beschäftigen sich kognitive Theorien in erster Linie mit der Aktualgenese von Emotionen in einer konkreten Situation.

6.1.1. Klassische Emotionstheorien der Psychologie

Eine eingehende Analyse der wichtigen Emotionstheorien in [ZS00] hat gezeigt, dass die meisten psychologischen Emotionstheorien lediglich als Teiltheorien anzusehen sind, da sie sich auf bestimmte Aspekte des Emotionsphänomens spezialisiert haben und ihren Erklärungsanspruch nur innerhalb dieses Geltungsbereichs sehen. In diesem Abschnitt werden die “klassischen“ Emotionstheorien von James-Lange (1894), Cannon-Bard (1929) und Schachter und Singer (1964) vorgestellt und bezüglich ihres Beitrages zur Erklärung einzelner Aspekte des emotionalen Prozesses (im Sinne des in Abb. 5.2 spezifizierten Referenzmodells) analysiert.

Die James-Lange-Theorie

Die James-Lange-Theorie (1894) beschäftigt sich vorwiegend mit der Beziehung zwischen dem subjektiven Gefühlszustand und peripheren physiologischen Reaktionen ([ZS00], vgl. auch Reaktionstrias-Konzept, in Kapitel 5.4).

Die zentrale Behauptung dieser Theorie besteht darin, dass den Emotionen körperliche Veränderungen vorangehen und Emotionen das bewusste subjektive Erleben der physiologischen Veränderungen darstellen. Während James für das bewusste Erleben einer Emotion *viszerale* Reaktionen verantwortlich macht, hält Lange (1885) ausschließlich *vasomotorische* Reaktionen für eine hinreichende Bedingung für die subjektive Wahrnehmung einer Emotion.

Das Netz in Abb. 6.1 veranschaulicht den Zusammenhang zwischen physiologischen Veränderungen und emotionaler Reaktion nach der James-Lange-Theorie.

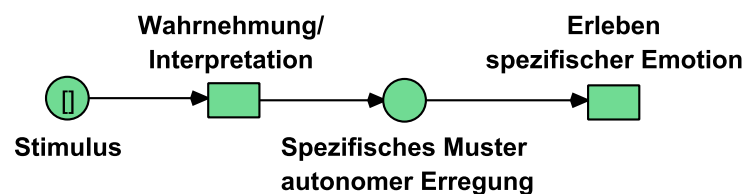


Abbildung 6.1.: Grafische Veranschaulichung der James-Lange-Theorie (1894).

Die Cannon-Bard-Theorie

Die zentralen Aussagen der James-Lange-Theorie wurden von sogenannten neo-jamesianischen Emotionstheorien in Frage gestellt [Can27].

Es wurde eine alternative Theorie vorgestellt, die *Cannon-Bard-Theorie*, nach der im Gegensatz zu den peripheren Prozessen der James-Lange-Theorie die Aktivierung bestimmter Strukturen des zentralen Nervensystems als Ursache für emotionale Zustände angenommen wird.

Durch die emotionsauslösenden Impulse werden die im Thalamus gespeicherten vorprogrammierten Reizmuster aktiviert. Dadurch wird das emotionale Erlebnis hervorgerufen, während gleichzeitig physiologische Veränderungen angeregt werden. Keine der beiden Reaktionen bedingt dabei die andere, vielmehr sind beide unabhängige Reaktionen auf einen wahrgenommenen Reiz.

Abb. 6.2 veranschaulicht diesen Sachverhalt: Körperliche Veränderung (Stelle **Generelles Muster autonomer Erregung**) und emotionales Erleben (Stelle **Erleben spezifischer Emotion**) entstehen infolge der Wahrnehmung und Interpretation des emotionsauslösenden Stimulus (Transition **Wahrnehmung / Interpretation**) simultan.

Auch die Cannon-Bard-Theorie kann nur als Teiltheorie angesehen werden, da sie sich vor allem mit den Zusammenhängen zwischen den subjektiven und den physiologischen Aspekten der Emotion beschäftigt (vgl. das in Kapitel 5.4 beschriebene Reaktionstrias-Konzept).

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

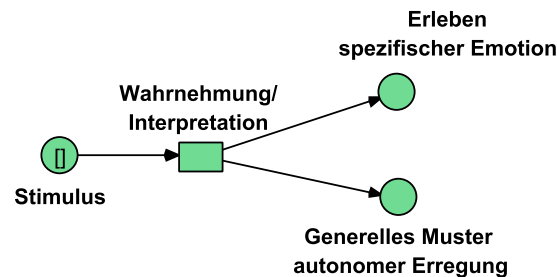


Abbildung 6.2.: Grafische Veranschaulichung der Cannon-Bard-Theorie (1929).

Die Emotionstheorie von Schachter und Singer

Auch die bis vor Kurzem sehr einflussreiche Emotionstheorie von Schachter und Singer (1962) hebt einen Teilaspekt der Emotionalität hervor, nämlich die Beziehung zwischen dem verbal etikettierten Gefühlszustand und den physiologischen Reaktionskomponenten [ZS00].

Obwohl die Theorie schon 1924 unter dem Namen *2-Komponenten-Theorie der Emotion* publiziert worden war und danach ähnliche Modelle zur Emotionsentstehung veröffentlicht wurden (z.B. Russel (1927), Duffy (1941)), war es die Theorie von Schachter und Singer, die aufgrund des ihr zugrunde liegenden Experimentaldesigns den größten Einfluss auf die Psychologie der folgenden zwanzig Jahre hatte und immer wieder zu Replikationsversuchen anregte.

Nach Schachter und Singer entstehen emotionale Erfahrungen aus der Interpretation, die ein Individuum, basierend auf der kognitiven Bewertung der Situation, seiner physiologischen Erregung (Arousal) zuschreibt.

Es werden zwei Dimensionen der Emotion unterschieden: Die *Emotionsintensität* ist abhängig vom Grad der physiologischen Erregung, die durch die Stimulussituation hervorgerufen wurde. Die *Emotionsqualität* ist durch die emotionsrelevante Einschätzung der Situation und die Kausalattribution der Erregung zu dieser Einschätzung bedingt. Abb. 6.3 veranschaulicht den Zusammenhang zwischen physiologischer Erregung und Emotionsgenerierung nach der Theorie von Schachter und Singer.

6.1.2. Einschätzungstheorien

Eine einflussreiche Unterklasse der kognitiven Emotionstheorien bilden die *Einschätzungs- oder Bewertungstheorien der Emotionen*. Gemeinsam ist diesen Emotionstheorien die Annahme eines kognitiven Bewertungsprozesses (*appraisal process*) als Ursache der Emotionsgenese. Ob ein Ereignis bei einer Person eine Emotion hervorruft, und wenn ja, welche (Freude, Trauer, Angst usw.) und mit welcher Intensität, hängt davon ab, wie die Person das Ereignis interpretiert, insbesondere, wie sie es relativ zu ihren Zielen und Wünschen bewertet. Diese alltagspsychologische Annahme wird in den Einschätzungstheorien präzisiert, indem die Qualität und Intensität der Emotionen gegenüber einem Objekt (einem Ereignis, einer

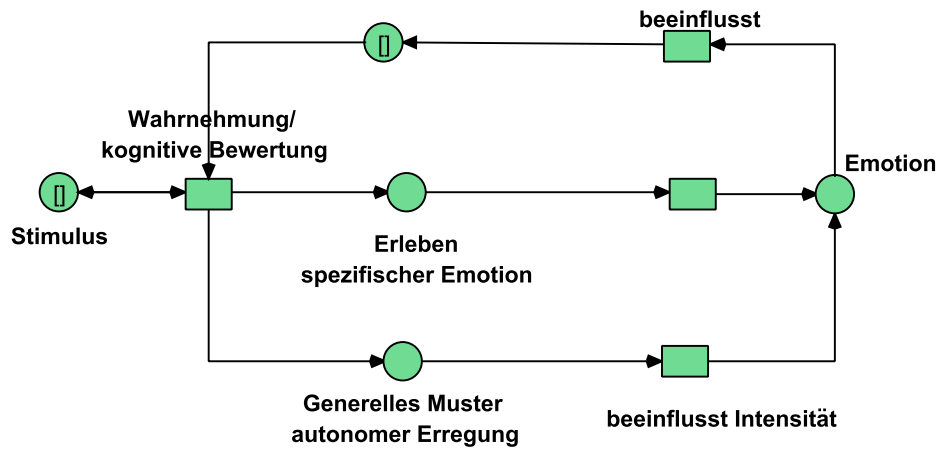


Abbildung 6.3.: Grafische Veranschaulichung der Emotionstheorie von Schachter und Singer (1962).

Situation, einer Person usw.) in Abhängigkeit zu verschiedenen Bewertungsdimensionen gesetzt werden.

Die Vertreter der einschätzungstheoretischen Tradition stimmen zwar in der grundlegenden Annahme überein, dass Emotionen durch kognitive (bewusste oder unbewusste) Bewertungsprozesse ausgelöst werden und diese die Art der emotionalen Reaktion bestimmen. Es bestehen jedoch wesentliche Unterschiede in Bezug auf die Ausgestaltung des Bewertungsprozesses. Insbesondere bleibt die Frage nach der Anzahl und Art der Bewertungsdimensionen umstritten. Viele Forscher haben ihre eigenen Komplexe von Bewertungskriterien vorgeschlagen (vgl. [vRS84], [Sch84], [SE85], [Fri86], [OCC88], [Laz91]). Ihre Ergebnisse variieren aber z.T. sehr stark und u.a. wegen der Schwierigkeit der empirischen Überprüfung konnte bisher bezüglich der anzunehmenden Bewertungsdimensionen keine Einigkeit erzielt werden.

Trotzdem gibt es nach [Rei00a] gegenwärtig in der psychologischen Emotionsforschung keine Alternative zu den Einschätzungstheorien, denn keine andere emotionstheoretische Tradition vermag die aktuellen empirischen Befunde und emotionspsychologischen Fakten umfassender zu erklären.

Die Mehrzahl der Computermodele von Emotionen, sofern sie sich ausdrücklich auf psychologische Theorienbildung beziehen, basieren auf den Einschätzungstheorien (vgl. Elliots “Affective Reasoner“ ([Ell92]), die “Oz“-Welt von Bates und Reilly ([Bat94], [BLR92a], [BLR92b]), Scherers “GENESE“ ([Sch93]), Bazzans “Simulations-Framework“ ([BAB02]) u.v.a.m.). Diese bieten sich für die informatische Modellierung insbesondere wegen ihrer einfachen Operationalisierbarkeit und dem engen Bezug zur Agentenmetapher an.

Einige Autoren (z.B. [OCC88]) haben ihre Theorien speziell im Hinblick auf die Implementierung in den Modellen der KI entwickelt: “[W]e would like to lay the foundation for a computationally tractable model of emotion. In other words, we would like an account of emotion that could in principle be used in an Artificial Intelligence (AI) system that would, for example, be able to reason about emotion“ ([OCC88], S. 2).

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Reisenzein (2000) unterscheidet folgende Grundkomponenten einer Einschätzungstheorie (nach [Rei00a], S. 120):

- Eine Menge von Einschätzungsdimensionen, welche verschiedene Arten von Objekteinschätzungen repräsentieren, wie zum Beispiel Valenz, Wahrscheinlichkeit, Ursächlichkeit oder Fairness. Die Einschätzungsdimensionen können qualitativ oder quantitativ, sowie unipolar oder bipolar sein.
- Eine Menge von Emotionen (Emotionstypen), welche die betreffende Strukturtheorie erklären soll.
- Eine oder mehrere (z.B. [OCC88]) Mengen von Objekten (je nach Theorie einzelne Ereignisse oder ganze Situationen), die bewertet werden und auf die sich die Einschätzungen und vermutlich auch die Emotionen beziehen.
- Eine Relation (häufig eine Funktion) R , die Muster von Einschätzungen, d.h. Kombinationen von Dimensionswerten, zu Emotionen in Beziehung setzt bzw. den Einschätzungsmustern Emotionen zuordnet. Häufig werden grafische Darstellungen der Relation R und damit auch der Menge der postulierten Einschätzungsdimensionen und der analysierten Emotionen verwendet. Tabellarische oder baumartige Darstellungen von Strukturtheorien der Einschätzung finden sich z.B. bei Ortony et al. 1988, Roseman 1984, Scherer 1993.

Ferner wird in [Rei00a] eine Klassifizierung der grundlegenden Postulate der Bewertungstheorien in *Strukturannahmen*, *Prozessannahmen* und *Annahmen zur Kognitions-Emotions-Beziehung* vorgeschlagen:

1. Strukturannahmen, die alle Bewertungstheorien beinhalten, sind die Annahmen über die Anzahl und Art der emotionsrelevanten Einschätzungsdimensionen und die Einschätzungsmuster, die mit unterschiedlichen Emotionen assoziiert sind.
2. Prozessannahmen sind die Annahmen über die Beschaffenheit der kognitiven Prozesse, durch die Einschätzungen gebildet werden, und
3. Annahmen zur Kognitions-Emotions-Beziehung präzisieren die Abhängigkeitsbeziehung zwischen Einschätzungen und Emotionen (z.B. kausal, Teil-Ganzes). „Mit dieser Präzisierung verbindet sich fast immer auch eine genauere Bestimmung der Natur der Emotionen selbst“ ([Rei00a], S. 119).

Zu diesen drei Klassen von Postulaten kommen in einigen Einschätzungstheorien noch weitere Annahmen hinzu, zum Beispiel über die Beziehung zwischen Einschätzung und dem Ausdruck von Emotionen (z.B. [Sch84]) oder zur evolutionären Grundlage der Einschätzungsmechanismen (z.B. [Laz91]).

Die verschiedenen Strukturannahmen und Bewertungsdimensionen der Einschätzungstheorien werden in den folgenden Abschnitten anhand der Theorien von Ortony, Clore und Collins (1988), Scherer (1984 /1993), Frijda (1986) und Lazarus (1991) vorgestellt. Um die Einschätzungstheorien mit Petrinetzen zu illustrieren, wird der sozionische Modellierungsansatz angewendet (vgl. Kapitel 4.4.2 und 4.5).

Die Einschätzungstheorie von Ortony, Clore und Collins

Die Theorie von Ortony, Clore und Collins (1988) (im Folgenden OCC-Theorie genannt) beschreibt den Prozess der kognitiven Einschätzung der Umgebung, der zur Generierung von Emotionen führt.

Die Autoren haben sich u.a. zum Ziel gesetzt, mit ihrer Theorie die Grundlage für ein Emotionsmodell zu liefern, das sich rechnergestützt simulieren lässt. Sie konzentrieren sich dabei auf die Frage, welche Rolle die Kognition bei der Entstehung von Emotionen spielt. Eine ihrer Grundannahmen bezüglich der Natur von Emotionen ist die Zurückführung einer Emotion auf die kognitive Einschätzung einer Situation durch den Beobachter ([OCC88], S. 1): “Taking the perspective of empirical psychology and cognitive science, we start with the assumption that emotions arise as a result of the way in which the situations that initiate them are construed by the experienter.”

Mit der OCC-Theorie können die kognitiven Einschätzungen von insgesamt 22 Emotionen explizit dargestellt werden. Neben der Angabe der Einschätzungsmuster für die 22 behandelten Emotionen präsentieren Ortony und Kollegen eine Theorie über die Determinanten der Intensität von Emotionen.

Die Autoren definieren Emotionen als “valenced reactions to events, agents, or objects, with their particular nature being determined by the way in which the eliciting situation is construed“ ([OCC88], S. 13). Ausgangspunkt der OCC-Theorie ist die Annahme, dass Akteure drei Arten von Stimuli in der Welt unterscheiden, auf die sie ihre Aufmerksamkeit richten und die sie einer kognitiven Einschätzung unterziehen (vgl. Abb. 6.5):

- Ereignisse und ihre Konsequenzen,
- Akteure und ihre Handlungen und
- Objekte (auch nichtmenschliche Lebewesen und Personen, sofern diese nicht als Akteure betrachtet werden) und ihre Eigenschaften.

Jede dieser drei Arten von Stimuli kann positiv oder negativ bewertet werden, wobei nach Ortony et al. jeweils unterschiedliche Bewertungskriterien zur Anwendung kommen: “When one focuses on events one does so because one is interested in their consequences, when one focuses on agents, one does so because of their actions, and when one focuses on objects, one is interested in certain aspects or imputed properties of them qua objects“ ([OCC88], S. 18). Je nachdem, wie eine Situation von einem Akteur anhand dieser Kriterien interpretiert wird, resultieren bestimmte diskrete Emotionen.

Der Einschätzungsprozess wird in der Theorie als sequenzieller Ablauf der Einschätzungen beschrieben (vgl. Abb. 6.4) und kann direkt mit einfachen P/T-Netzen modelliert werden.

Das Netz in Abb. 6.5 veranschaulicht die Einschätzung der drei grundlegenden Emotionsklassen, die von der OCC-Theorie identifiziert werden: Zuerst werden die in der Umwelt beobachteten Stimuli als *Ereignis*, *Akteur* oder *Objekt* identifiziert (Transition **observe environment**). Jeder dieser Klassen von Entitäten ist eine Einschätzungsdimension zugeordnet. So werden Ereignisse und ihre Konsequenzen nach dem Grad ihrer Erwünschtheit (*desirability*) bewertet (Transition **appraise desirability**), die Handlungen von Akteuren nach

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

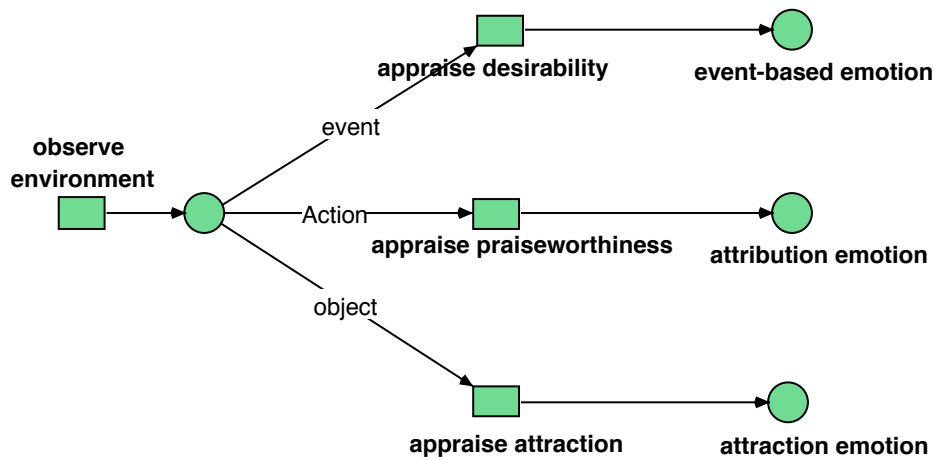


Abbildung 6.4.: Emotionsgenerierung in der OCC-Theorie (nach [OCC88], S. 19).

dem Grad ihrer Lobwürdigkeit (*praiseworthiness*) (Transition *appraise praiseworthiness*) und Objekte nach dem Grad ihrer Attraktivität (*appealingness*) (Transition *appraise attraction*) beurteilt ([OCC88], S. 33).

Infolge dieser Einschätzungen werden je nach Auslösestimulus entsprechend ereignisfundierte (Stelle *event-based emotion*), akteurfundierte (Stelle *attribution emotion*) oder objektfundierte Emotionen (Stelle *attraction emotion*) generiert.

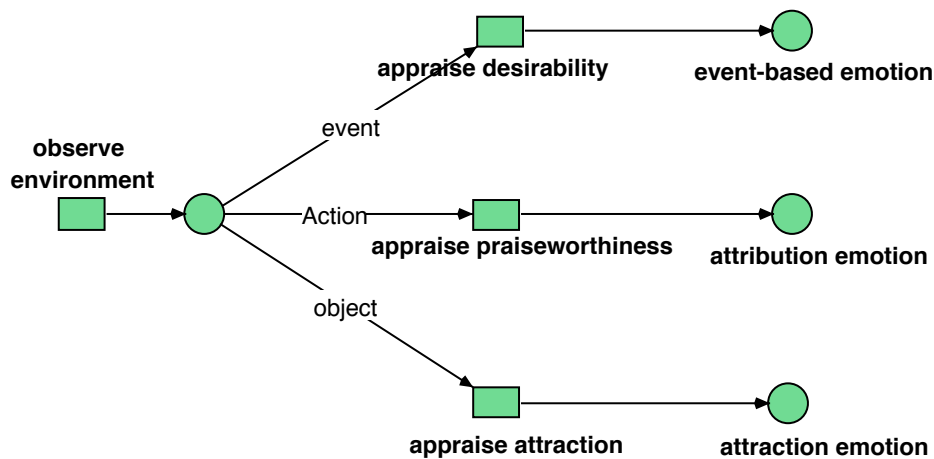


Abbildung 6.5.: OCC-Überblick.

Um der Diskretheit der Emotionsgenerierung nach [OCC88] Rechnung zu tragen, wird in Abb. 6.6 eine zusätzliche Stelle *diskr* eingeführt, durch die eine simultane Ausführung der Einschätzungen verhindert wird. Auf diese Weise können nie zwei verschiedene Emotionsarten (z. B. akteurfundierte und objektfundierte Emotionen) zur gleichen Zeit generiert

werden. Erst wenn eine Emotion abklingt (Transition **decay**), wird eine neue Einschätzung vorgenommen. Allerdings wird die Forderung nach einer diskreten Ausführung der Einschätzungen in der OCC-Theorie nicht explizit formuliert, sondern kann lediglich aus einigen Formulierungen und insbesondere aus der Aufteilung der Stimuli in drei Klassen bezüglich der Einschätzung gefolgert werden.

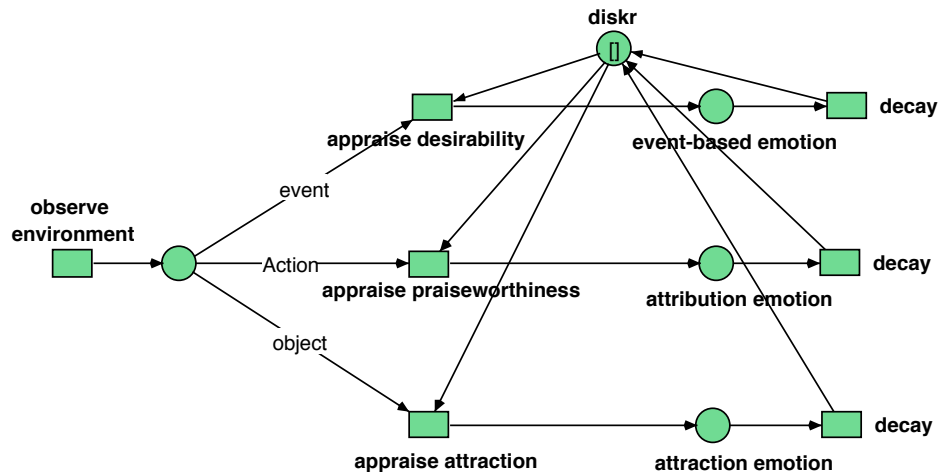


Abbildung 6.6.: OCC-Überblick: Diskrete Generierung von Emotionen.

Abb. 6.7 stellt eine Verfeinerung des Modells aus Abb. 6.5 dar, bei der die stimullenspezifischen Einschätzungsdimensionen in Anlehnung an Abb. 6.4 konkretisiert werden. Im Folgenden werden diese Einschätzungsdimensionen für jede der drei Emotionsklassen näher beschrieben. Auf die Modellierung der Diskretheit bei der Emotionsgenerierung wurde in diesem Netz verzichtet, um es übersichtlicher zu halten. Die simultane Ausführung der Einschätzungen kann netztechnisch mit einfachen Mitteln (z.B. durch Einführung zusätzlicher Puffer-Stellen und Transitionen für die Emotionszerfall wie in Abb. 6.6) verhindert werden.

Ereignisfundierte Emotionen. Ereignisfundierte Emotionen sind Ausdrücke der Zufriedenheit oder Unzufriedenheit (Stelle **pleased / displeased**) über die Konsequenzen eines bewerteten Ereignisses (Transition **appraise desirability**). Die ereignisfundierten Emotionen werden in mehrere Untergruppen in Abhängigkeit vom Ereignisfokus (Stelle **focus on** - man selbst vs. jemand anders ist betroffen), der subjektiven Ereigniswahrscheinlichkeitseinschätzung (Stelle **prospects relevance** - das Ereignis ist sicher vs. unsicher) sowie, bei unsicheren Ereignissen, der Erwartungskonformität des Ereignisses (Stelle **prospects confirmation** - erwartungsbestätigend, erwartungswidrig) unterteilt.

Betrifft das Ereignis hauptsächlich einen selbst (Transition **appraise consequences for self**), so resultieren in Abhängigkeit von der Erwünschtheit des Ereignisses Freude oder Sorge (Stelle **joy / distress**), wenn das Ereignis als sicher eingeschätzt wird und ereignisbezogene Erwartungen keine Rolle spielen (Transition **prospects irrelevant**). Wenn das Ereignis als unsicher eingeschätzt wird (Transition **prospects relevant**), entstehen zunächst Hoffnung

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

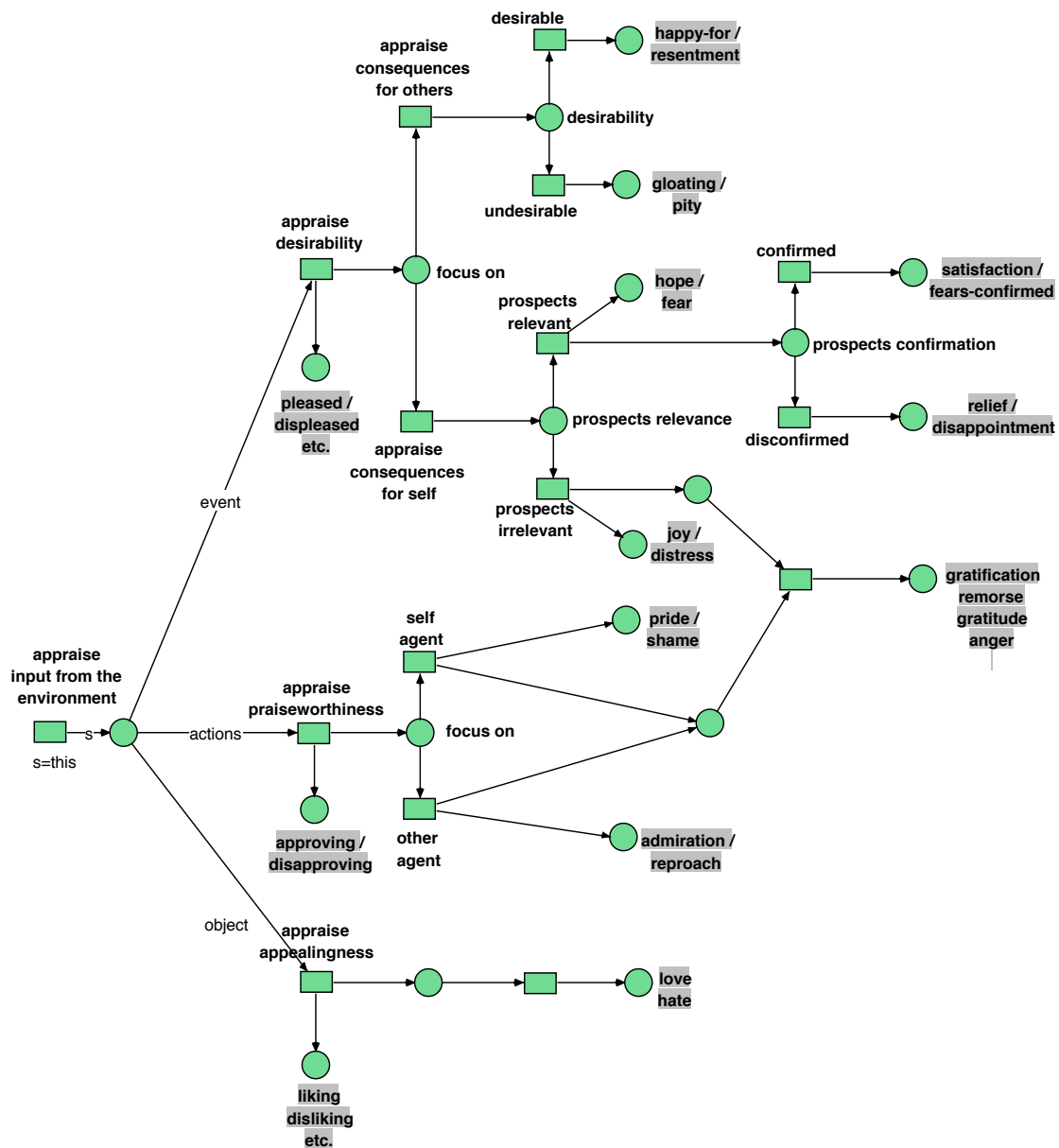


Abbildung 6.7.: Einschätzungsprozess nach [OCC88] (die eingeschätzten Emotionen sind grau unterlegt).

oder Angst (Stelle hope / fear). In Abhängigkeit davon, ob das Ereignis tatsächlich eingetreten ist (Stelle prospects confirmation), werden diese Emotionen durch Erleichterung oder Enttäuschung (Stelle relief / disappointment) abgelöst, falls das Ereignis erwartungswidrig ist (Transition disconfirmed), bzw. durch Befriedigung oder die sog. bestätigte Furcht (Stelle satisfaction / fears-confirmed), falls die Erwartungen bezüglich dieses Ereignisses sich

bestätigen (Transition **confirmed**).

Betrifft ein Ereignis in erster Linie eine andere Person (Transition **appraise consequences for others**), so resultieren in Abhängigkeit von der Erwünschtheit des Ereignisses (Stelle **desirability**), die Emotionen Mitfreude oder Schadenfreude (Stelle **happy-for / resentment**), falls man das Ereignis selbst als erwünscht bewertet (Transition **desirable**), bzw. Neid oder Mitleid (Stelle **gloating / pity**), falls man das Ereignis als unerwünscht bewertet (Transition **undesirable**).

Akteurfundierte Emotionen. Akteurfundierte Emotionen gründen auf der Emotion der Billigung bzw. Missbilligung (Stelle **approving / disapproving**) einer als lobenswert bzw. als tadelnswert bewerteten Handlung (Transition **appraise praiseworthiness**). Wenn man sich selbst für die einzuschätzende Handlung verantwortlich macht (Transition **self agent**), werden, je nach den Ergebnissen der Bewertung der Lobenswürdigkeit dieser Handlung (Transition **appraise praiseworthiness**), Stolz oder Schuld generiert (Stelle **pride / shame**). Macht man eine andere Person für die betreffende Handlung verantwortlich (Transition **other agent**), entstehen Bewunderung oder Empörung (Stelle **admiration / reproach**).

In Abb. 6.4 kann eine Reihe von Emotionen (gratification, remorse, gratitude, anger) weder als ereignisfundierte noch als akteurfundierte Emotionen klassifiziert werden. Die eindeutige Zuweisung zu einer Emotionsklasse wird dadurch unmöglich, dass diese Emotionen sowohl die Einschätzung der Erwünschtheit (*desirability*), die für ereignisfundierte Emotionen vorausgesetzt wird, als auch die Einschätzung der Lobwürdigkeit (*praiseworthiness*), die der Generierung der akteurfundierten Emotionen vorangeht, erfordern (vgl. Abb. 6.7). Dazu wird in [OCC88] erläutert, dass ereignisfundierte Emotionen häufig gemeinsam mit handlungsfundierten Emotionen auftreten, weil man Ereignisse mit einer Handlung und ihrem Urheber in Verbindung bringt. Zum Beispiel kann das Verhalten einer anderen Person, je nach dem kognitiven Fokus der Person, einmal als Handlung eines Akteurs oder ganz allgemein als Ereignis aufgefasst werden, wobei dann jeweils akteur- bzw. ereignisfundierte Emotionen erlebt werden.

Objektfundierte Emotionen. Objektfundierte Emotionen (auch Attraktivitätsemotionen genannt) können allgemein beschrieben werden als Mögen bzw. Nichtmögen (Stelle **liking / disliking etc.**) eines als attraktiv bzw. als abstoßend bewerteten Objekts (Transition **appraise appealingness**). Spezielle Unterformen des Mögens sind zum Beispiel Zuneigung, Verehrung und Liebe; spezielle Formen des Nichtmögens - Abneigung, Ekel und Hass.

Emotionsintensität. Zur Berechnung der Intensität einer emotionalen Empfindung wird eine Reihe von globalen und lokalen Intensitätsvariablen eingesetzt.

Sense-of-reality, proximity, unexpectedness und *arousal* sind die vier globalen Variablen, die über alle drei Emotionskategorien hinweg wirken.

Das Netz in Abb. 6.8 veranschaulicht die Berechnung der globalen Variablen (Transition **appraise global variables**), die je nach einer Emotionsklasse, der weiteren spezifischen Einschätzung vorangestellt ist. Das Netz stellt eine Verfeinerung des Netzes aus Abb. 6.6 dar.

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

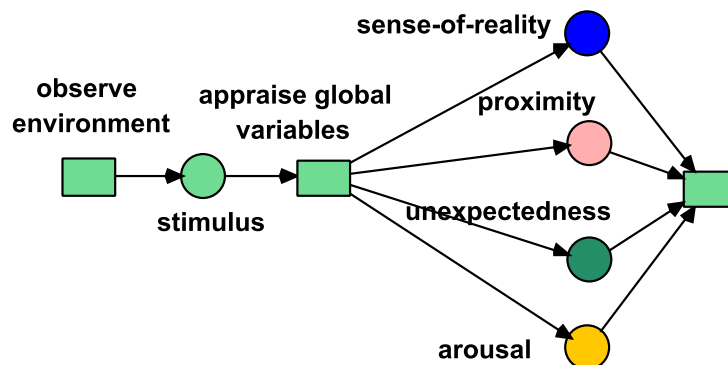


Abbildung 6.8.: Emotionsintensität in Abhängigkeit von globalen Variablen.

Die lokalen Variablen werden separat für jede Emotionskategorie - Ereignisse, Akteure und Objekte - definiert und anhand von allgemeinen Standardkriterien evaluiert:

- Erwünschtheit (*desirability*) ist verknüpft mit der Reaktion auf Ereignisse und wird im Hinblick auf Ziele (*goals*) des Individuums bestimmt;
- Lobwürdigkeit (*praiseworthiness*) ist verknüpft mit der Reaktion auf Aktionen von Akteuren und wird im Hinblick auf Normen und Standards (*standards*) evaluiert, wobei die jeweilige Aktion als lobenswert oder tadelnswert bewertet wird;
- Attraktivität (*Appealingness*) ist verknüpft mit der Reaktion auf Objekte und bewertet diese anhand der Einstellungen (*attitudes*) eines Akteurs als anziehend oder abstoßend.

Jeder dieser Variablen wird im konkreten Fall ein Wert (*value*) und eine Gewichtung (*weight*) zugewiesen. Zudem gibt es für jede Emotion einen Schwellenwert (*threshold*), unterhalb dessen eine Emotion nicht subjektiv empfunden wird.

Unter Zielen (*goals*) werden in [OCC88] insbesondere kurz- und mittelfristige Ziele verstanden, die ein Akteur verfolgt. Darüber hinaus werden aktiv verfolgte Ziele (*active-pursuit goals*), Interessen (*interest goals*) und regelmäßig wiederkehrende Ziele (*replenishment goals*) unterschieden.

- **Active-pursuit goals** zeichnen sich dadurch aus, dass sie eine bestimmte Veränderung in der Umwelt oder im Selbst eines Akteurs herbeiführen sollen: “[Active-pursuit goals] represent the kinds of things one wants to get done“ ([OCC88], S. 41). Dies beinhaltet insbesondere die Befriedigung bestimmter (biologischer) Bedürfnisse, die Aufrechterhaltung eines spezifischen Status quo und die Herbeiführung eines neuen Umweltzustandes, der der Erfüllung bestimmter Wünsche und dem Verlangen eines Akteurs dient etc.
- **Interest goals** sind die Ziele eines Akteurs, die von genereller, mittelfristiger Natur sind und auf die ein Akteur in den meisten Fällen keinen unmittelbaren Einfluss

nehmen kann. “[interest goals] represent the kinds of things one wants to see happen“ ([OCC88], S. 41). Dazu gehört beispielsweise die Absicht, gesund zu bleiben genauso wie der Wunsch, dass die favorisierte Fußballmannschaft ihr nächstes Spiel gewinnen möge.

- **Replenishment goals** stellen eine Unterklasse der aktiv verfolgten Ziele dar, unterscheiden sich jedoch dadurch, dass sie, nachdem sie einmal erreicht wurden, weiterhin aufrecht erhalten werden und somit zyklischer Natur sind ([OCC88], S. 42). Zu dieser Kategorie zählen z.B. das regelmäßige Nachfüllen des Benzintanks oder der Bezug einer monatlichen Gehaltszahlung.

Standards stellen [OCC88] zufolge gesellschaftlich normierte, moralische und ethische Prinzipien und Richtlinien dar, an die sich Akteure bei der Bewertung des eigenen Verhaltens und dem Verhalten anderer orientieren. Standards und soziale Normen werden sozial erlernt und dienen in erster Linie der Erfüllung gesellschaftlicher und nicht individueller Ziele. Standards reflektieren im Einschätzungsprozess also die Einbettung des Akteurs in seine soziale Umwelt und werden vor allem bei der Entstehung komplexer sozialer Emotionen wie etwa Scham oder Stolz relevant.

Einstellungen (*attitudes*) werden als relativ dynamische und ungebundene Vorlieben oder Abneigungen gegenüber physischen Objekten und deren Eigenschaften beschrieben. Gemäß dieser Definition wären beispielsweise Stil und Geschmack als Einstellungen zu bezeichnen. Zwischen Einstellungen auf der einen und Zielen und Standards auf der anderen Seite besteht Ortony et al. zufolge keinerlei Zusammenhang (vgl. [OCC88], S. 46).

Das Netz aus Abb. 6.6 kann nun durch die lokalen Variablen ergänzt werden, so dass die Einschauungskriterien der jeweiligen Emotionsklassen integriert werden ([OCC88], S. 58).

Die Netze aus den Abb. 6.8 und 6.9 können kombiniert werden, so dass der Einschätzungsprozess in einem Netz abgebildet wird.

Unter Anwendung des Referenznetze-Formalismus kann man das ursprüngliche Modell aus Figur 6.5 vereinfachen. Die verschiedenen Klassen von Stimuli (*event, action, object*) werden durch drei verschiedene Farben repräsentiert, die der Stelle **Stimulus** zugewiesen sind. Entsprechend werden Ziele, Standards und Einstellungen als Farben (*goals, standards, attitudes*) der Stelle **actor's goals, standards, attitudes** zugewiesen. Der Einschätzungsprozess wird über einen synchronen Kanal an der Transition **appraise** initiiert, indem, in Abhängigkeit von den Werten der Parameter (x,y) , eine der Transitionen **appraise desirability**, **appraise praiseworthiness** oder **appraise attraction** aufgerufen wird.

Die Einschätzung kann für jede Emotionsklasse unter Berücksichtigung des in Abb. 6.7 dargestellten Modells der OCC-Einschätzungsprozesse detaillierter repräsentiert werden. Das Netz in Abb. 6.11 zeigt eine entsprechende Verfeinerung.

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

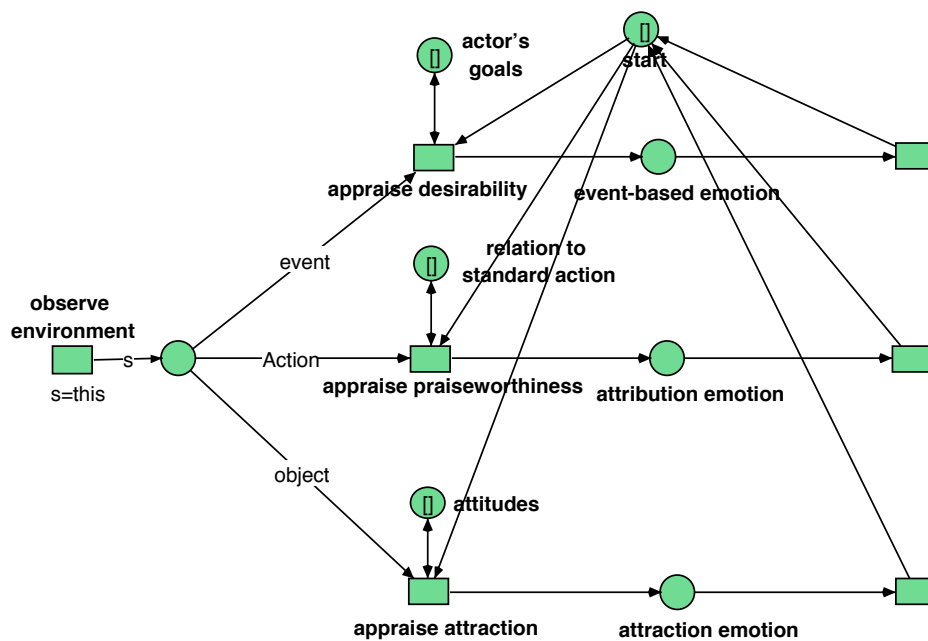


Abbildung 6.9.: Emotionseinschätzung nach OCC-Theorie unter Berücksichtigung der lokalen Variablen.

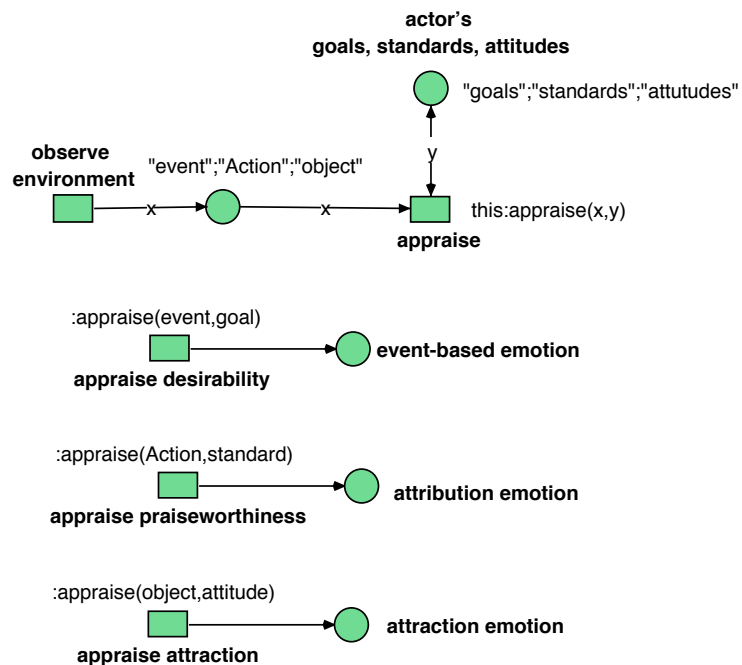


Abbildung 6.10.: OCC als Referenznetz.

6.1. Psychologische Theorien der Emotion

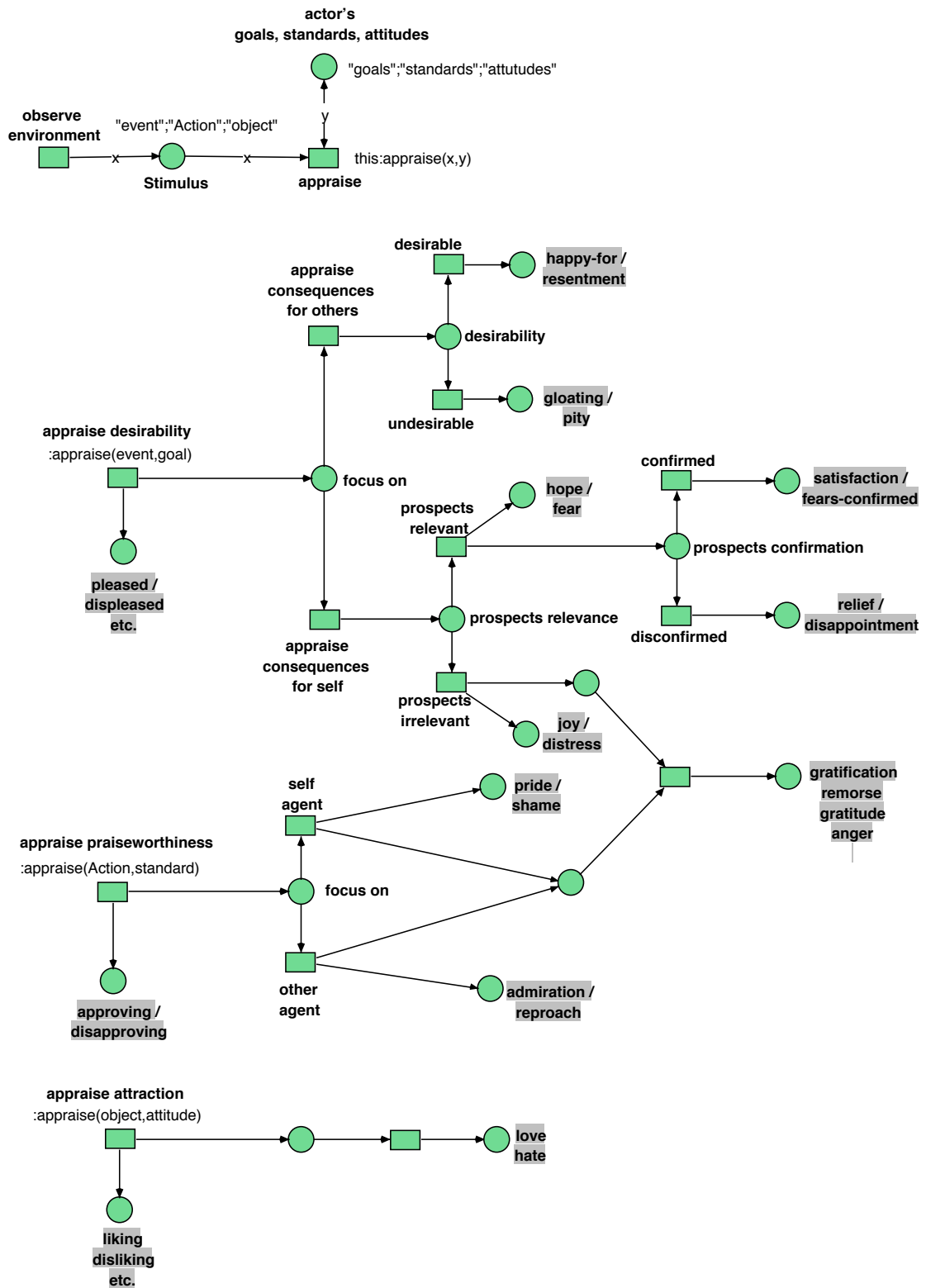


Abbildung 6.11.: OCC als Referenznetz: Verfeinerung im Bezug auf die Einschätzungsprozesse.

Auswertung der OCC-Theorie aus Sicht der informatischen Modellierung

Ein großer Vorteil des Modells von Ortony und Kollegen aus Sicht der Informatik ist seine leichte Umsetzbarkeit in ein regelbasiertes System, wobei Emotionen aufgrund der kognitiven Einschätzungsregeln von der Form

“ if <condition> then <action>“

generiert werden. Die linke Seite der Regel prüft entweder die Erwünschtheit *desirability* einer Ereigniskonsequenz, oder die Lobwürdigkeit *praiseworthiness* einer Agentenaktion oder die Attraktivität *appealingness* eines Objekts. Die rechte Seite der Regel setzt die Generierung eines entsprechenden emotionalen Zustandes in Gang.

Dank ihrer guten Operationalisierbarkeit wurde die OCC-Theorie bisher besonders oft in den informatischen Emotionsarchitekturen umgesetzt: Nach Staller und Petta (2000) ist sie bis heute “the most popular “reference model“ of appraisal used in agent architectures“ ([SP00], S. 5).

Ein Beispiel der Implementierung des OCC-Modells als regelbasiertes System stellt ein Framework für Multiagentensysteme von Bazzan et al. (2002) dar. Die objektfundierte Emotion des Mögens (*liking*) wird in diesem Modell folgendermaßen definiert (nach [BAB02], S. 2):

$A(p, o, t)$ sei die Attraktivität eines Objekts, die die Person p einem Objekt o zum Zeitpunkt t zuschreibt. Ferner sei $P_M(p, o, t)$ das Potenzial, einen emotionalen Zustand des Mögens zu generieren, $G(V_{g1}, \dots, V_{gn})$ eine Kombination der globalen Intensitätsvariablen, $I_M(p, o, t)$ die Intensität des Mögens, $T_M(p, t)$ der Schwellenwert und $f_M(\cdot)$ eine fürs Mögen spezifische Funktion. Dann sieht die Regel zur Generierung des emotionalen Zustandes des Mögens folgendermaßen aus:

$$\begin{aligned} &\text{IF } P_M(p, o, t) > T_M(p, t) \\ &\text{THEN set } I_M(p, o, t) = P_M(p, o, t) - T_M(p, t) \\ &\text{ELSE set } I_M(p, o, t) = 0 \end{aligned}$$

Diese Regel wird durch eine andere ausgelöst:

IF $A(p, o, t) > 0$ THEN set $P_M(p, o, t) = f_M(A(p, o, t), G)$

Ein wesentliches Problem für die Modellierung der OCC-Theorie besteht im Fehlen der genauen Spezifizierung der qualitativen / quantitativen Ausprägung von Einschätzungsprozessen: In [OCC88] sind keine Formalisierungen für alle definierten Emotionen gegeben, vielmehr wird eine Reihe von Beispielen präsentiert. Ortony et al. behaupten jedoch, dass jede Emotion in entsprechend formaler Notation formuliert werden kann, wenn sich dies bei vielen Emotionen auch als sehr komplex erweist.

Die genaue Spezifikation von zahlreichen implementationsrelevanten Details der Emotionsberechnung wird in der OCC-Theorie nicht explizit beschrieben. Dies bereitet Probleme bei der Umsetzung des Einschätzungsprozesses in informatische Modelle. Zum Beispiel berichten Bazzan et al. von ihren Schwierigkeiten, eine geeignete spezifische Generierungsfunktion f für jede Emotion zu finden ([BAB02], S. 3). An dieser Stelle müssen oft weitere Emotionstheorien hinzugezogen werden, um den Einschätzungsprozess genauer zu beschreiben.

Auf der allgemeinen Ebene ergab die Modellierung der Emotionstheorie von Ortony et al. folgende Fragen:

1. Disjunktivität der Emotionsklassen

Die Petrinetz-Modellierung deckte eine Reihe von Mehrdeutigkeiten auf, die in den Formulierungen der Theorie bezüglich der Differenzierung von drei Emotionsklassen enthalten sind.

Die OCC-Theorie liefert keine eindeutige Aussage darüber, ob es sich bei der Zuweisung einer Emotion zu einer Emotionsklasse um eine Entweder-Oder-Beziehung handelt. In Abb. 6.4 wird z.B. eine Reihe von Emotionen (*gratification*, *remorse*, *gratitude*, *anger*) durch beides - die Einschätzung der Erwünschtheit und die Einschätzung der Lobwürdigkeit generiert. Dadurch können diese Emotionen weder der Klasse der ereignisbasierten Emotionen noch den aktEURfundierten Emotionen zugeordnet werden. Ortony et al. erläutern, dass ereignisfundierte Emotionen häufig gemeinsam mit handlungsfundierten Emotionen auftreten, weil man Ereignisse mit der Handlung und ihrem Urheber in Verbindung bringt. Zum Beispiel kann das Verhalten eines anderen Akteurs, je nach dem kognitiven Fokus der Person, einmal als Handlung eines Akteurs oder ganz allgemein als Ereignis aufgefasst werden, wobei dann jeweils aktEUR- bzw. ereignisfundierte Emotionen erlebt werden. In diesem Fall wird jedoch nicht klar, ob es sich bei den generierten Emotionen um aktEURs- bzw. ereignisfundierte Emotion oder gar um eine "Mischemotion" handelt. Letzteres würde bedeuten, dass die drei Emotionsklassen (ereignisfundiert, aktEURfundiert und objektfundiert) nicht disjunkt sind, was einer grundlegenden Aussage der OCC-Theorie widerspricht.

2. **Klassifizierung der Stimuli** In den Netz-Modellen der OCC-Theorie wurde die Klassifizierung in Ereignis, Akteur oder Objekt durchgehend nicht auf die Stimuli bezogen, sondern auf deren Interpretation. Die verschiedenen Interpretationsmöglichkeiten eines Stimulus und die damit verbundenen unterschiedlichen Einschätzungsarten wurden durch alternative Netzpfade in Petrinetzen bzw. synchrone Kanäle in Referenznetzen realisiert.

Die Ereignis / Akteur / Objekt- Klassifizierung bezieht sich hierbei nicht auf die Art des emotionsauslösenden Stimulus, sondern auf die Dimensionen, anhand derer die Entitäten dieses Klassifizierungssystems eingeschätzt werden. Der Stimulus wird dementsprechend je nach (Situations-)Kontext bzw. persönlichem (Aufmerksamkeits-)Fokus als "Akteur", "Ereignis" oder "Objekt" in die Umgebung eingebettet und dementsprechend bezüglich seiner Konsequenzen, Aktionen oder Eigenschaften eingeschätzt. Wichtig für den Einschätzungsprozess und damit für die resultierende Emotion ist dann nicht die Identifizierung des Stimulus als Ereignis, Akteur oder Objekt, sondern die jeweilige Dimension der Einschätzung, nämlich *praiseworthiness*, *desirability* bzw. *attractiveness*.

3. **Gleichzeitige Generierung mehrerer (widersprüchlicher) Emotionen** Ein weiteres Modellierungsproblem ist damit verbunden, dass die OCC-Theorie keine eindeutige Aussage darüber macht, ob die verschiedenen generierten Emotionen gleichzeitig bestehen können oder sich gegenseitig ausschließen. Insbesondere wird nicht klar, ob die Emotionen einer Emotionsklasse, die im Laufe der Einschätzungsprozesses

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

durch weitere Emotionen gleicher Klasse ergänzt bzw. erweitert werden, parallel zueinander existieren oder einander ablösen. Im Falle der ereignisfundierten Emotionen (vgl. Abb. 6.11) wird z.B. bei der Ereignisseinschätzung die Emotion der Zufriedenheit (*pleased*) im ersten Einschätzungsschritt (Transition **appraise desirability**) generiert, die im weiteren Verlauf des Einschätzungsprozesses je nach Ergebnis der weiteren Einschätzungsschritte (Stellen **focus on**, **prospects relevance**, **prospects confirmation**) durch weitere Emotionen (Stellen **hope / fear**, **joy / distress**, **relief / disappointment** usw.) ergänzt bzw. spezifiziert wird. Ebenso wird aus der OCC-Theorie nicht klar, wie sich Emotionen verschiedener Klassen (z.B. ereignis- vs. akteurfundierte Emotionen) zueinander verhalten und ob in diesem Fall ein gegenseitiger Ausschluss oder eine nebenläufige Emotionsgenerierung zu modellieren ist.

4. Zeitlicher Ablauf der Einschätzungen

Damit verbunden ist die Frage, ob die Einschätzungsprozesse für verschiedene Stimuli parallel zueinander laufen können. Die Forderung nach einer diskreten Ausführung der Einschätzungen ist in der OCC-Theorie nicht explizit enthalten, sondern kann lediglich aus der Aufteilung der Stimuli in drei Klassen bezüglich der Einschätzung und aus einigen sprachlichen Formulierungen gefolgert werden (vgl. [BodAHN03]).

Die Diskretheit der Emotionsgenerierung konnte netztechnisch durch die Einführung zusätzlicher Synchronisationskonstrukte (vgl. z.B. Transition **decay** in Abb. 6.6) dargestellt werden, durch die eine simultane Ausführung der Einschätzungen ausgeschlossen wird. Auf diese Weise können nie zwei verschiedene Emotionen zur gleichen Zeit generiert werden. Erst wenn eine Emotion abklingt, wird eine neue Einschätzung vorgenommen. Durch das Weglassen dieser Konstrukte kann eine nicht diskrete Ausführung von Einschätzungen realisiert werden.

5. Zeitlicher Ablauf der erzeugten Emotionen

Des Weiteren macht die OCC-Theorie keine Angaben über die Dauer respektive den zeitlichen Verlauf einer Emotion erzeugten Emotionen. Reilly (1996) beschäftigte sich mit dem zeitlichen Verhalten und stieß dabei auf zwei konzeptionelle Probleme [Rei96]: Erstens scheint die Zeitlichkeit einer Emotion unter anderem von dem Ziel abzuhängen, das der Agent verfolgt und das der Grund für die Emotion ist. Zweitens scheinen sich die Emotionen auch in ihrem zeitlichen Verlauf gegenseitig zu beeinflussen (vgl. [BodAHN03]).

Die Untersuchung, ob in der OCC-Theorie ein bestimmter zeitlicher Ablauf von Einschätzungen bzw. der generierten Emotionen (nebenläufig bzw. sequenziell) implizit vorausgesetzt wird, ist im Rahmen dieser Arbeit nicht möglich.

In Bezug auf unser Forschungsziel der Bewertung der Petrinetze als Modellierungstechnik für sozialwissenschaftliche Theorien können wir jedoch festhalten, dass sich offensichtlich beide Möglichkeiten mit dem Petrinetzformalismus modellieren ließen (vgl. z.B. Abb. 6.5 und 6.6). Ferner ließ die Darstellung der Theorieaussagen mit Netzmodellen die in der OCC-Theorie enthaltene Mehrdeutigkeit bezüglich der Differenzierung der Emotionsklassen, des kausalen Ablaufs der Einschätzungsprozesse und der generieren Emotionen zunächst deutlich zum Vorschein kommen.

Es existieren mehrere Systeme, die auf der OCC-Theorie aufbauen. Insbesondere wird das OCC-Modell zur Implementation der Agentenemotionen in zwei der bekanntesten Multiagentenarchitekturen umgesetzt: dem *Affective Reasoner* ([Ell92]) und der Emotionskomponente Em als Teil des Oz-Projekt ([Bat94], [Rei96]) (siehe Kapitel 7 für eine detaillierte Übersicht über die beiden Systeme).

Beide Architekturen haben die OCC-Theorie nicht in ihrer ganzen Komplexität implementiert, was zum Teil auch darauf zurückzuführen ist, dass die OCC-Theorie einige konzeptionelle Probleme beinhaltet, die einer umfassenden Implementierung im Wege stehen. Große Probleme bereitete beiden Autoren die Möglichkeit des gleichzeitigen Auftretens mehrerer widersprüchlicher Emotionen (s.o.). Da diese sich in beiden Systemen jedoch weder bei ihrer Entstehung noch in ihren zeitlichen Verlauf gegenseitig beeinflussen, wurden am Ende Ad-hoc-Lösungen verwendet, um ein konsistentes und nachvollziehbares Agentenverhalten im Sinne eines *Broad-but-shallow*-Ansatzes (s.u.) hervorbringen zu können (siehe Kapitel 7.3.1).

Da die OCC-Theorie lediglich den Prozess der Emotionsgenerierung betrachtet und hinsichtlich der Funktionen der generierten Emotionen innerhalb des Gesamtsystems keine Aussagen macht, müssen zu ihrer Umsetzung – z.B. in einem Agentensystem – weitere Theorien hinzugezogen werden. Bei vielen Umsetzungen der OCC-Theorie werden an dieser Stelle intuitive Konstrukte der Entwickler eingesetzt, um die fehlende Spezifizierung der Auswirkungen der Emotion auf das Agentenverhalten zu ersetzen: “[R]eification of emotions as identifiable system components [...] engenders the problem of how to proceed from these emotions for further system processing, leading to the adoption of ad-hoc constructions of dubious validity“ ([SP00], S. 6).

Die Einschätzungstheorie von Scherer

In der Theorie von Scherer wird eine spezifische Anzahl von Einschätzungsdimensionen definiert, anhand derer Ereignisse vom Akteur eingeschätzt werden. Die Ausdifferenzierung und die Bestimmung der Emotionen wird durch den Prozess der kognitiven Ereignisseinschätzung beschrieben, wobei als Einschätzungskriterien die Bedeutung des Ereignisses für die individuellen Ziele und Wertestrukturen (vgl. *Beliefs* und *Desires* eines Agenten), das persönliche Handlungspotenzial (vgl. *coping* [Laz91]) und die sozial-normative Ereignissignifikanz eine Rolle spielen [Sch99a].

Scherer definiert Emotion als “eine Episode zeitlicher Synchronisation aller bedeutender Subsysteme des Organismus, die fünf Komponenten bilden (Kognition, physiologische Regulation, Motivation, motorischer Ausdruck [motor expression] und Monitoring / Gefühl), und die eine Antwort auf die Bewertung eines externalen oder internalen Reizereignisses als bedeutsam für die zentralen Bedürfnisse und Ziele des Organismus darstellt“ ([Sch93], S. 4; siehe auch Kapitel 5.7).

Es werden folgende fünf funktionalen Subsysteme unterschieden, welche durch emotionale Prozesse kontrolliert werden [Sch93]:

1. Ein kognitives Subsystem (*cognitive component*) evaluiert den Reiz durch Wahrnehmung, Gedächtnis, Vorhersage und Evaluation verfügbarer Informationen.
2. Ein neurophysiologisches Subsystem (*neurophysiological component*) reguliert den internen Zustand durch die Kontrolle von neuroendokrinen, somatischen und autonomen Zuständen.
3. Ein motivationales Subsystem (*motivational component*) plant, bereitet Aktionen vor und wählt zwischen konkurrierenden Motiven aus.
4. Ein handelndes Subsystem (*expressive component*) kontrolliert motorischen Ausdruck und sichtbares Verhalten.
5. Ein Monitor-Subsystem (*monitoring / feeling component*) schließlich kontrolliert die Aufmerksamkeit, die den gegenwärtigen Zuständen zuteil wird und leitet das resultierende Feedback an die anderen Subsysteme weiter.

Diese Subsysteme funktionieren normalerweise relativ unabhängig voneinander. In Notfallsituationen bzw. in emotionalen Episoden (*emotion episodes*) werden sie jedoch aufeinander abgestimmt.

In diesem Zusammenhang ist das kognitive Subsystem (*cognition*) nach Scherer von besonderem Interesse [Sch93]: Nach Scherer’s Definition wird eine emotionale Episode grundsätzlich als Ergebnis einer *kognitiven Stimulus-Einschätzung*, d.h. der kognitiven Auswertung der Eigenschaften eines Stimulusereignisses aufgefasst. Die Emotion erfolgt als Reaktion auf ein externes oder internes Stimulusereignis, dessen Relevanz für zentrale Ziele des Organismus als hoch eingeschätzt wird. Dieser Sachverhalt ist in Abb. 6.12 festgehalten. Die Struktur des zugehörigen Petrinetzes entspricht der Struktur des Netzes aus Abb. 5.7, das eine allgemeine Emotionsdefinition nach Scherer beschreibt (siehe auch Kapitel 5.4).

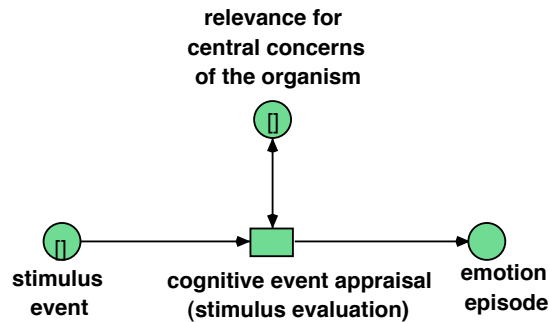


Abbildung 6.12.: Emotionale Episode als Folge der Ereigniseinschätzung.

Eine emotionale Episode erfordert die Mobilisierung der substantiellen Ressourcen des Organismus, um eine Adaption oder eine aktive Reaktion auf den veränderten internen Zustand oder die externe Situation zu ermöglichen. Sie besteht nach Scherer's Definition in einer temporären Synchronisation aller funktionalen Subsysteme ([Sch93], [Sch99a]). Die emotionale Episode (vgl. Abb. 6.13) beginnt mit der Synchronisierung der Subsysteme nach einem speziellen Muster (Transition synchronize), der von der Stimuluseinschätzung infolge des Einschätzungsprozesses (Transition cognitive event appraisal) abhängt, und endet mit dem Rückzug der Subsysteme auf ihr voneinander unabhängiges Funktionieren (Transition return to independent functioning of subsystems).

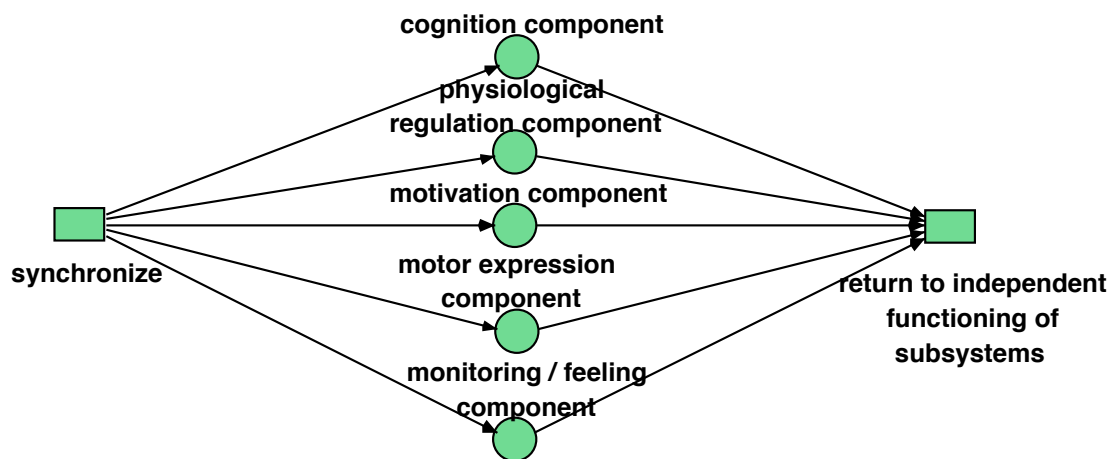


Abbildung 6.13.: Emotionale Episode als Synchronisierung aller Subsysteme.

Scherer liefert somit zwei Arten von Definitionen für eine emotionale Episode, die durch Netzmodelle explizit dargestellt werden konnten: Einerseits beschreibt eine strukturelle Definition eine emotionale Episode als Ergebnis der kognitiven Stimuluseinschätzung (vgl. Abb. 6.12) und andererseits stellt eine prozedurale bzw. funktionale Definition eine emotio-

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

nale Episode als temporäre Synchronisation aller organischen Subsysteme dar (vgl. Abb. 6.13). Dem Zusammenhang zwischen diesen zwei Definitionen wird Rechnung getragen, indem sie in Abb. 6.14 integriert dargestellt werden. Die Integration der beiden Definitionen kann durch die Verfeinerung der Stelle **emotion episode** aus Abb. 6.12 in Bezug auf den temporären Synchronisationsprozess erfolgen (Abb. 6.13).

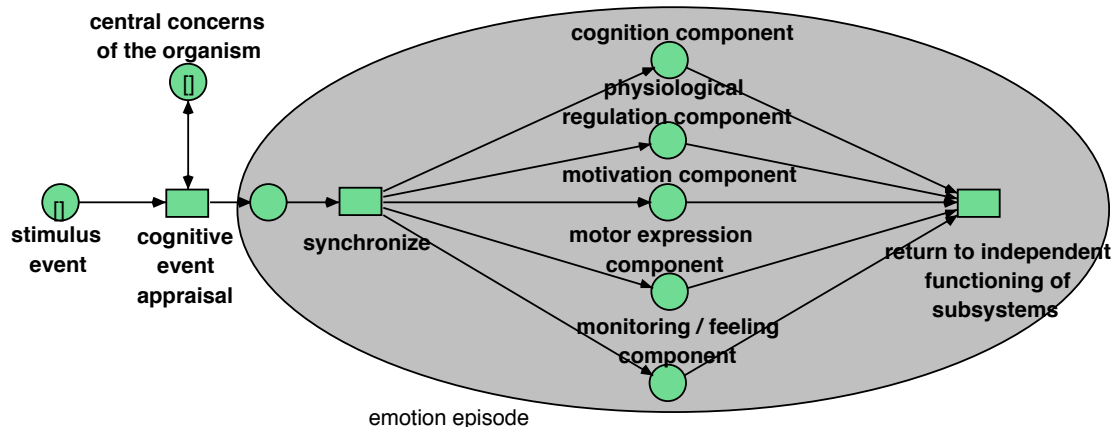


Abbildung 6.14.: Definition der emotionalen Episoden als Momente der Synchronisation aller Subsysteme.

Scherer nimmt an, dass das Ergebnis der Stimulusauswertung jedes Subsystem direkt beeinflusst. Wegen der engen Interaktion der Subsysteme während einer emotionalen Episode ist die Regulation sehr komplex, sie wird durch multiple Feedback- und Feedforward-Prozesse gewährleistet. Daraus resultiert nach Scherer eine unüberschaubare Anzahl hoch differenzierter emotionaler Zustände, so dass die aktuell registrierte Emotion nur eine zentrale Tendenz bzw. einen Cluster der regulär wiederkehrenden modalen Zustandsmuster (*modal states*) darstellt. Diese modalen Zustandsmuster werden in Scherers Emotionstheorie durch die gängigen Label für Emotionen nur im Sinne von zentralen Tendenzen beschrieben und weisen eine weit darüber hinausgehende Vielfalt auf.

Der Theorie von Scherer zufolge werden Emotionen durch eine Sequenz von Einschätzungen, die „*stimulus evaluation checks*“ (SEC) ausgelöst, die jeder Stimulus im Zuge der Informationsverarbeitung durchläuft ([Sch93], [Sch99b], [Sch99a], [Sch01]).

Die SECs laufen nach Scherer sequenziell ab und prüfen bestimmte Eigenschaften eines Stimulus, wobei zeitlich spätere Einschätzungen auf die Ergebnisse vorangegangener Checks angewiesen sind. Für den sequenziellen Ablauf einzelner Einschätzungen sprechen erste empirische Untersuchungen [Sch99a].

Scherer unterscheidet innerhalb des kognitiven Subsystems fünf wesentliche *stimulus evaluation checks* (SEC), von denen vier weitere *subchecks* besitzen:

- Der *novelty check* bewertet, ob sich äußere oder innere Reize verändert haben; seine *subchecks* sind Plötzlichkeit (*suddenness*), Vertrautheit (*familiarity*) und Vorhersagbarkeit (*predictability*).

6.1. Psychologische Theorien der Emotion

- Der *intrinsic pleasantness check* bestimmt, ob der Reiz angenehm oder unangenehm ist, und verursacht entsprechende Annäherungs- oder Vermeidungstendenzen.
- Der *goal significance check* stellt fest, ob das Ereignis für die Ziele der Person förderlich oder hinderlich ist; seine *subchecks* sind Relevanzüberprüfung (*concern relevance*), Wahrscheinlichkeitsbewertung (*outcome probability*), Erwartungskonsistenz (*expectation / discrepancy*), Förderlichkeitsüberprüfung (*conductiveness*) und Dringlichkeit (*urgency*).
- Der *coping potential check* ermittelt, inwieweit die Person glaubt, die Ereignisse unter Kontrolle zu haben; seine *subchecks* sind Kausalitätsprüfung (*cause agent, cause motiv*), Kontrollpotenzial (*control*), Macht (*power*) und Anpassungsfähigkeit (*adjustment*).
- Der *compatibility with standards check* vergleicht das Ereignis mit inneren und äußeren Standards und Normen; seine *subchecks* prüfen die Kompatibilität mit externen Normen (*external standards*) und internen Vorgaben (*internal standards*).

Je nachdem, mit welchen Ergebnissen der sequenzieller Einschätzungsprozess die einzelnen SECs durchläuft, entstehen nach Scherer unterschiedliche Emotionen. Jede Emotion kann, so Scherer, eindeutig durch eine Kombination der SECs und *subchecks* bestimmt werden.

Der sequenzielle Ablauf der SECs ist in einem einfachen Petrinetz in Abb. 6.15 dargestellt.

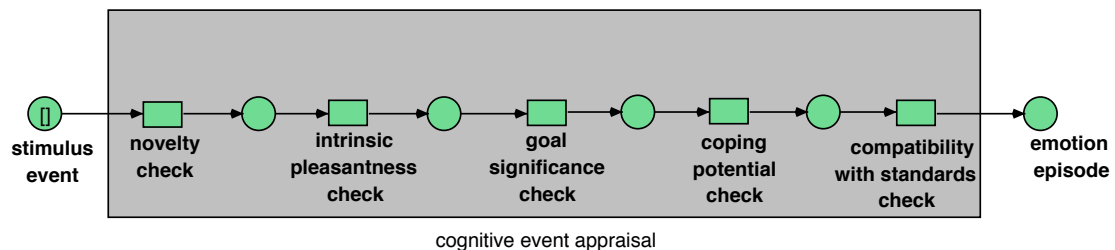


Abbildung 6.15.: Stimulus Evaluation Checks (SECs).

Das Resultat der SECs ist eine Emotion, die nach Scherers Emotionsdefinition wiederum Veränderungen in den anderen Subsystemen auslöst (vgl. Abb. 6.14). Die Transition **cognitive event appraisal** im Netz aus Abb. 6.14, in der eine emotionale Episode als Moment der Synchronisation aller Subsysteme dargestellt wird, kann durch das Netz aus Abb. 6.15 ersetzt werden. So kann der gesamte Ablauf der Emotionsgenerierung vom Stimulus bis zur emotionalen Episode in einem Netz dargestellt werden.

Den Ausgangspunkt eines Einschätzungsprozesses bildet ein internes oder externes Ereignis, das eine Zustandsänderung signalisiert, die entweder durch das Auftreten eines unerwarteten Ereignisses oder das Ausbleiben eines erwarteten Ereignisses ausgelöst werden kann.

Im ersten Schritt des Einschätzungsprozesses wird im Rahmen des *novelty check* bewertet, inwieweit ein Stimulus als neu bzw. unerwartet einzustufen ist. Der Ablauf des *novelty check*

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

wird in Abb. 6.16 als Petrinetz dargestellt. Das Netz aus Abb. 6.15, das den allgemeinen Ablauf der SECs zeigt, wird in Bezug auf den *novelty check* verfeinert (vgl. die gleichnamige Transition in Abb. 6.15).¹

Beim *novelty check* wird zunächst entschieden, ob das Ereignis gerade eingesetzt hat (Transition **suddenness**) und die Auslösung einer Adaptionsreaktion erfordert. Als Nächstes wird die Vertrautheit (Transition **familiarity**) des Ereignisses bewertet. Plötzlich auftretende unbekannte Ereignisse hätten somit ein großes Potenzial zur Auslösung einer Emotion. Diese Einschätzung kann sich unter Umständen auch auf die Geschwindigkeit der folgenden Einschätzungen auswirken, vor allem bei unbekannten Reizen [Sch84]. Ein letzter Subcheck prüft die Vorhersagbarkeit eines Ereignisses (Transition **predictability**).

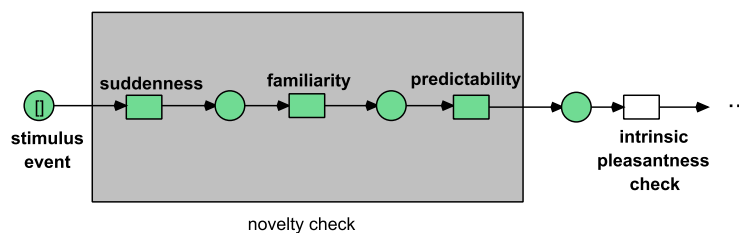


Abbildung 6.16.: Spezifizierung des ersten Einschätzungsschrittes (Novelty Check).

Der zweite Einschätzungsschritt *intrinsic pleasantness check* bezieht sich auf die intrinsische, affektive Valenz eines Reizes und entspricht in etwa der Einschätzung der Attraktivität eines Objekts (*Appealingness*) in der Theorie von [OCC88] (s.o.). Die Bewertung bezieht sich auf dem Stimulus *inhärente* Eigenschaften (im Gegensatz etwa zu den für die Zielerreichung relevanten Eigenschaften). Trotz der Bedeutung inhärenter Faktoren basiert diese Einschätzung auf angeborenen Merkmalsdetektoren ebenso wie auf erlernten Eigenschaften ([LS87], S. 15; [CSC94], S. 352; [Sch84]).

Die Überprüfung, ob das Ereignis *intern* oder *extern* ist, bildet einen Teil dieses Einschätzungsschrittes. Diese Unterscheidung ist insofern nicht trivial, als bei einer speziellen Art der Repräsentation von Objekten, den sog. *Selbstobjektrepräsentanzen*, ein Objekt der Außenwelt so behandelt wird, als sei es ein Teil des Selbst. Ereignisse, die dem Selbstobjekt zugeordnet werden, sind zwar de facto externe Ereignisse, werden aber als intern evaluiert.

Der *intrinsic pleasantness check* besitzt keine Subchecks; er wird im Netz in Abb. 6.17 dargestellt.

Der dritte SEC überprüft die Bedeutung eines Ereignisses hinsichtlich der Ziele, Wünsche und Bedürfnisse des Akteurs (*goal/need significance check*). Dieser Einschätzungsschritt besteht aus fünf Subchecks (vgl. Abb. 6.18):

1. Die Einschätzung der Relevanz des Ereignisses in Bezug auf die Handlungspläne und Ziele (Transition **concern relevance**) leitet den eigentlich zentralen Teil des Bewertungsprozesses ein. Sie besteht darin, dass aus der Menge der gerade verfolgten Ziele

¹Entsprechend wird im Folgenden mit den weiteren vier SECs verfahren, d.h. der jeweils betrachtete SEC wird im Rahmen des Gesamtablaufs des Einschätzungsprozesses (Abb. 6.15) verfeinert dargestellt.

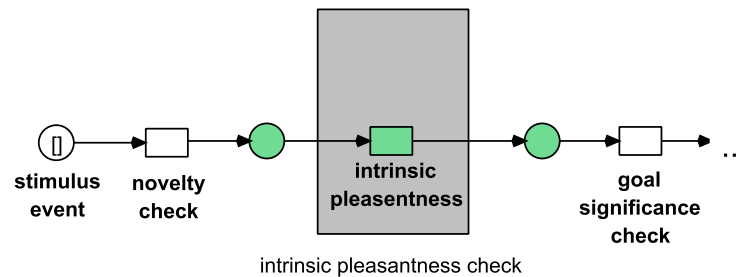


Abbildung 6.17.: Spezifizierung des 2. Einschätzungsschrittes (Intrinsic Pleasantness Check).

oder Bedürfnisse diejenigen ausgewählt werden, die von dem Ereignis betroffen sind. Zusätzlich wird unterschieden, ob das Ereignis Konsequenzen nach sich zieht, die für einen selbst oder für repräsentative Andere relevant sind.

2. Die Art der Konsequenzen, insbesondere die Wahrscheinlichkeit ihres Eintretens und Wirksamwerdens, geht als weitere Größe in den Einschätzungsprozess ein (Transition *outcomes probability*).
3. Durch Einschätzung der Erwartungskonsistenz (Transition *expectation discrepancy*) eines Ereignisses wird seine Erwartungskonformität bewertet. Die Erwartungskonsistenz wird als Subcheck in den Einschätzungsprozess eingeführt, da es z.B. sein kann, dass ein zielbehinderndes (non-conductiv) Ereignis wiederholt zu einem bestimmten, nun vorhersehbaren Umgebungszustand führt. Wenn dies nun in einem Fall entgegen der Erwartung des Akteurs nicht eintritt, kommt es z.B. statt zu Resignation zu Freude [Sch88].
4. Als vierter Subcheck liefert die Förderlichkeitsüberprüfung (Transition *conductiveness*) Aussagen darüber, ob die momentan verfolgten Ziele durch das eingeschätzte Ereignis direkt gefördert bzw. behindert werden.
5. Der letzte Subcheck - der Dringlichkeitscheck (Transition *urgency*) - prüft, ob das Ereignis eine sofortige Reaktion erfordert (vgl. [LS87], S. 15).

Abb. 6.18 zeigt den *goal/need significance check* mit seinen Subchecks als Netz.

Scherer unterscheidet bewusst zwischen der Zielförderlichkeit und der intrinsischen Valenz eines Objekts, denn auch intrinsisch zunächst angenehme Objekte oder Ereignisse können letztendlich die Erreichung eines Ziels blockieren und aus diesem Grund negativ eingeschätzt werden. Der Zahnarztbesuch beispielsweise birgt intrinsisch unangenehme affektive Reaktionen, ist jedoch dem Ziel des Erhalts der Gesundheit in der Regel förderlich (vgl. auch [CSC94], S. 352).

Eine ähnliche Möglichkeit sieht auch die OCC-Theorie ([OCC88]) vor, bei der eine auf Zielen basierende Einschätzung durchaus mit der Einschätzung des intrinsischen Wohlbefindens

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

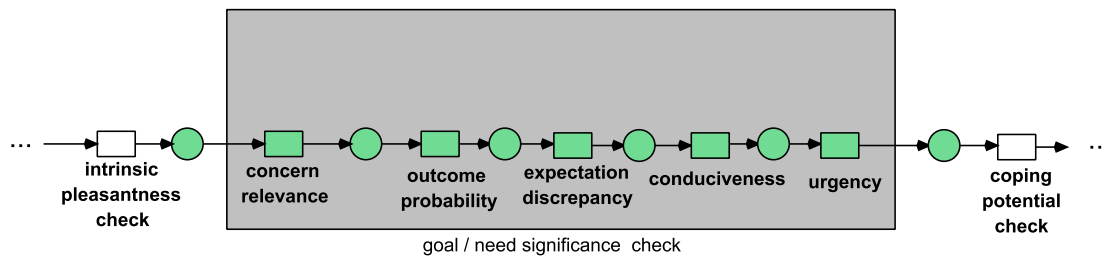


Abbildung 6.18.: Spezifizierung des 3. Einschätzungsschrittes (Goal Significance Check).

kollidieren kann (vgl. [CSC94]). Auch weitere von Scherer spezifizierte Einschätzungsdimensionen (z.B. Ereigniswahrscheinlichkeit, Erwartungsrelevanz- und konsistenz) findet man in ähnlicher Form in der OCC-Theorie ([OCC88], s.o.) wieder.

Der nächste Schritt im Bewertungsprozess (Abb. 6.19) bestimmt das Bewältigungspotenzial (*coping potential check*) und beinhaltet wiederum fünf Subchecks [Sch93]:

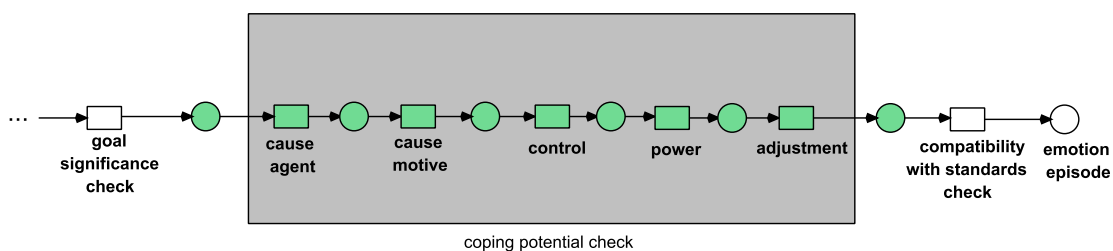


Abbildung 6.19.: Spezifizierung des 4. Einschätzungsschrittes (Coping Potential Check).

- Die Kausalitätsprüfung (Transition **cause agent**) bewertet die Ursache des Ereignisses, d.h. wer oder was für das Ereignis verantwortlich ist. Unterschieden wird dabei hauptsächlich zwischen dem Selbst als Verursacher (interne Kausalität), anderen (externe Verursachung) und einer Verursachung durch "natural agents" (Umstände, Zufall usw.). Dies entspricht in etwa der Einschätzung des Ereignisfokus in der Theorie von Ortony et al. (6.5), wobei jedoch in der OCC-Theorie nur die Möglichkeit der Verursachung durch das Selbst oder andere gegeben ist.
- Die Frage nach den Gründen, die den Verursacher leiten (Transition **cause motive**), wurde von Scherer ebenfalls in die Bewertung miteinbezogen, da sie für den Einschätzungsprozess und sein Ergebnis vor allem im Hinblick auf Bewältigungsmöglichkeiten mindestens ebenso wichtig ist wie die Frage nach dem Verursacher.
- Das Kontrollpotenzial (Transition **control**) bewertet, ob die Konsequenzen eines Ereignisses absehbar sind und ob sie kontrolliert bzw. beeinflusst werden können. Die Überprüfung des Kontrollpotenzials bezieht sich, wie Scherer selbst ausführt, auf

die Energie, d.h. die unterschiedlichsten Arten von Ressourcen (materiell, kulturell, physiologisch), die einem Akteur zur Verfügung stehen.

- Die Macht (*power*) beinhaltet die Einschätzung der eigenen Kraft im Hinblick auf mögliche Hindernisse bei der Bewältigung / Beeinflussung der Konsequenzen eines Ereignisses. .
- Die Anpassungsfähigkeit (*adjustment*) bewertet, ob es für den Akteur möglich ist, eigene Ziele an irreversibel veränderte Umstände anzupassen.

Der fünfte und letzte Bewertungsschritt (*compatibility check*) vergleicht die *Kompatibilität* eines Ereignisses bzw. einer Handlung mit den Normen der Moral, Legitimität, und Gerechtigkeit (Abb. 6.20). Überprüft wird hierbei einerseits, ob eine eigene oder fremde Handlung den sozialen Normen, kulturellen Konventionen, Regeln oder Erwartungen signifikanter anderer entspricht (Transition **external standards subcheck**). Beim Transition **internal standards subcheck** wird eingeschätzt, inwieweit eine Handlung mit den eigenen Normen und Standards - als Teil des Selbstkonzepts oder Selbstideals - übereinstimmt. Diese Einschätzungsdimension erlangt Bedeutung vor allem im Zusammenhang mit "sozialen" Emotionen (vgl. Definition von Damasio in Kapitel 5.5) wie Verachtung, Scham, Schuld und Ärger.

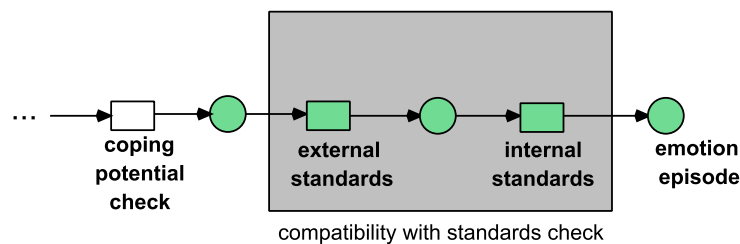


Abbildung 6.20.: Spezifizierung des 5. Einschätzungsschrittes (Norm / Self Compatibility Check).

Scherer geht davon aus, dass alle zu Emotionen fähigen Spezies über die ersten vier SECs verfügen, jedoch lediglich Menschen und möglicherweise einige wenige Arten höherer Säugetiere den Norm- und Selbstkompatibilitätscheck durchführen ([Sch84], S.307f.; siehe auch [CSC94], S. 352).

Eine Reihe von empirischen Studien hat die einzelnen Aspekte von Scherers Theorie bislang unterstützt z. B. [Sch99b]), obwohl Scherer selbst darauf hinweist, dass die empirische Bestimmung von Bewertungsprofilen grundsätzlich problematisch ist, da die Bewertungsschritte meist sehr schnell, automatisch und unbewusst ablaufen und somit kein Zugang zu diesen Prozessen mit konventionellen Verifikationsmethoden, wie z.B. Fragebogen, hergestellt werden kann [Sch99a]. Alternativ wurde in der Geneva Emotion Research Group ein interaktives Computerspiel entwickelt, das eine bessere Auswertungsumgebung für die Untersuchung emotionaler Reaktionen als Resultat einzelner Bewertungsschritte bereitstellt.

Auswertung der Theorie von Scherer aus der Sicht der informatischen Modellierung

Scherer ([Sch93], [Sch99a]) definiert Emotionen als eine Episode temporärer Synchronisation aller Systeme des organischen Funktionierens, die er durch fünf funktionale Subsysteme repräsentiert (Kognition, physiologisches Arousal, Motivation, motorischer Ausdruck, subjektives Erleben). Die Synchronisation erfolgt als Reaktion auf ein externes oder internes Stimulusereignis, das für zentrale Ziele des Organismus relevant ist. Die von Scherer angenommenen Einschätzungsdimensionen stimmen zwar mit den Einschätzungsdimensionen anderer Emotionstheoretiker überein, jedoch werden diese in keiner der behandelten Theorien so differenziert und umfassend spezifiziert. Der sequentielle Ablauf der einzelnen Einschätzungsschritte wurde in [Sch99b] indirekt nachgewiesen und in [Sch01] theoretisch begründet.

Von den betrachteten psychologischen Emotionstheorien berücksichtigen lediglich die Theorien von Scherer (1993) und Frijda (1986) (s.u.) in der einen oder anderen Form explizit die Rolle der sozialen Phänomene auf verschiedenen Stufen des Emotionsprozesses.

Emotionstheorie von Lazarus

Emotionen sind nach Lazarus organisierte psychophysiologische Reaktionen auf Veränderung der Überzeugungen einer Person hinsichtlich der Bedeutung der Person-Umwelt-Beziehungen für ihr persönliches Wohlergehen. "Emotionen sind im Grunde organisierte kognitiv-motivational-relationale Konfigurationen, deren Status sich mit Änderungen der Person-Umwelt-Beziehung, wie sie wahrgenommen und bewertet wird, ändert" ([Laz91], S. 38; vgl. auch Kapitel 5.6).

Ausgangspunkt des Prozesses der Emotionsentstehung nach Lazarus ist die Person-Umwelt-Beziehung, wie sie von der Person konstruiert bzw. wahrgenommen wird (vgl. Abb. 6.21). In diesen Konstruktionsprozess fließen einerseits die objektiven *Fakten über die Umwelt* und andererseits subjektive Merkmale der Person ein, nämlich ihr *Wissen über die Ziele und Werte*, denen sie sich verpflichtet fühlt. Diese Faktoren, die als Grundlage für die Emotionseinschätzung dienen, stehen den Konzepten der BDI-Architektur sehr nahe: Allgemeinwissen eines Akteurs über seine Umwelt entspricht in etwa den *Beliefs* eines BDI-Agenten, während das Wissen um seine Ziele / Werte den *Desires* eines BDI-Agenten gleicht.

Die wahrgenommene Person-Umwelt-Beziehung wird in Bezug auf eine vergleichsweise kleine Anzahl spezifischer, genetisch festgelegter Gesichtspunkte bewertet, die nach Lazarus das Entdecken der grundlegenden Anpassungsprobleme ermöglichen, welche Emotionen auslösen. Genau genommen handelt es sich dabei um insgesamt sechs Bewertungsgesichtspunkte, wovon Lazarus jeweils drei primären und sekundären Bewertungsprozessen zuordnet. Für den Ablauf der einzelnen Schritte den primären bzw. den sekundären Bewertungsprozesse spezifiziert Lazarus keine feste kausale oder temporäre Reihenfolge. Vielmehr behauptet er, dass diese Prozesse sehr schnell und unwillkürlich verlaufen und sich gegenseitig beeinflussen. Darüber hinaus können viele der in der Vergangenheit gemachten Bewertungen durch geeignete Hinweisreize aus dem Gedächtnis abgerufen werden.

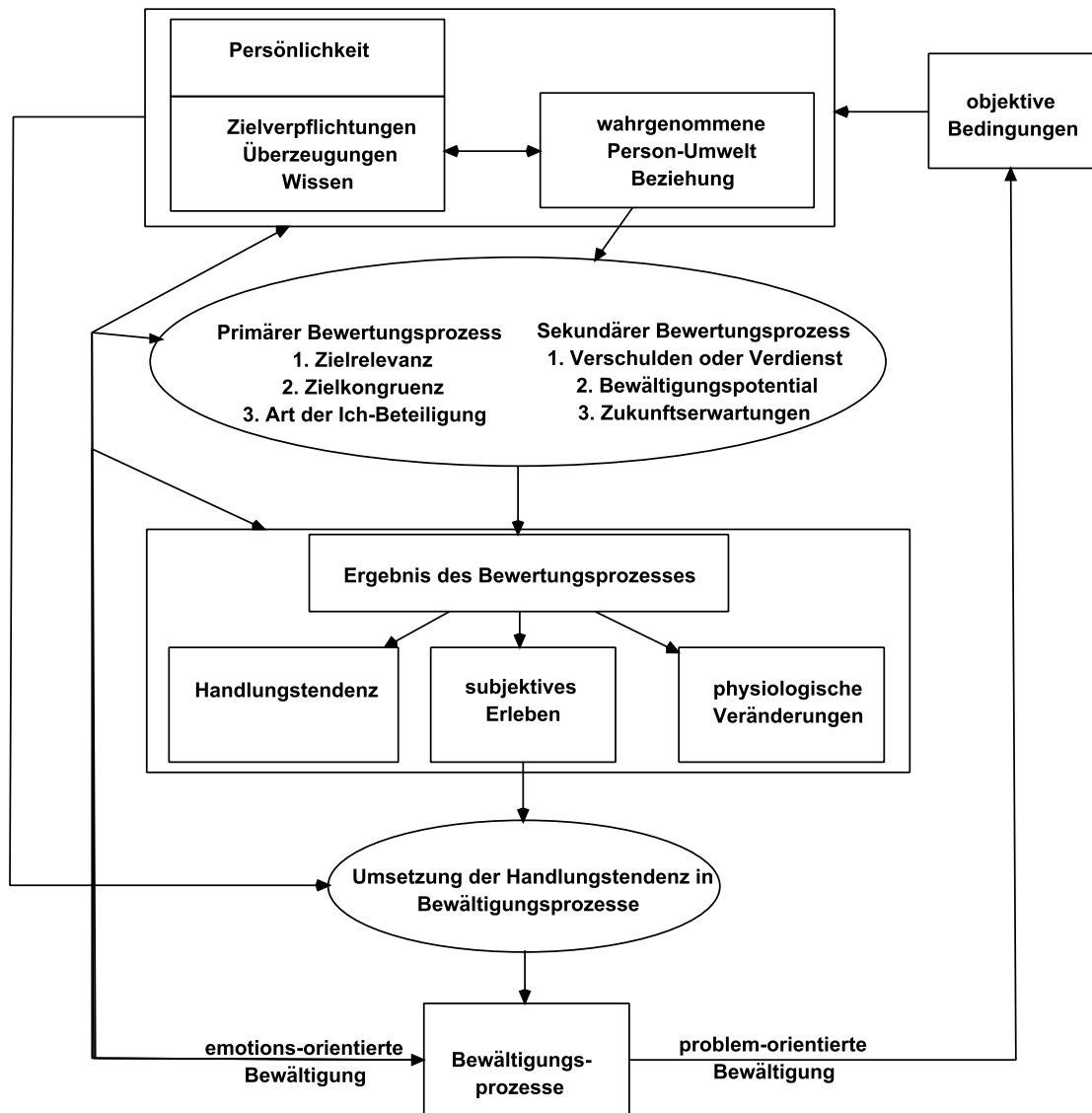


Abbildung 6.21.: Das kognitiv-motivational-emotionale System nach Lazarus (1991).

Wegen der unscharfen Definition der Ablaufreihenfolge der Bewertungen ist die prozessorientierte Darstellung der Emotionsgenerierung zunächst nicht direkt aus der Theorie von Lazarus ableitbar. Bei der Betrachtung der Bewertungsschritte im Einzelnen lässt sich jedoch feststellen, dass obwohl keine kausale Reihenfolge der Einschätzungen in der Theorie explizit spezifiziert wird, viele Bewertungsschritte aufeinander aufbauen und an manchen Stellen die Ergebnisse der vorangegangenen Bewertung für die weitere Einschätzung sogar zwingend benötigt werden. So muss bei der Bewertung der Zielkongruenz (zweiter Bewertungsschritt des primären Bewertungsprozesses, s.u.) das Ergebnis der Zielrelevanzbe-

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

wertung (erster Bewertungsschritt des primären Bewertungsprozesses) bereits vorliegen, da bei einer negativen Zielrelevanzbewertung keine Emotion generiert wird und in dem Fall auch die Zielkongruenzbewertung entfällt.

Gleichermaßen muss bei der Bestimmung der Art der Ich-Beteiligung (dritter Bewertungsschritt des primären Bewertungsprozesses) das Ergebnis der Zielkongruenz-Bewertung bekannt sein, da von diesem Ergebnis die Art der zu generierenden Emotion abhängt. Auf die gleiche Weise kann die kausale Reihenfolge der Bewertungsschritte aus der Beschreibung der Emotionsgenerierung für jede Einschätzungsdimension der primären Bewertungsprozesse abgeleitet werden.

Die kausale Ordnung der Bewertungsschritte des Emotionsentstehungsprozesses wurde als Grundlage für die Petrinetz-Repräsentation des primären Bewertungsprozesses in Abb. 6.22 genommen. Der Emotionsgenerierungsprozess konnte so als sequenzielle Abfolge der einzelnen Bewertungsschritte dargestellt werden.

Als Ausgangsbasis wurde das Netz aus Abb. 5.6 genommen, in der die Emotionsdefinition von Lazarus visualisiert wurde. Demnach wird die Emotionsgenerierung durch eine Änderung der Überzeugungen über die Bedeutung der aktuellen Person-Umwelt-Beziehungen (Stelle **Änderung der Überzeugungen über die Person-Umwelt- Beziehung**) ausgelöst. Eine Emotion entsteht als Ergebnis einer kognitiven Bewertung (Transition **kognitive Bewertung**) dieser Änderungen in Bezug auf das persönliche Wohlergehen bzw. die eigenen Ziele (Stelle **Ziele**). In Abb. 6.22 wird der kognitive Bewertungsprozess (Transition **kognitive Bewertung** im Netz 5.6) in Bezug auf die einzelnen Bewertungskriterien des primären Bewertungsprozesses nach Lazarus verfeinert.

Primärer Bewertungsprozess Die Aufgabe des primären Bewertungsprozesses besteht darin festzustellen, ob und in welcher Weise die wahrgenommene Person-Umwelt-Beziehung für das eigene Wohlergehen bedeutsam ist. Dazu wird sie in Bezug auf ihre Zielrelevanz, ihre Zielkongruenz und ihre Art der Ich-Beteiligung überprüft.

Die Überprüfung der Zielrelevanz (Transition **Bewertung der Zielrelevanz**) entscheidet darüber, ob die Person mit einer Emotion reagiert oder nicht. Wird die Person-Umwelt-Beziehung als relevant für persönliche Ziele eingestuft, folgt darauf grundsätzlich eine Emotion (Stelle **Emotion**), deren Intensität mit der Stärke und der Wichtigkeit des Zieles zunimmt. Erweist sich die Person-Umwelt-Beziehung als irrelevant für persönliche Ziele, entsteht keine Emotion (Stelle **keine Emotion**).

Von der Bewertung der Zielkongruenz (Transition **Bewertung der Zielkongruenz**) hängt ab, ob die Emotion positiv oder negativ sein wird. Eine zielkongruente Bewertung führt zu positiven (Stelle **zielkongruent**) und eine zielinkongruente Bewertung resultiert in negativen Emotionen (Stelle **zielinkongruent**).

Die dritte Komponente des primären Bewertungsprozesses, gilt der Feststellung der Art der Ich-Beteiligung (Transition **Art der Ich-Beteiligung**). Dazu wird nach Lazarus geprüft, welche Aspekte der Ich-Identität von der Person-Umwelt-Beziehung betroffen sind. Lazarus unterscheidet sechs Aspekte der Ich-Identität (vgl. gleichnamige Transitionen):

- (1) Selbst- und soziale Achtung;
- (2) moralische Werte;

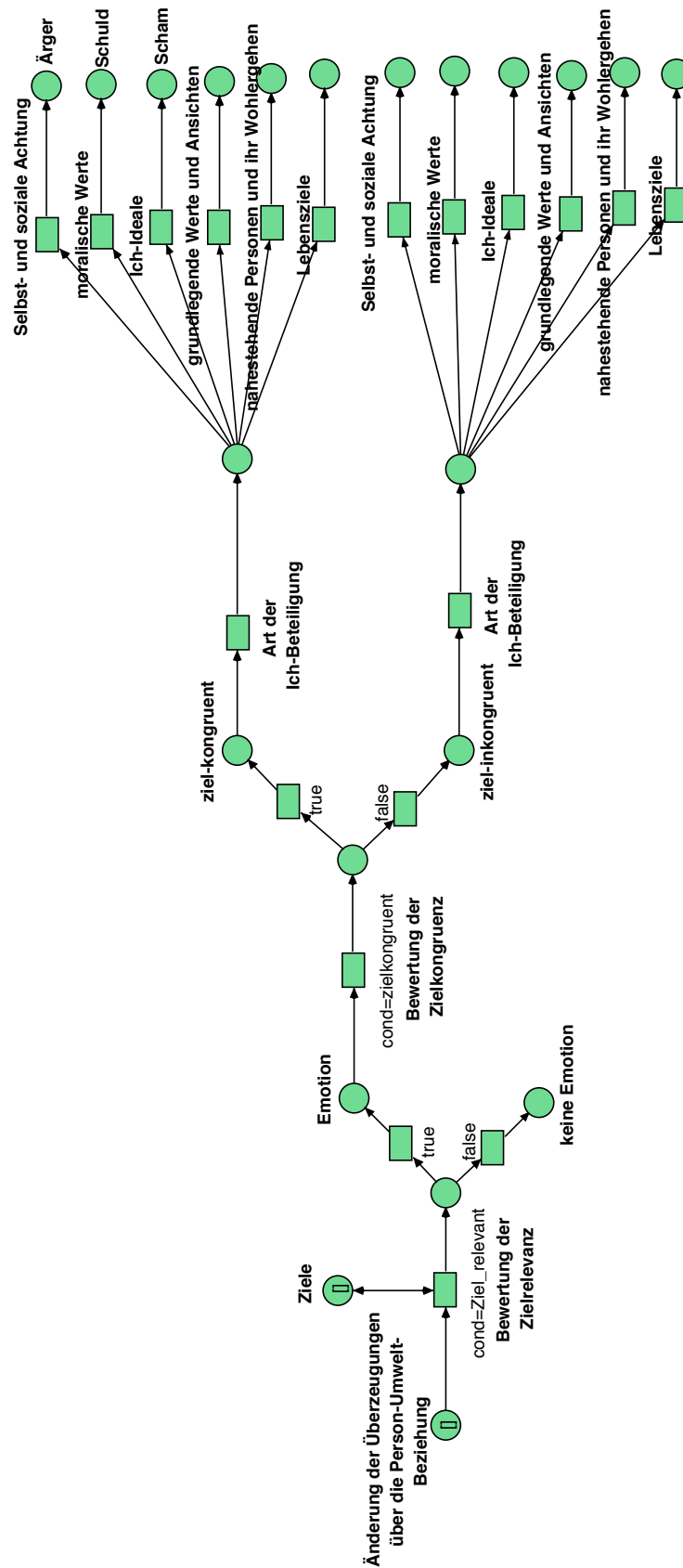


Abbildung 6.22.: Petrinetz-Darstellung des primären Bewertungsprozesses nach Lazarus.

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

- (3) Ich-Ideale;
- (4) grundlegende Werte und Ansichten;
- (5) nahestehende Personen und ihr Wohlergehen sowie
- (6) Lebensziele.

Für die Entstehung von Ärger (Stelle **Ärger**) hält es Lazarus beispielsweise für erforderlich, dass eine zielinkongruente Person-Umwelt Beziehung als eine Bedrohung der Selbst- oder der sozialen Achtung eingeschätzt wird; Schuld soll nach Lazarus die Bewertung voraussetzen, moralische Werte verletzt zu haben (Stelle **Schuld**); Scham dagegen die Bewertung, einem Ich-Ideal nicht gerecht geworden zu sein (Stelle **Scham**).

Sekundärer Bewertungsprozess Die drei Komponenten des sekundären Bewertungsprozesses umfassen die Bewertung, ob ein Verschulden oder ein Verdienst vorliegt sowie Bewertungen des Bewältigungspotentials und der zukunftsbezogenen Erwartungen. Der sekundäre Bewertungsprozess setzt auf den Ergebnissen der primären Bewertung auf (vgl. Abb. 6.23).

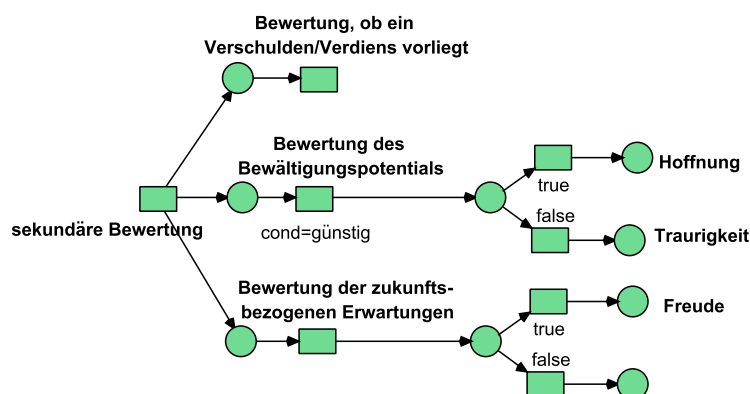


Abbildung 6.23.: Referenznetz-Darstellung des sekundären Bewertungsprozesses nach Lazarus.

Die Entscheidung über Verschulden oder Verdienst (Transition **Bewertung, ob ein Verschulden / Verdienst vorliegt**) beruht nach Lazarus auf kausalen Überzeugungen oder Attributionen über das Zustandekommen der ziel(in)kongruenten Person-Umwelt-Beziehung. Die kausalen Überzeugungen rechnet Lazarus dem kognitiven nicht emotionalen Wissen einer Person zu. Sie werden nur dann emotional gefärbt, wenn sich aus dem für das Zustandekommen einer ziel(in)kongruenten Person-Umwelt-Beziehung verantwortlichen Ursachenfaktor ein Verschulden oder ein Verdienst ableiten lässt, das auf eine persönliche Verantwortlichkeit schließen lässt. Diesen Schluss lassen nach Lazarus nur solche Ursachefaktoren zu, die durch die betreffende Person kontrollierbar sind. Da in der Theorie von Lazarus die Auswirkungen dieser Bewertungsdimension für den Emotionsentstehungsprozess bzw. für die Emotionsdifferenzierung nicht näher spezifiziert werden, können die Auswirkungen der Transition **Bewertung, ob ein Verschulden / Verdienst vorliegt**, innerhalb des Referenznetzes in Abb. 6.23 nicht modelliert werden.

Die Bewertung des Bewältigungspotenzials (Transition Bewertung des Bewältigungspotenzials) bezieht sich auf die Möglichkeiten, die der Person zur Verfügung stehen, um zielinkongruente Person-Umwelt-Beziehungen zu beseitigen oder zu verbessern bzw. zielkongruente Person-Umwelt-Beziehungen zu bewahren. Das Ergebnis dieser Bewertungskomponente ist lediglich für die Entstehung einer Emotion notwendig, nämlich für Traurigkeit (Stelle Traurigkeit), die bei einer ungünstigen Bewertung des Bewältigungspotenzials für einen persönlichen Verlust entsteht. Werden die Möglichkeiten, den Verlust zu bewältigen, dagegen als günstig beurteilt, entsteht keine Traurigkeit, sondern eher Hoffnung. Wichtiger als für die Differenzierung einzelner Emotionen ist das Ergebnis dieser Bewertungskomponente allerdings für die Intensivierung bzw. Abschwächung bestimmter negativer (nicht aber positiver) Emotionen. Nach Lazarus lässt sich durch eine günstige Bewertung des Bewältigungspotenzials Schuld durch Entschuldigung oder Wiedergutmachung lindern, Scham dagegen durch die Versprechung, die Anstrengungen zu verdoppeln, um einem Selbst-Ideal gerecht zu werden.

Ähnlich wie die Bewertung des Bewältigungspotenzials trägt auch die Bewertung der zukunftsbezogenen Erwartungen (Transition Bewertung der zukunftsbezogenen Erwartungen) eher zur Intensivierung bzw. Abschwächung einiger negativer (und auch einiger positiver) emotionaler Reaktionen als zu deren Differenzierung bei. Bei der Bewertung der zukunftsbezogenen Erwartungen überprüft die Person, inwieweit sich die Person-Umwelt-Beziehung zum Besseren oder zum Schlechteren verändern wird. Lediglich für Freude ist ein günstiges Ergebnis dieser Bewertungskomponente notwendig (Stelle Freude). Was die Abschwächung bzw. Intensivierung von Emotionen betrifft, nimmt Lazarus für eine positive Bewertung der zukunftsbezogenen Erwartungen ganz ähnliche Auswirkungen an wie für eine günstige Bewertung des Bewältigungspotenzials. Schuld und Scham sollen zum Beispiel dadurch abgeschwächt werden, während Traurigkeit mit Hoffnung statt mit Hoffnungslosigkeit und Depression verknüpft werden soll.

Die Emotionskomponenten und ihre Funktionen. Neben dem Ergebnis des Bewertungsprozesses machen drei weitere Komponenten eine Emotion aus, nämlich eine für jede Emotion charakteristische subjektive Erlebensqualität (ein Gefühl), eine emotionsspezifische Handlungstendenz sowie physiologische Veränderungen, die mit der Aktivierung der jeweiligen Handlungstendenz einhergehen. Die Emotionskomponenten sind in Abb. 6.24 in einem Petrinetz dargestellt. Dabei wurde die Emotionsdefinition von Lazarus (vgl. Abb. 5.6) mit seiner Spezifizierung der Emotionskomponenten kombiniert.

Wenn eine Handlungstendenz nicht unterdrückt wird, führt sie zur Handlung. Die physiologische Reaktion, die mit der Aktivierung der jeweiligen Handlungstendenz einhergeht, soll helfen, die Handlungen besser umsetzen zu können.

Die emotionsspezifischen Handlungstendenzen gelangen jedoch nicht notwendigerweise zur Ausführung, da wir normalerweise durchaus in der Lage sind, die jeweilige Handlungstendenz zu unterdrücken und aus einer Vielzahl unterschiedlicher verfügbarer Bewältigungsstrategien die auszuwählen, die unter den gegebenen Umständen die größten Erfolgsaussichten verspricht, zielkongruente Person-Umwelt-Beziehungen zu bewahren bzw. zielinkongruente Person-Umwelt-Beziehungen zu verbessern.

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

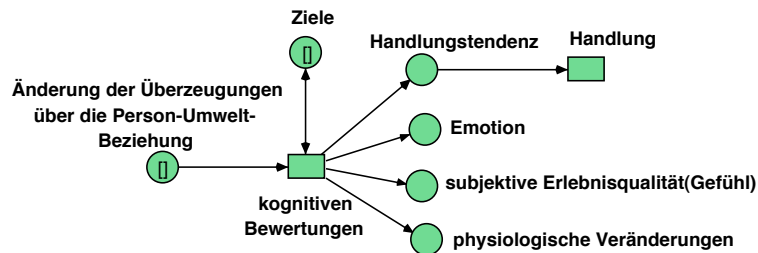


Abbildung 6.24.: Emotionskomponenten nach Lazarus.

Die Annahme von Lazarus, dass Handlungstendenzen notwendige Bestandteile von Emotionen seien, wurde in der Literatur kritisch diskutiert (z.B. [RWUS96]). Die Mehrheit der Autoren sieht Handlungstendenzen als Begleiterscheinungen oder Ergebnisse einer Emotion (vgl. exemplarisch [Fri86], [Fri07]). Nach dem Reaktionstrias-Konzept geht eine Emotion einer Handlung(stendenz) voraus (vgl. Kapitel 5.4).

Bewältigungsprozesse. Lazarus unterscheidet grundsätzlich zwei Formen von Anpassungs- bzw. Bewältigungsstrategien (*coping*), nämlich nach problemorientierter und emotionsorientierter Bewältigung.

Problemorientierte Bewältigungsbemühungen zielen auf konkrete Verbesserungen der Person-Umwelt-Beziehung ab, wobei versucht wird, durch direktes Handeln die emotionsauslösende Situation zu verändern.

Wenn die Situation durch Aktionen nicht verändert wurde oder werden konnte, greift der Bewältigungsprozess (*coping*) auf die emotionsorientierte Bewältigung zurück. Diese zielt auf eine Veränderung der subjektiven Sichtweise der Person-Umwelt-Beziehung ab, ohne diese Beziehung allerdings direkt zu beeinflussen. In der agentenorientierten Sichtweise hieße dies, dass bei der emotionsorientierten Bewältigungsstrategie die Bewertung der Umweltsituation bezüglich der *Beliefs* und *Desires* eines Agenten verändert bzw. umgewandelt werden müsste, um eine Anpassung an die bestehende Agent-Umwelt-Beziehung zu erreichen.

Nach Lazarus gibt es zwei Ansatzpunkte für die emotions-orientierten Bewältigungsprozesse. Zum einen ist es möglich, die Aufmerksamkeit von der Person-Umwelt-Beziehung ab- und anderen Dingen zuzuwenden. Zum anderen kann eine günstigere Umbewertung der Bedeutung der Person-Umwelt-Beziehung für das eigene Wohlergehen vorgenommen werden. Beide Formen der Bewältigungsstrategien führen zu einer Neubewertung (Reappraisal) der Person-Umwelt-Beziehung und in Abhängigkeit von den Ergebnissen dieser Bewertungen möglicherweise zu neuen Emotionen. Der Emotionsprozess ist nach Lazarus also ein sich ständig verändernder Prozess, in dem sich die beteiligten Komponenten wechselseitig beeinflussen.

Auswertung der Theorie von Lazarus aus der Sicht der informatischen Modellierung

Emotionen sind für Lazarus Reaktionen auf den Status der langfristigen Ziele und Werte. Nur wenn die Erreichung der Ziele oder Werte bedroht oder begünstigt wird, entsteht eine Emotion: “Wir reagieren nicht bei unwichtigen Dingen mit einer Emotion, sondern bei Werten und Zielen, zu denen wir eine starke Verpflichtung eingegangen sind.“ ([Laz91], S. 819f.).

Wie eine Situation bewertet wird, hängt einerseits von den Zielen, Wünschen oder Motiven einer Person, andererseits von ihrem aktuellen Wissenstand ab. Lazarus betont die Notwendigkeit eines allgemeinen *faktischen Wissens* für das Entstehen von Emotionen. Unter faktischem Wissen versteht Lazarus die Überzeugungen einer Person, “wie die Welt im allgemeinen und in einem besonderen Kontext funktioniert“ ([Laz91], S. 354; vgl. die *Beliefs* eines Agenten). Dieses Wissen ist nicht emotional und beeinflusst das Entstehen von Emotionen nur indirekt. Erst die Bewertung des Gewussten in Bezug auf seine Bedeutsamkeit für das persönliche Wohlergehen hat direkten kausalen Einfluss auf die Emotionsentstehung.

Die Emotionstheorie von Lazarus weist offensichtlich eine relativ hohe Kompatibilität mit den Konzepten der BDI-Agentenarchitektur auf. Das von Lazarus definierte kognitive, nicht emotionale Wissen korrespondiert mit den *Beliefs* eines BDI-Agenten, während die “Ziele und Werte“, denen sich ein Akteur verpflichtet, weitestgehend den *Desires* eines Agenten entsprechen.

Lazarus (1991) definiert die Handlungstendenzen im Unterschied zu den meisten anderen Emotionstheoretikern nicht als unter Einfluss von Emotionen gebildete Aktionspläne, sondern als Komponenten einer Emotion. Der Vergleich der entsprechenden Petrinetz-Darstellungen hat gezeigt, dass dies dem grundlegenden emotionspsychologischen Konzept der Reaktionstrias widerspricht (vgl. Abb. 5.4 und 6.24). In der psychologischen Literatur wurde die Annahme von Lazarus, Handlungstendenzen seien notwendige Bestandteile von Emotionen, mehrfach kritisch diskutiert und widerlegt (z.B. [RWUS96]).

Die Planungstätigkeit wird oberhalb der Aktionstendenzgenerierung platziert, nämlich bei der sekundären Einschätzungskomponente, die das Bewältigungspotenzial (*coping*) durch das Planen komplexer Anpassungs- und Bewältigungsprozesse bewertet. Die Modellierung des Bewertungsprozesses mit Petrinetzen hat gezeigt, dass die Theorie von Lazarus im Vergleich zu anderen Einschätzungstheorien weniger gut prozessorientiert darstellbar ist, da Lazarus nicht von den sequenziell ablaufenden Einschätzungsschritten ausgeht, sondern von sehr schnellen, unwillkürlichen und kausal nicht vorhersagbar verlaufenden Prozessen. Die Einschätzungsdimensionen sind bei Lazarus im Vergleich zu denen der anderen betrachteten Theorien weniger präzise definiert.

Emotionstheorie von Frijda

Fijda sieht Emotionen als adaptive Mechanismen, die durch ihre Funktionalität zur Anpassung des Organismus an die Umwelt beitragen. In [MF00] werden Emotionen definiert als “concern-activated response patterns which, depending on context evaluation, take control (precedence) in order to prepare the organism for a suitable change in action tendencies“ ([MF00], S. 60).

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Im Mittelpunkt von Frijdas Theorie steht der Begriff des Anliegens (*concern*). Ein Anliegen ist die Disposition eines Systems, bestimmte Zustände der Umwelt und des eigenen Organismus der Abwesenheit solcher Zustände vorzuziehen. Anliegen repräsentieren somit Ziele und Präferenzen des Systems: “A concern can be thought of as a representation of a desired state or event“ [MF00], S. 60). Somit ist der Begriff des Anliegens im Prinzip mit dem Konzept der Wünsche eines BDI-Agenten gleichbedeutend: Nach der Definition von [Rao95] repräsentieren “Wünsche“ (*Desires*) die von dem Agenten angestrebten zukünftigen Umweltzustände (vgl. Kapitel 3.2.5).

In [FM93] wird Anliegen mit den Zielen eines Agenten direkt in Verbindung gebracht: “[C]oncern, which is a deep need the system has, sometimes thought of as a permanent top-level goal“ ([FM93], S. 2). Ziele (*Goals*) werden als eine in sich konsistente Teilmenge von Intentionen definiert, durch die eine Intention in konkrete, direkt ausführbare Aktionen (Handlungsprimitiven) übersetzt wird (siehe Kapitel 3.2.5). Intentionen sind wiederum eine zur direkten Umsetzung ausgewählte Teilmenge von Wünschen. Es liegt also nahe, Frijdas Konzept des Anliegens mit dem Konzept der “Wünsche“ (*Desires*) eines BDI-Agenten zu vergleichen.

Ist ein Ereignis für die Anliegen des Systems relevant, signalisiert es das emotionale Subsystem an alle anderen Komponenten des Systems. Um die Anliegenrelevanz zu bestimmen, müssen externe Informationen aus der Umwelt und interne Signale aus dem eigenen System wahrgenommen und in Bezug auf die Systemanliegen ausgewertet werden. Die Auswertung, d.h. Bestimmung der Anliegenrelevanz (*concern relevance detection*) erfolgt im ersten Teil des Emotionsprozesses im Rahmen der Einschätzung. Die Stärke einer Emotion wird im Wesentlichen durch die Stärke des oder der assoziierten Anliegen bestimmt.

Der Emotionsprozess nach Frijda. Frijda spezifiziert in seiner Theorie nicht nur die Entstehung der Emotion infolge der kognitiven Einschätzungen, sondern den gesamten emotionalen Prozess, wobei insbesondere die funktionale Rolle der Emotion bei der Handlungsgenerierung beschrieben wird.

Der Emotionsprozess kann nach Frijda als eine sequenzielle Ausführung von mehreren Teilprozessen dargestellt werden (vgl. Abb. 6.25). Ausgehend von dieser Darstellung, bilden wir im Folgenden den Emotionsprozess nach Frijda mit Petri- bzw. Referenznetzen ab. Die Repräsentation des Emotionsprozesses als Referenznetz ist in Abb. 6.26 zu sehen.

Nach der Perzeption und der internen Repräsentation eines Stimulus (Transition *perception and encoding*) ist die Einschätzung der erste Schritt des emotionalen Prozesses. Die Einschätzung ist ein zweistufiger Prozess mit den Teilprozessen Relevanzeinschätzung (Transition *relevance appraisal*) und Kontexteinschätzung (Transition *context appraisal*).

Relevanzeinschätzung Bei der Relevanzeinschätzung - in ([MF00], S. 60) auch *primäre* Einschätzung genannt - wird die Bedeutung eines Reizes für das Anliegen des Systems eingeschätzt. Es handelt sich um eine schnelle automatische (d.h. unbewusste) Prüfung, ob zwischen dem Stimulusereignis und einem der Anliegen des Organismus (Stelle *concern*) Kongruenz besteht.

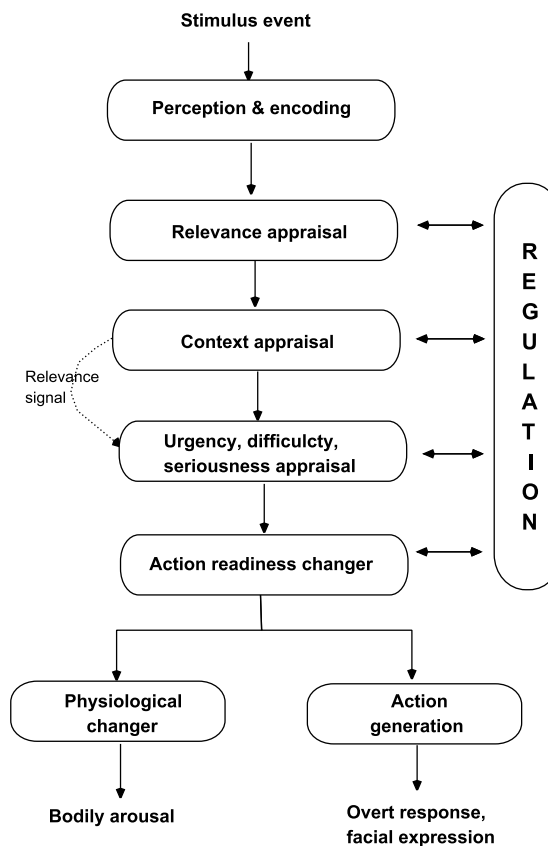


Abbildung 6.25.: Emotionsprozess nach Frijda (aus [MF00]).

Wenn eine Übereinstimmung festgestellt wird, wird das Ereignis als relevant für das jeweilige Anliegen eingeschätzt. Frijda erklärt, dass Stimulusereignisse kontinuierlich als relevant für die Anliegen des Organismus eingeschätzt werden, wobei oft mehrere Anliegen gleichzeitig angesprochen werden können. Dafür spricht die Überlegung, dass ein Organismus eine sehr große Menge von Anliegen haben kann (darunter sowohl die angeborenen als auch die aus der Erfahrung erlernten), und die Beobachtung, dass ein unvorgesehenes Ereignis sehr schnell eine emotionale Reaktion auslösen kann.

Falls bei der Relevanzeinschätzung die Übereinstimmung zwischen dem Stimulusereignis und dem Systemanliegen sehr genau ist, es sich um ein besonders wichtiges Anliegen handelt oder das Ereignis für mehrere Anliegen gleichzeitig relevant ist, wird ein Signal mit einer höheren Intensität ausgelöst (Transition generate concern-relevance signal).

Im nächsten Schritt wird der Stimulus als positiv oder negativ im Hinblick auf das Erreichen des Anliegens bewertet. Somit besitzt das als Ergebnis der primären Einschätzung generierte Signal - *“concern-relevance signal“* oder *“charger“* - die Parameter *Intensität* und *Valenz* und korrespondiert nach Frijda mit dem Affekt (vgl. 5.1). Es ermöglicht dem Organismus eine erste schnelle Einordnung der Situation und bereitet ihn auf die weiteren

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

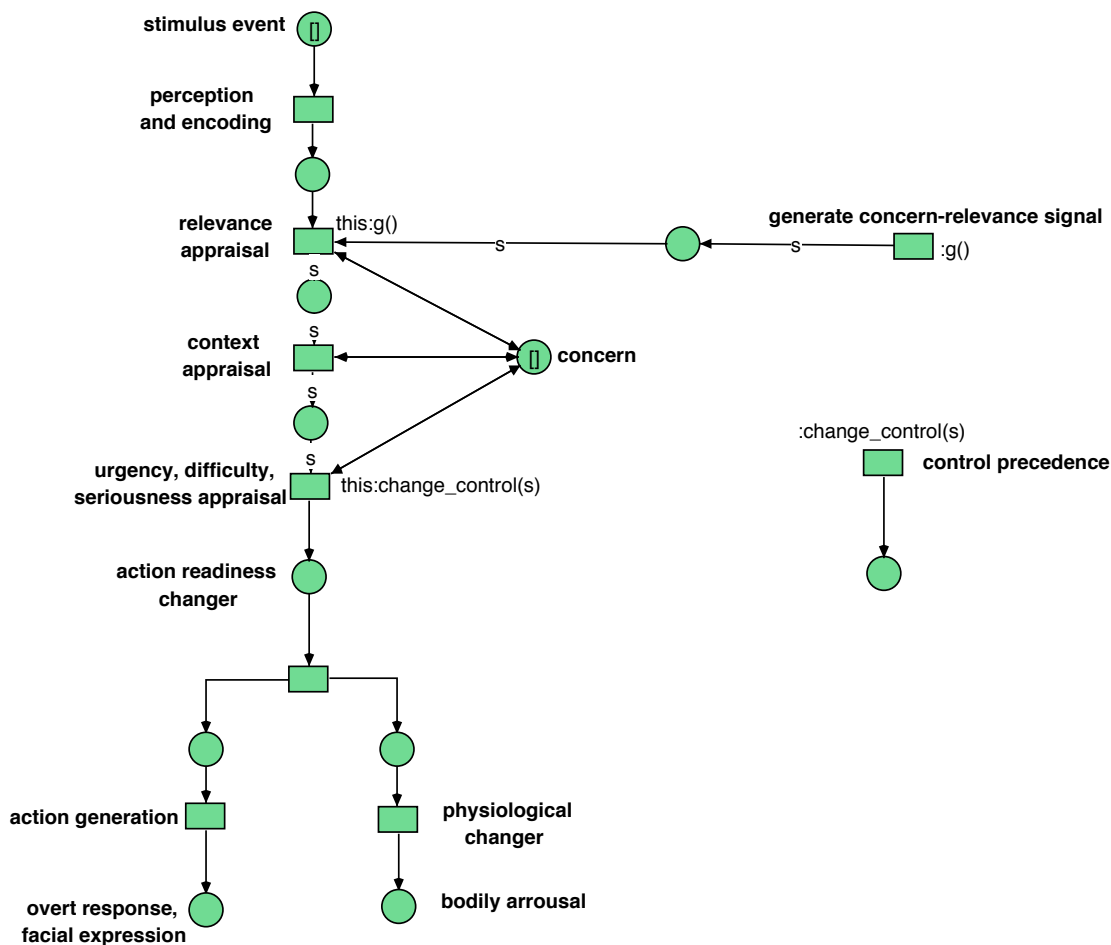


Abbildung 6.26.: Petrinetz-Darstellung des Emotionsprozesses nach Frijda.

Verarbeitungsprozesse vor.

Wenn das *concern-relevance signal* einen bestimmten Schwellenwert erreicht, kann es die laufenden kognitiven Verarbeitungsprozesse unterbrechen und die Aufmerksamkeit des Organismus auf den Auslösestimulus ziehen. Dieser Prozess wird Kontrollpriorität (Transition control precedence) genannt. Falls er durch eine entsprechend hohes Intensität des concern-relevance signal ausgelöst wird, verändert er die Kontrolle über das Verhalten und die Ressourcen sowie über die Prioritäten der Wahrnehmung und der Verarbeitung.

Kontexteinschätzung Bei der Kontexteinschätzung (auch *sekundäre* Einschätzung genannt) erfolgt die Einschätzung des Stimulus nach einer Reihe von Einschätzungsdimensionen. Diese Einschätzungsdimensionen beinhalten u.a. die Einschätzung von Valenz (*valence*), Erwartungskonformität (*expectedness*), Sicherheit (*certainty*), Kontrollierbarkeit (*contollability*), Bekanntheit (*familiarity*), Verursacher (*agency*) und Zeitbezug eines

Ereignisses. In [MF00] wird darauf hingewiesen, dass alle von Einschätzungstheoretikern definierten Einschätzungskriterien (insbesondere nach [Fri07], [Laz91] [Sch88]) als mögliche Einschätzungsdimensionen zur Kontexteinschätzung eingesetzt werden können.

Als besonders wichtige Einschätzungsdimension wird in [MF00] die Einschätzung des Bewältigungspotenzials (*coping potential*) hervorgehoben, bei dem die eigene Fähigkeit des Organismus eingeschätzt wird, mit dem Stimulusereignis / -situation umzugehen bzw. sie zu bewältigen. Die Autoren stellen die Hypothese auf, dass ein niedriger Wert bei der Einschätzung des Bewältigungspotenzials zu einer höheren Kontrollpriorität für die sich entwickelnde Emotion führt und dadurch noch mehr Aufmerksamkeit und kognitive Verarbeitungsressourcen beansprucht wird. [MF00] halten die Einschätzung des Bewältigungspotenzials für die wichtigste Einschätzungsdimension bei der sekundären Einschätzung, durch die die Emotion als solche definiert wird: Wird die Situation als einfach zu bewältigen eingeschätzt, entsteht nur eine affektive Reaktion, weil keine Notwendigkeit besteht, weitere Ressourcen für die Bewältigung freizusetzen. Wird dagegen die Bewältigung der Situation als schwer eingeschätzt, d.h. ist das eingeschätzte Wert des Bewältigungspotenzials negativ, so entwickelt sich entsprechend eine negative Emotion. Hat sich das Wert des Bewältigungspotenzials im Vergleich zu einer früheren Einschätzung verbessert, wird eine positive Emotion generiert.

Handlungsbereitschaft und Aktionstendenzen Im folgenden Schritt des Emotionsprozesses erfolgt die *Veränderung der Handlungsbereitschaft* (Transition *action readiness change*). Dies stellt, laut Frijda, das Herzstück der emotionalen Reaktion dar. Aus seiner Sicht können Emotionen sowohl als individuelle Erlebnisse verschiedener Einschätzungsmuster als auch als Zustände der besonderen Handlungsbereitschaft betrachtet werden. Veränderung der Handlungsbereitschaft bedeutet Veränderung in der Zuteilung von Verarbeitungs- und Aufmerksamkeitsressourcen sowie Generierung von Tendenzen, bestimmte Arten von Aktionen auszuführen - sog. Aktionstendenzen (*action tendencies*). Diese werden definiert als “states of readiness to achieve or maintain a given kind of relationship with the environment. They can be conceived of as plans or programs to achieve such ends, which are put in a state of readiness“ ([Fri86], S. 75). Aktionstendenzen sind eng mit emotionalen Zuständen verknüpft und dienen der Durchsetzung von Anliegen (*concern realization*). In [FKT89] wurden signifikante Abhängigkeitsbeziehungen zwischen bestimmten Einschätzungsmustern und Aktionstendenzen herausgearbeitet.

Tabelle 6.1 zeigt die von Frijda definierten Aktionstendenzen mit den zugehörigen Emotionen.

Der abschließende Schritt im emotionalen Prozess ist die Generierung der physiologischen Reaktionen (Transition *physiological changes*) oder faktischen Aktionen (Transition *Action generation*), u.a. in Form der expressiven Verhaltensweisen wie dem Gesichtsausdruck.

Das Emotionssystem überwacht alle Prozesse des Gesamtsystems und die Ereignisse der Umwelt, um rechtzeitig regulatorische Prozesse zu aktivieren, die steuernd eingreifen können. Frijda betont die Bedeutung der Emotionsregulation, da alle vorgestellten Schritte des Emotionsprozesses Gegenstände der Einwirkung durch regulatorische Prozesse sind. Regulatorische Prozesse können u.a. die Einschätzung beeinflussen (z.B. eine erneute Einschätzung einer Situation auslösen), die Impulse kontrollieren (z.B. durch Unterdrückung einer Akti-

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Aktionstendenz	Emotion
Approach	Desire
Avoidance	Fear
Being-with	Enjoyment, Confidence
Attending	Interest
Rejecting	Disgust
Nonattending	Indifference
Agonistic (Attack / Threat)	Anger
Interrupting	Shock, Surprise
Dominating	Arrogance
Submitting	Humility, Resignation

Tabelle 6.1.: Abhängigkeitsbeziehung zwischen Einschätzungsmustern und Aktionstendenzen nach Frijda ([Fri86], S. 88).

onstendenz) oder Handlungen modifizieren (z.B. durch Abmilderung oder Einstellung der expressiven Verhaltensweisen).

Rolle sozialer Normen Das Emotionssystem ist darauf eingestellt, in einer vorwiegend sozialen Umwelt zu operieren. Viele Einschätzungskategorien sind deshalb sozialer Natur. Die Definition des Einschätzungskriteriums “Legitimität“ basiert nach Frijda auf sozialen Normen. Dabei wird eingeschätzt, inwiefern eine fremde oder eigene Handlung mit den gesellschaftlichen Normen und Konventionen im Einklang steht. Viele Emotionen sind durch Missachtung der sozialen Normen bzw. die Konformität mit diesen bedingt. Beispiele für solche Emotionen sind etwa “comfort in one’s sense of propriety, pride in one’s outstanding achievements, admiration for those of others; shame and guilt upon one’s own infringements and distrust, anger, and indignation upon those of others“ ([Fri86], S. 311).

Auch für die Veranlassung des Emotionsregulationsprozesses sind soziale Normen wichtig: Sie bestimmen die Angemessenheit einer Emotion oder eines Emotionsausdrucks. Das Individuum wird zum normkonformen Emotionsausdruck motiviert, indem dieser von der Gesellschaft belohnt und dadurch verstärkt wird, wogegen Norm-abweichendes emotionales Verhalten als sozial unakzeptabel befunden und sanktioniert wird. Der regulatorische Prozess wird nach Frijda durch unerwünschte externe oder interne Ergebnisse der Handlungen, u.a. durch unkontrolliertes emotionales Verhalten, ausgelöst. Ein Beispiel einer unerwünschten externen Reaktion auf das unkontrollierte emotionale Verhalten ist die Bestrafung durch die (soziale) Umwelt, “when the environment retaliates, envies, disapproves, or despises because of emotions shown“ ([Fri86], S. 409), wobei die Missachtung sozialer Normen offensichtlich Hauptgrund für die Bestrafung ist. Soziale Normen liegen auch den “Gewissensbissen“ und dem Gefühl für soziale Angemessenheit (“the calls of conscience and the sense of propriety“ ([Fri86], S. 409)) zugrunde, indem sie unerwünschte interne Resultate bestimmter Verhaltensweisen signalisieren. Die entsprechenden internen Signale bestehen im Antizipieren der Emotionen der Scham oder Schuld, die nach Frijda durch Normmissachtung verursacht werden.

Moffat und Frijda (2000) illustrieren die Generierung und Regulation von Emotionen nach Frijdas Modell am Beispiel von Ärger. Ärger entsteht, wenn die primäre Einschätzung (Relevanzeinschätzung) eine Relevanz eines Ereignisses für die Anliegen des Akteurs ergibt und die sekundäre Einschätzung (Kontexteinschätzung) eine Behinderung für die Erfüllung dieses Anliegens feststellt. Als Aktionstendenz für Ärger ergibt sich nach Frijdas Abbildungsfunktion „Agonistic (Attack / Threat)“ (vgl. Abb. 6.1). Das Ziel dieser Aktionstendenz ist die Beseitigung der Hindernisse auf dem Weg zur Erfüllung des Anliegens. Wenn diese Aktionstendenz nicht reguliert, d.h. unterdrückt, wird, resultiert sie in einem aggressiven Verhalten gegenüber demjenigen Akteur, der das Erreichen des Anliegens verhindert hat bzw. als Bedrohung für das Anliegen eingestuft wurde.

Wenn jedoch Aggression keine sozial akzeptierte Verhaltensnorm darstellt, wird die Aktionstendenz zum aggressiven Verhalten durch Regulationsprozesse unterdrückt. Die Regulationsprozesse werden dann aktiv, wenn auf die vorangegangenen Episoden des gleichen emotionalen Verhaltens bzw. unkontrollierten Emotionsausdrucks negative Sanktionen seitens anderer Akteure folgen und deshalb auch die erneute Demonstration des gleichen Verhaltensmusters aversive Ergebnisse erwarten lässt. Wenn also aggressives Verhalten keine sozial akzeptierte emotionale Reaktion ist, wird die Aggressionsdemonstration z.B. durch Verachtung, Empörung, Ablehnung oder Zurückweisung sanktioniert. Diese (*externen aversiven Signale*) bilden die Grundlage für die Initiierung der Regulationsprozesse, die die gleichen Aktionstendenzen in Zukunft unterdrücken sollen, um erneute Bestrafungsmaßnahmen zu vermeiden. Dem gleichen Ziel dienen (*interne aversive Signale*), z.B. Schuld- und Schamgefühle, die ebenfalls sozial unzulässige Emotionen oder unangemessenen Emotionsausdruck signalisieren. Dadurch wird die emotionale Reaktion auf zweifache Weise auf ihre soziale Angemessenheit hin kontrolliert und ggf. reguliert.

Auswertung der Theorie von Frijda aus Sicht der informatischen Modellierung

Die Theorie von Frijda weist einige besondere Vorteile im Vergleich zu den anderen betrachteten Einschätzungstheorien auf.

Zum einen beschreibt sie nicht nur die Emotionseinschätzung, sondern den gesamten emotionalen Prozess. Speziell wird die funktionale Rolle der Emotion bei der Handlungsgenerierung ausführlich erläutert.

Zum anderen ist die Theorie gut operationalisierbar, da alle Prozessbeschreibungen genügend detailliert sind.

Schließlich lassen sich zentrale Konzepte wie Anliegen, Anliegenrelevanz usw. relativ gut mit den Konzepten der Agentenorientierung in der BDI-Architektur wie etwa Ziele und Motive eines Agenten zusammenbringen.

Damit bietet Frijdas Emotionstheorie neben einer funktionalen Emotionsdefinition eine sehr genaue Abdeckung aller Schritte des Emotionsprozesses im Sinne des in Abb. 5.2 spezifizierten Referenzmodells. Sie ist deshalb im Hinblick auf die Emotionsmodellierung in informatischen Systemen besonders gut geeignet.

Der besondere Vorteil von Frijdas Emotionstheorie liegt in ihrer Berücksichtigung der sozialen Komponente der Emotionen, die sowohl bei der Einschätzung als auch bei der Handlungstendenzgenerierung und den Regulationsprozessen zur Geltung kommt. Diese

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Eigenschaften machen Frijdas Emotionstheorie besonders für solche informatischen Emotionsmodelle interessant, die eine Berücksichtigung der sozialen Emotionsaspekte erfordern oder die Zusammenhänge zwischen Emotionen und sozialen Makrophänomenen erforschen wollen (vgl. z.B. die TABASCO-Architektur [SP98]; siehe auch Kapitel 7.2.1).

Da sich die generelle Beschreibung des emotionalen Systems nach Frijda besonders einfach formalisieren lässt, hat sie bereits als Grundlage für mehrere informatische Emotionsmodelle gedient, u.a. für die psychologischen Modelle ACRES und Will, die Frijda in Zusammenarbeit mit Swagerman (1987) bzw. Moffat (1995) entwickelt hat. Die Evaluation von ACRES führte Frijda anschließend dazu, eine Reihe von Modifikationen an seinem theoretischen Ansatz vorzunehmen.

6.1.3. Zusammenfassung und Gesamtbewertung der psychologischen Emotionstheorien im Hinblick auf informatische Modellierung

Die Betrachtung in diesem Kapitel hat gezeigt, dass die meisten psychologischen Emotionstheorien lediglich als Teiltheorien anzusehen sind. Im Sinne des in Kapitel 5 spezifizierten Referenzmodells betrachten sie lediglich einzelne Aspekte des emotionalen Prozesses und begrenzen auf diesen ihren Erklärungsanspruch.

In informatischen Emotionsmodellen werden am häufigsten die psychologischen Einschätzungstheorien implementiert.

Einschätzungstheorien erklären Emotionsgenerierung als einen Bewertungsprozess, bei dem bestimmte Umgebungseigenschaften aufgrund von spezifischen, zwischen den Theorien variierenden Einschätzungsdimensionen in Bezug auf Überzeugungen und Ziele eines Akteurs bewertet werden. Mit Einschätzungstheorien können neben der Emotionsgenerierung emotionale Phänomene wie die Relation zwischen Ereignis und Emotion oder die Entstehung verschiedener Emotionen in scheinbar gleichen situationalen Kontexten erklärt werden. Vorteile der Einschätzungstheorien im Hinblick auf die Modellierung der Emotion in Agentensystemen sind u.a. ihre leichte Operationalisierbarkeit, ihre funktionale Sicht auf Emotion sowie die gute Anbindung ihrer grundlegenden Konzepte an die Agentenmetapher, vor allem an die BDI-Architektur.

Alle betrachteten Emotionstheorien setzen die Bewertung der Relevanz des einzuschätzenden Ereignisses / der Handlung für die Ziele des Akteurs / der Agenten an die erste Stelle des Bewertungsprozesses und damit in direkte Beziehung zur Emotionsgenerierung (vgl. den Einfluss der Ziele eines Akteurs auf die Einschätzung der Erwünschtheit eines Ereignisses in [OCC88], die Einschätzung der Anliegerrelevanz in [Fri86], die Zielrelevanz in [Laz91] bzw. den *goal significance check* in [Sch93]). Ist das Ereignis für die Ziele des Agenten nicht relevant, wird keine Emotion ausgelöst ([Laz91], S. 819). Damit entsteht die Möglichkeit der Integration des Emotionsgenerierungsprozesses in die BDI-Architektur, etwa indem die aktuellen Ziele (*Desires*) eines Agenten seine Emotionsentstehung leiten.

Die Emotionstheorien von Lazarus und Frijda scheinen eine besonders gute Anbindung für BDI-Konzepte zu bieten: Das von Lazarus definierte “kognitives, nicht emotionale[s...] Wissen“ entspricht den *Beliefs*, die “Ziele und Werte“ den *Desires* eines Agenten (s.o.).

Die Theorie von Frijda erscheint für die Modellierung der Emotion zur technischen Systemoptimierung aufgrund mehrerer Gesichtspunkte besonders geeignet: Sie beschreibt nicht

nur die Einschätzung, sondern den gesamten emotionalen Prozess, ist gut operationalisierbar und ihre zentralen Konzepte wie Anliegen, Anliegenrelevanz usw., lassen sich relativ gut mit den Konzepten der Agentenorientierung, wie etwa den Zielen und Motiven eines Agenten in der BDI-Architektur, zusammenbringen.

Obwohl die Spezifizierung des Einschätzungsmechanismus in den meisten Einschätzungstheorien den größten Raum einnimmt, bildet die Einschätzung nur den ersten Schritt des gesamten Emotionsprozesses ([Fri86], [Fri07] - vgl. das Referenzmodell in Abb. 5.2). Für die informatische Modellierung ist jedoch insbesondere der zweite Schritt - nämlich die Funktion der Emotion, insbesondere ihre Auswirkung auf das Verhalten und die Kognition - interessant. Dieser Teil des Emotionsprozesses soll beschreiben, wie die durch die Einschätzung generierten Emotionen ihre regulatorischen Funktionen wahrnehmen und die Anpassungsreaktionen im Organismus auslösen. Von den betrachteten Einschätzungstheorien werden nur in der Theorie von Frijda [Fri86] bzw. [Fri07] die an die Einschätzung anschließenden Prozesse thematisiert. Die Theorien von Ortony et al., Lazarus und Scherer betrachten lediglich die Emotionseinschätzung (Emotionsgenerierung), was offensichtlich einen Nachteil dieser Theorien in Bezug auf die ganzheitliche Emotionsmodellierung darstellt.

So wird z.B. in der OCC-Theorie lediglich der Prozess der kognitiven Einschätzung der Ereignisse / Akteure / Objekte in Bezug auf bestimmten Dimensionen spezifiziert. Damit wird die Emotionsgenerierung von der Theorie hinreichend detailliert beschrieben, jedoch nicht die Funktion der generierten Emotionen für den weiteren Verlauf der Systemprozesse, wie etwa die Wahrnehmung, die Informationsverarbeitung oder die Handlungswahl. Will man die OCC-Theorie in einem Agentensystem modellieren, müssen zuerst die Relationen der Emotionen und ihre Auswirkungen im Gesamtsystem von den Modellentwicklern entworfen werden. Bei den Emotionsmodellen, die die OCC-Theorie als Grundlage benutzen ([Ell92], [Bat94], [Rei96]), wurde deshalb die Funktion der Emotionen im System meist intuitiv nachgebaut (bzw. ignoriert).

Die Theorie von Lazarus stimmt in vielen Beziehungen mit Frijdas theoretischen Annahmen überein. So setzt Lazarus ebenfalls die Existenz der mit Emotionen zusammenhängenden physiologischen Veränderungen und der in [Fri86] postulierten flexiblen Handlungstendenzen voraus.

Nach Lazarus bilden diese Handlungstendenzen zwar die Grundlage für die Einschätzung des Bewältigungspotenzials, er definiert sie jedoch in Unterschied zu Frijda nicht als unter Einfluss der Emotionen gebildete Aktionspläne. Vielmehr platziert er die Planungstätigkeiten oberhalb der Aktionstendenzgenerierung bei der sekundären Einschätzungskomponente, die das Bewältigungspotenzial (*coping*) durch das Planen komplexer Anpassungs- und Bewältigungsprozesse und das Antizipieren ihrer Ergebnisse bewertet.

Zusammenfassend lässt sich feststellen, dass die Theorie von Lazarus im Vergleich zu Frijdas Theorie weniger gut prozessorientiert darstellbar ist, da Lazarus nicht von sequenziell ablaufenden Einschätzungsschritten ausgeht, sondern von sehr schnellen, unwillkürlichen und kausal nicht vorhersagbar verlaufenden Prozessen. Die Einschätzungsdimensionen sind bei Lazarus im Vergleich zu denen der anderen betrachteten Theorien weniger präzise definiert.

Scherer definiert Emotion als eine Episode temporärer Synchronisation aller Systeme des

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

organischen Funktionierens [Sch93], die er durch fünf funktionale Subsysteme repräsentiert: Kognition, physiologisches Arousal, Motivation, motorischer Ausdruck, subjektives Erleben. Die Synchronisation erfolgt als Reaktion auf ein externes oder internes Stimulusereignis, das für zentrale Ziele des Organismus relevant ist. Viele der von Scherer angenommenen Einschätzungsdimensionen werden zwar auch in anderen Einschätzungstheorien genannt, jedoch werden sie in keiner der behandelten Theorien so differenziert und umfassend spezifiziert.

Von den betrachteten psychologischen Emotionstheorien berücksichtigen lediglich die Theorien von Frijda [Fri86] und Scherer [Sch93] explizit die Rolle der sozialen Phänomene auf verschiedenen Stufen des Emotionsprozesses. Insbesondere Frijda geht auf die Rolle der sozialen Normen bei der Generierung und Regulation von Emotionen ein. Nach [Sch93] wird durch den *compatibility check* die Kompatibilität eines Ereignisses bzw. einer Handlung mit den Gruppennormen bewertet. Diese Bewertungsdimension ist für die Generierung sozial wirksamer Emotionen wie Verachtung, Scham, Schuld und Ärger verantwortlich. Die grundlegende Frage an die psychologische Emotionsforschung, wie die wechselseitige Korrelation zwischen sozialen Phänomenen und Emotionen auf der psychologischen Ebene beschaffen ist, bleibt jedoch offen. Die betrachteten Ansätze illustrieren, dass die meisten emotionspsychologischen Forscher "gegenüber soziologischen Befunden weitgehend immun geblieben sind - sie schließen nur wenige und allenfalls eher dekorativ soziologische Beiträge ein" (Collins 1995, Barbalet 1999) (nach [Fla03], S. 118).

Die gemeinsamen Nachteile der Einschätzungstheorien liegen in ihrer schwierigen empirischen Überprüfung, den nicht gesicherten Einschätzungsdimensionen sowie der vielerorts mangelhaften Spezifikation der Mechanismen des Einschätzungsprozesses (vgl. [Rei00a]).

6.2. Neurobiologische Grundlagen von Emotion

Experimentelle Beweise der Beteiligung der Emotionen an kognitiven Denk- und Entscheidungsprozessen kommen in den letzten Jahren vor allem aus der neurowissenschaftlichen Forschung. Die Untersuchung der neurobiologischen Grundlagen der menschlichen Emotion hat zur Entwicklung von Damasio's *Somatic Marker Hypothese* über die wegweisende Rolle der Emotion bei der Lenkung von Entscheidungsprozessen geführt, die anschließend experimentell belegt werden konnte.

Moderne Neurowissenschaft präsentiert Emotion aus lokaler Sicht als Werkzeug der internen Selbstregulation des Organismus und aus globaler Perspektive als Werkzeug der externen sozialen Gesellschaftsregulation. Es gilt als neurowissenschaftlich erwiesen, dass Emotion unter günstigen wie ungünstigen Bedingungen ein integraler Bestandteil von Denk- und Entscheidungsprozessen ist [Dam03], [FADT11]. Weitere Untersuchungsergebnisse weisen auf einen engen Zusammenhang zwischen den sozialen Emotionen als Instrument der internen neurobiologischen Regulation auf der Mikroebene und der Regulation sozialer Phänomene auf der Makroebene hin ([Ado03], [R.10]). Das Zusammenwirken der kognitiven und der emotionalen Komponente bei der menschlichen Informationsverarbeitung ist in der aktuellen neurowissenschaftlichen Forschung unumstritten.

Aus Sicht der Informatik bedeuten diese Ergebnisse, dass der Zugang zum künstlichen Intellekt nicht allein durch Nachbilden der kognitiven Mechanismen unter Vernachlässigung der emotionalen möglich ist.

Bevor wir in diesem Kapitel auf die Hypothese der somatischen Marker (*Somatic Marker Hypothese*) eingehen, die die Rolle der Emotionen für die Entscheidungsfindung beschreibt, stellen wir kurz den Stand der neurowissenschaftlichen Emotionsforschung zur Emotion dar. Anschließend wird aus neurowissenschaftlicher Perspektive auf die Problematik der Gesellschaftsstrukturierung und Normetablierung durch Emotion eingegangen, die als spezieller Schwerpunkt dieser Theorienübersicht spezifiziert wurde (vgl. Kapitel 1.3).

6.2.1. Neurowissenschaftlicher Forschungsstand zur Emotion

Ebenso wie in der Psychologie waren Emotionen in den Neurowissenschaften bis vor wenigen Jahren kein expliziter Forschungsgegenstand. Ein Grund dafür lag darin, dass man Emotionen in den stammesgeschichtlich alten Zentren des Hirnstammes ortete, wogegen die "höchsten" kognitiven Hirnleistungen in Neocortex angesiedelt wurden und lange Zeit im Mittelpunkt des Forschungsinteresses standen.

In seinem Buch "The Emotional Brain" [LeD96a] kritisierte Joseph LeDoux die Vernachlässigung der Emotionalität in der Neurologie und die dadurch entstehende Eindimensionalität des Bewusstseinsbildes. Antonio Damasio behauptete ebenfalls, dass "Bewusstsein und Emotion nicht zu trennen" seien ([BDD00], S. 28). Die Arbeiten von LeDoux, Damasio, Adolphs und einigen weiteren amerikanischen Neurobiologen haben im Laufe der neunziger Jahre des 20. Jahrhunderts innerhalb der Neurowissenschaften schließlich eine Wende hin zur expliziten Betrachtung der Emotionen als wichtigste menschliche Verhaltensdeterminante herbeigeführt (vgl. [HJW07]).

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

In der Neurobiologie bestand die dominierende Herangehensweise zur Erforschung der Emotion darin, die für ihre Entstehung zuständigen Gehirnareale zu identifizieren. Die wichtige Rolle des Hypothalamus im emotionalen Geschehen wurde bereits 1929 durch die Experimente von Cannon und Bard bestätigt (siehe auch Kapitel 6.1.1). Später erweiterte James Papez die Hypothese vom Hypothalamus als Emotionszentrum [Pap37] und versuchte die Entstehung des subjektiven Erlebens der Emotionen durch einen Kreis anatomischer neuronaler Verbindungen zu erklären (*Papez-Kreis*). Nach heutiger Kenntnis spielt dieses neuronale Netzwerk allerdings für Emotionen eine untergeordnete Rolle. Zeitgenössische Forscher ([Pan98], [Pan06], [DP07], [LeD96a]) nehmen an, dass emotionale Erregung auf verschiedenen Ebenen entsteht und sich entlang verschiedener neuronaler Verbindungen ausbreitet.

Paul MacLean fasste das viscerale Gehirn und einige weitere Hirnstrukturen (Amygdala, Septum, Mamillarkörper, präfrontaler Cortex) unter den Begriff des *limbischen Systems* zusammen [Mac49]. Dieses betrachtete er als die stammesgeschichtlich älteste Hirnregion, die für die Sicherung des Überlebens und die damit verbundenen affektiven Verhaltensweisen (Nahrungserwerb, Fortpflanzungstrieb, Flucht und Kampf) zuständig ist. Emotionen beruhen MacLean zufolge auf der Verknüpfung von Sinnesempfindungen, die aus der Umwelt kommen, mit visceralen Empfindungen aus dem Körperinneren, und diese Integration erfolgt im limbischen System, dessen Zentrum er im Hippocampus sieht [Mac96].

Mehrere moderne Forscher propagierten die Distanzierung von dem Konzept des limbischen Systems, da viel Evidenz gesammelt werden konnte, dass das limbische System keine abgeschlossene neuroanatomische Einheit darstellt, die allein für Emotionen verantwortlich ist (vgl. exemplarisch [LeD00], S. 159). Vom Konzept des limbischen Systems ist man deshalb mangels unabhängiger Definitionskriterien sowie seiner Unschärfe und Missverständlichkeit inzwischen abgerückt.

Nach heutigem Erkenntnisstand basiert Emotion sowohl auf corticalen (in der Großhirnrinde zu findenden) als auch auf subcorticalen Strukturen. Die an Emotion beteiligten Prozesse sind hierarchisch strukturiert und umfassen verschiedene Hirnteile, vom sensorischen Apparat bis zum Neocortex. Auf jeder Ebene gibt es automatische und selbst regulierende Mechanismen, die eine vollständige Kontrolle durch ein übergeordnetes System verhindern. Die regulatorische Funktion der Emotionen wird dem orbitofrontalen Cortex zugeschrieben [Dam94].

Die Art der neuronalen Verbindungen zwischen den corticalen Strukturen erklärt die nur bedingte Kontrollierbarkeit der Emotionen durch das kognitive System: “Der bewußte Einfluß auf die Emotionen ist schwach, doch umgekehrt können Emotionen das ganze Bewußtsein überfluten. Der Grund: Die Verdrahtung des Gehirns wurde an dem entsprechenden Punkt unserer Evolutionsgeschichte so gestaltet, dass die Verbindungen von den emotionalen Systemen zu den kognitiven Systemen stärker sind als die Verbindungen in umgekehrter Richtung“ [LeD98].

Abweichend vom Großteil der neurowissenschaftlichen Forscher vertritt LeDoux die Meinung, dass verschiedenen Emotionsklassen unterschiedliche neuronale Strukturen zugrunde liegen: “[D]ie verschiedenen Klassen von Emotionen [werden] von eigenen neuralen Systemen vermittelt, die sich aus je eigenen Gründen entwickelt haben. Um uns einer Gefahr zu

6.2. Neurobiologische Grundlagen von Emotion

erwehren, benutzen wir ein anderes System als etwa bei der Fortpflanzung, und die mit der Aktivierung dieser Systeme entstehenden Gefühle - Angst bzw. sexuelle Lust - haben keinen gemeinsamen Ursprung. So etwas wie ein Emotionvermögen gibt es nicht, und es gibt kein Hirnsystem, das sich mit dieser Phantomfunktion befasst. Wenn wir die verschiedenen Phänomene verstehen wollen, für die wir den Ausdruck "Emotion" benutzen, müssen wir uns auf bestimmte Klassen von Emotionen beschränken" [LeD98].

Ein entscheidender Fortschritt beim Verständnis der neuronalen Grundlagen von Emotionen erbrachte das Studium von Läsionen im ventromedialen Frontallappen. Diese Läsionen führen zu Gefühlsarmut und Beeinträchtigungen des normalen sozialen Verhaltens, die meisten kognitiven Fähigkeiten sind dadurch aber typischerweise nicht betroffen: Intelligenz, Wortschatz, räumliches Vorstellungsvermögen, Gedächtnis, Lernen und Aufmerksamkeit bleiben meist intakt. Die Betroffenen verändern sich jedoch stark in ihrer Persönlichkeit; sie fallen durch störrisches Verhalten, Stehlen, Wutausbrüche, einen Mangel an Schuldbewusstsein oder Mitgefühl auf und missachten die grundlegenden ethischen und moralischen Normen der Gesellschaft (*Frontallappensyndrom*).

Es wurde ferner eine Abnormalität bei den emotionalen Prozessen festgestellt. Die Patienten sind nicht in der Lage, Emotionen zu produzieren, die von der Gesellschaft mit bestimmten Kategorien sozialer Situationen assoziiert werden, sodass sie zum Beispiel in "peinlichen" Situationen kein Gefühl der Verlegenheit zeigten ([Dam96], [BDTD05]). Emotionale Reize, die nicht angeboren sind, sondern erst zusammen mit dem Hintergrundwissen Gefühle auslösen, blieben ebenfalls wirkungslos. Auch änderte sich der elektrische Hautwiderstand (elektrodermale Aktivität) dieser Patienten nicht, was bei emotionaler Erregung normalerweise der Fall ist.

Schließlich schienen Patienten mit einer solchen Läsionsart unfähig zu sein, in komplexen Situationen rationale Entscheidungen zu treffen. Sie konnten die Vertrauenswürdigkeit einer Person nicht richtig einschätzen, zeigten keine Sympathie, Empathie, Scham oder Reue bei eigenen Vergehen, konnten keine Verpflichtungen einhalten usw. Gleichzeitig zeigten solche Patienten eine vollkommen normale Wahrnehmung der sozialen Situation, konnten Prämissen des Problems und mögliche Handlungsalternativen darstellen, kurz- und langfristige Folgen abschätzen und durch dieses Wissen logisch navigieren. Jedoch stellte für sie dieses Wissen in den realen Entscheidungssituationen keine Entscheidungshilfe dar.

Da also weder die Zugriffsprobleme auf soziales Wissen noch Insuffizienzen beim logischen Schlussfolgerungsmechanismus noch allgemeine Aufmerksamkeits- und Lernprobleme als Erklärungen für das Frontallappensyndrom in Frage kamen, stellte Damasio die These auf, dass mangelnde Emotionalität für die beobachteten Verarbeitungsstörungen verantwortlich sein könnte([Dam94], [Dam96], [BDTD05], [FADT11]). Er vermutete, dass die Patienten aufgrund der Läsion die im Laufe des Lebens akkumulierten emotionalen Erfahrungen bei ihrem Entscheidungsprozess nicht anwenden konnten ([Dam03], S. 145).²

²Die Amygdala spielt allgemein bei der Generierung der somatischen (emotionalen) Zustände eine Rolle, deshalb wirken sich Läsionen der Amygdala auf die Entscheidungsfindung ebenso negativ aus, wie Läsionen des ventromedialen präfrontalen Cortex.

Die Somatic Marker Hypothese (Damasio)

Nach Damasio's *Somatic Marker Hypothese* werden Emotionen in Form von physiologischen Körperzuständen (Blutdruck, Herzschlag, Hormonspiegel usw.) im Gehirn abgebildet und zusammen mit den Eigenschaften der emotionsauslösenden Situation gespeichert. In einer Entscheidungssituation können diese gespeicherten körperlichen Zustände (*somatische Marker*), die an vorherige positive oder negative Erfahrungen des Individuums gekoppelt sind, die potenziellen Handlungsergebnisse antizipieren, so dass langfristig die günstigere Handlungsstrategie vorgezogen wird. "The key idea of this hypothesis is that decision making is a process that is influenced by marker signals that arise in bioregulatory processes, including those that express themselves in emotions and feelings" ([BDD00], S. 295). Somatische Marker werden vom emotionalen Gedächtnis erzeugt und beeinflussen wesentlich die Entscheidungsfindung in komplexen Situationen, insbesondere im sozialen Kontext, sofern die Situation nicht völlig unvertraut ist. Einmal etabliert, kann die Überlagerung von kognitiven Inhalten und somatischen Markern auch ohne körperliche Vorgänge erfolgen. Im Gehirn werden dann die Vorstellungen kontrafaktischer Situationen generiert, in Form von "Als-ob-Schleifen", die die Körperrepräsentation im Gehirn beeinflussen, gleichsam in Form einer Simulation der körperlichen Rückmeldung. Dies erlaubt eine emotionale Bewertung auch fiktiver Situationen, was für Planungsprozesse von großer Hilfe ist.

Der ventromediale präfrontale Cortex stellt einen Lagerplatz für die Verknüpfungen zwischen Eigenschaften der vom Individuum erlebten Situationen und den mit diesen zusammenhängenden bioregulatorischen Zuständen dar. Die ventromedialen kortikalen Strukturen ermöglichen die Assoziation bestimmter Klassen komplexer Situationen mit verschiedenen Arten der bioregulatorischen Reaktion (einschließlich emotionaler Reaktionen), die, entsprechend der persönlichen Erfahrung in der Vergangenheit, in der Regel mit diesen Situationsklassen einhergehen. "The ventromedial sector holds linkages between the facts that compose a given situation, and the emotion previously paired with it in an individual's contingent experience" [BDD00]. Die Verknüpfung zwischen Situationseigenschaften und Emotionen nennt Damasio "dispositional" in dem Sinne, dass nicht die Repräsentationen der Situationsmerkmale oder der emotionellen Zustände selbst explizit festgehalten werden, sondern eher das Potenzial, eine Emotion durch Ansprechen entsprechender kortikaler und subkortikaler Strukturen zu reaktivieren.

In einer komplexen Situation machen wir bestimmte Beobachtungen über die Situation als Ganzes sowie insbesondere ihre charakteristischen Komponenten, z.B. über die beteiligten Akteure und ihre Konfiguration, Handlungsfolgen, Menge der Reaktionsmöglichkeiten, Menge der kurz- und langfristigen Ergebnisse für jede Reaktionsoption usw. Diese Beobachtungen werden verarbeitet und in Form einer Situationskategorisierung mit zugehörigen Handlungsdispositionen im Gedächtnissystem aufgezeichnet. "[B]ecause the experience of some of those components has been associated with emotional responses, which were triggered from cortical and subcortical sites that are dispositionally prepared to respond, it is proposed that the ventromedial prefrontal cortex establishes a linkage between the disposition for a certain aspect of a situation (for instance, the long-term outcome for a type of response option), and the disposition for the type of emotion that in past experience has been associated with the situation" ([BDD00], S. 297).

6.2. Neurobiologische Grundlagen von Emotion

Wenn eine Person in einer Situation ist, für die (aufgrund vorheriger Erfahrung) eine Zuordnung zu einer bestimmten Situationsklasse mit bereits kategorisierten Situationsfaktoren möglich ist, wird zusätzlich zur Reaktivierung dieser Faktoren der mit ihnen verknüpfte Apparat der emotionalen Dispositionen aus dem ventromedialen präfrontalen Cortex mitreaktiviert. Als Ergebnis wird der situationsgebundene emotionale Zustand rekonstruiert.

Analog gilt, dass wenn die Person in einer Situation eine Entscheidung treffen muss, ihr nicht nur die faktischen Erfahrungen aus der Vergangenheit (wie die Gesamtmenge möglicher Entscheidungen, frühere Entscheidungen in gleichem Kontext samt ihrer kurz- und langfristigen Ergebnisse usw.) zur Verfügung stehen, sondern auch die Bewertung dieser faktischen Erfahrungen durch die mit ihnen verknüpften emotionalen Zustände. Diese reaktivierten emotionalen Zustände (*somatic marker*) sind die Randbedingungen, die den Entscheidungsprozess auf mehreren Ebenen und in mehrere Richtungen lenken. Sie markieren Handlungsoptionen und Ausgangsalternativen mit positiven oder negativen Signalen, verkleinern so den Entscheidungsraum und erhöhen damit einerseits die Wahrscheinlichkeit einer optimalen Entscheidung, andererseits die Effizienz des Entscheidungsprozesses. “For instance, when the somatosensory image, which defines a certain emotional response is juxtaposed to the images which describe a related scenario of future outcome, and which triggered the emotional response via the ventromedial linkage, the somatosensory pattern marks the scenario as good or bad” ([BDD00], 297). Somit können die Zukunftsalternativen anhand von früheren emotionalen Erfahrungen antizipiert werden und zwar schneller und zuverlässiger, als es mit einem kognitiven Entscheidungsprozess je möglich wäre.

In einer Reihe von Experimenten unter Beteiligung von Patienten mit einer Läsion des ventromedialen präfrontalen Cortex konnten die einzelnen Postulate von Damasio's Hypothese erfolgreich bestätigt werden (z.B. [BDTD97], [BDTD05]). Eine spezielle Testaufgabe wurde entwickelt, um die Schwere der Schädigung des Entscheidungsfindungsmechanismus der Patienten jeweils zu bestimmen und zu messen: “the gambling task” ([BDDA94], [BDTD05]). In Experimenten mit Spielkarten (z.B. als Wahl zwischen zwei Stapeln mit unterschiedlichen Gewinnen und Verlusten) lernten gesunde Versuchspersonen schnell, auf Intuition und die Gefühle “im Bauch” zu hören und verhielten sich alsbald richtig - und zwar häufig sogar, bevor ihnen bewusst wurde, welche Entscheidung die bessere ist. Den Versuchspersonen mit Schädigung des frontalen Cortex gelang dies nicht und oft handelten sie sogar wider besseres Wissen. Die Positronenemissionstomografie-Aufnahmen zeigten, dass der frontale Cortex umso aktiver war, je unkontrollierbarer die Aufgabe, d.h. je mehr eine Entscheidung von intuitiven Einschätzungen abhing.

Da die Emotion in natürlichen intelligenten Systemen nachweislich an Entscheidungsprozessen beteiligt ist und deren Effizienz steigert, kann man von den gleichen Auswirkungen der sachgerecht implementierten emotionalen Mechanismen auf die künstlichen Entscheidungs- und Handlungswahlprozesse ausgehen. Die Somatic Marker Hypothese ist zu einem wichtigen Argument der KI-Forscher geworden, die in der Nachbildung von emotionalen Mechanismen in KI-Systemen eine Möglichkeit der Optimierung rein kognitiver Entscheidungs- und Handlungswahlalgorithmen sehen.

Emotionales und kognitives Prozessieren (LeDoux)

Joseph LeDoux und Michael Gazzaniga konnten in Experimenten an Split-Brain-Patienten zeigen, dass die emotionale und kognitive Verarbeitung eines Stimulus durch unterschiedliche neuronale Systeme erfolgt, die jedoch parallel zueinander arbeiten [LeD96a]. In einem Versuch konnten durch Stimuli bei einem Split-Brain-Patienten bewusste Emotionen hervorgerufen werden, die dieser nicht bewusst wahrnahm. LeDoux zeigte damit, dass die linke Gehirn-Hemisphäre in der Lage ist, auf Reize, die sie nicht erkennt bzw. die ihr nicht präsentiert werden, emotional zu reagieren. Er erklärt diese Dissoziation durch eine Art Gabelungseffekt. Der Weg, den der Reiz nimmt, nachdem der Stimulus der rechten Hemisphäre präsentiert wird, gabelt sich auf: Die rechte Hemisphäre identifiziert den Stimulus, aber diese Information kann aufgrund der durchtrennten Nervenbahnen der linken nicht übermittelt werden. Diese erste begriffliche Identifikation des Stimulus ist jedoch nicht alles: Der Stimulus wird weiterverarbeitet in den Bereichen der rechten Hemisphäre, die seinen emotionalen Gehalt auswerten. Offensichtlich erreicht diese Information trotz der durchtrennten neuronalen Verbindungen die linke Gehirnhälfte, wo sie als bewusste Emotion wahrgenommen und verbalisiert werden kann. Die zentralen neuralen Bahnen, auf denen die affektive Signifikanz äußerer Stimuli weitergegeben bzw. verarbeitet wird, sind nicht dieselben, auf denen die objektiven Merkmale des Stimulus verarbeitet werden. Das bedeutet, dass voneinander verschiedene Verarbeitungsfunktionen, die durch unterschiedliche neuronale Systeme erfolgen, die affektive und die kognitive Reizverarbeitung leisten ([LeD96a], S. 15ff.). Somit liefern die durch die Split-Brain-Experimente gewonnenen Erkenntnisse das hirnpfysiologische Korrelat zur grundlegenden Dichotomie zwischen Emotion und Kognition und geben erste Hinweise darauf, dass emotionale und kognitive Verarbeitungsprozesse nach der Wahrnehmung eines Stimulus parallel ablaufen und sich gegenseitig ergänzen.

In weiteren Arbeiten hat sich LeDoux speziell mit der Entschlüsselung der neuronalen Grundlagen von Angst und Furcht beschäftigt. Nach seinem Modell der Angstkonditionierung lösen emotionsevozierende Stimuli zwei Prozesse im Zentralnervensystem aus, die beide ihren Ursprung im Thalamus haben und danach jeweils entlang auf einem eigenständigen Pfad verlaufen. Das *emotionale Prozessieren* ist der schnelle Abgleich des Reizes mit groben Reizmustern zur Kategorisierung als gefährlich / ungefährlich. Dieser Verarbeitungspfad dient zur Vorbereitung schneller Reaktionen (z.B. Flucht) und ist fehleranfällig. Für das emotionale Prozessieren ist hauptsächlich die Amygdala zuständig (vgl. auch [FADT11]). Die Kategorisierung findet nicht auf Grundlage angeborener, sondern aufgrund erlernter Reizmuster statt. Die Amygdala habituiert erfahrungsbedingt, also kann man von der Amygdala als einem emotionalen Gedächtnis sprechen.

Den zweiten Prozess nennt LeDoux *kognitives Prozessieren*. Er dient zur Kontrolle der beim emotionalen Prozessieren gewonnenen Information und ist zeitaufwendiger. Dieser Prozess beginnt am Thalamus und verläuft über den präfrontalen Cortex sowie auf einer Nebenroute über den Hippocampus. Der präfrontale Cortex empfängt die verarbeiteten sensorischen Signale, integriert sie mit Gedächtnisinhalten und emotionalen Bewertungen und initiiert auf dieser Basis Handlungen. Er wird als oberstes Kontrollzentrum für eine situationsangemessene Handlungssteuerung angesehen und ist gleichzeitig intensiv an der Regulation emotionaler Prozesse beteiligt. Ggf. werden die auf dem Wege des emotionalen

Prozessierens in der Amygdala erzeugten emotionalen Reaktionen durch die kognitive Verarbeitung im präfrontalen Cortex korrigiert bzw. gehemmt.

Das Zusammenwirken kognitiver und emotionaler Verarbeitungsprozesse wird in Abb. 6.27 durch ein Petrinetz veranschaulicht. Beide Prozesse werden jeweils durch einen Stimulus (eine natürliche Bedrohung im Falle von Angstgenerierung) initiiert, der an die entsprechenden Gehirnareale (Stelle Amygdala bzw. präfrontaler Cortex, Hippocampus) übertragen wird (Transition Reizübertragung). Die emotionalen und kognitiven Verarbeitungsprozesse (Transitionen kognitives Prozessieren bzw. emotionales Prozessieren) laufen parallel und mobilisieren automatisch neuronale und hormonelle Reaktionen (Transitionen automatische neuronale Erregung bzw. hormonelle Reaktionen), die eine schnelle Verhaltensantwort (z.B. Flucht- oder Angriffsverhalten) vorbereiten sollen.

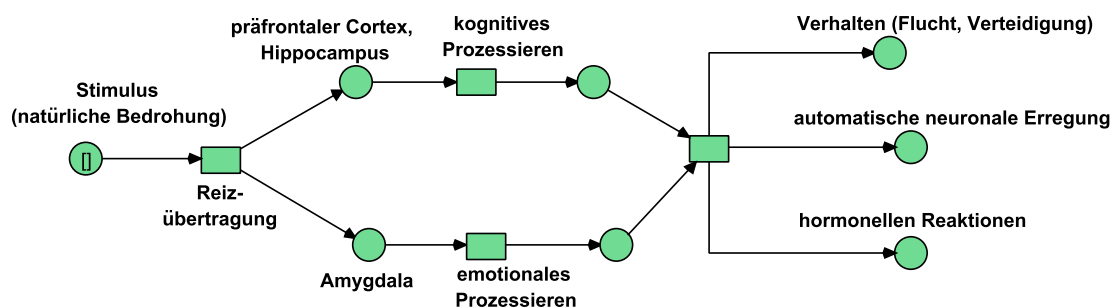


Abbildung 6.27.: Emotionalen und kognitiven Verarbeitungsprozesse bei Angst nach [LeD96a].

Social cognition und Social neuroscience

Adolphs untersuchte, welche neuronalen Strukturen für die soziale Urteilsbildung bzw. die Perzeption und Interpretation des sozialen Einflusses von anderen Personen (in der Sozialpsychologie mit dem Term *person perception* bezeichnet) verantwortlich sind. Er konnte zeigen, dass dieselben Gehirnstrukturen, die sich als wichtig für emotionale Prozesse herausgestellt haben, ebenfalls bei der Generierung des sozialen Verhaltens beteiligt sind (vgl. [Ado03]). Damit ist ein neues Forschungsfeld in der Neurowissenschaft entstanden, das sich schwerpunktmäßig mit der Erforschung von kognitiven neuronalen Grundlagen des sozialen Verhaltens beschäftigt, *social neuroscience* (z.B. [HJW07]), auch *social cognition* genannt ([Ado03], S. 119).

6.2.2. Emotion und soziale Norm: Soziale Emotionen bei der Regulation sozialen Verhaltens

In der Neurowissenschaft werden Emotionen als Mechanismen der Bioregulation (Homeostase) gesehen, welche die Anpassung des Organismus an die komplexe Umgebung steuern: "Emotions can be thought of as states that coordinate homeostasis in a complex, dynamic environment" ([Ado03], S. 165). Damasio unterteilt Emotionen in drei Klassen: Hintergrundemotionen, primäre und soziale Emotionen. Emotionen aller drei Klassen dienen

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

der optimalen Anpassung des Organismus an die Umweltbedingungen ("survival with well-being"). Damasio behauptet, dass soziale Emotionen, zusammen mit den Trieben und den primären Emotionen, auf die sie aufbauen, einen ausschlaggebenden Einfluss auf das soziale Verhalten haben. Auch Adolphs sieht die Funktion der sozialen (bzw. "moralischen") Emotionen in der Steuerung des sozialen Verhaltens: "[I]n so far as one aspect of the environment is social, emotions will participate in regulating social behaviour. In fact, one class of emotions – the so-called social or MORAL EMOTIONS – serve specifically in this capacity and probably guide altruistic helping and punishment" ([Ado03], S. 165, Herv.i.O.). Diese Annahme wird durch die Untersuchungsergebnisse von Damasio et al. unterstützt, die belegen, dass die Defizite durch Amnesien des ventromedialen frontalen Lappens, die die sozialen Emotionen hemmen, sich auf soziales Verhalten negativ auswirken [BDD00].

Damasio sieht die gesellschaftlichen Normen als einen speziellen Mechanismus der Bioregulation (Homeostase) - neben metabolischer Regulation, Trieben und Motivationen sowie allen drei Arten von Emotionen (vgl. Kapitel 5.5) -, der jedoch auf der globalen Ebene einer Akteurgemeinschaft wirkt. Bei Menschen ist die Komplexität der physischen und sozialen Umgebung dermaßen hoch, dass diese nicht allein durch automatische Anpassungsprozesse wie Emotionen bewältigt werden kann. Die sozialen Konventionen und ethischen Verhaltensregeln sowie die Institutionen, die sie etablieren - Religion, Gerechtigkeit und sozialpolitische Organisationen -, werden auf der Ebene der sozialen Gruppe zu zusätzlichen nicht automatischen homeostatischen (Kontroll-)Mechanismen.

In [Dam03] stellt Damasio eine eigene Theorie über die Prinzipien des Funktionierens der sozialen Regulation in einer menschlichen Gemeinschaft vor. Er behauptet, dass die menschliche Tendenz zum sozialen Einvernehmen mit den anderen eine ontogenetische Erweiterung des Selbsterhaltungstriebes sei und durch den regulatorischen Apparat der Emotionalität zumindest zum Teil biologisch verankert ist. "We happen to be biologically structured in a certain way - mandated to survive and to maximize pleasurable rather than painful survival - and from that necessity comes a certain social agreement." ([Dam03], S. 173).

Auf Basis der Hypothese der somatischen Marker begründet Damasio [Dam96] die Rolle der sozialen Emotionen für die Kategorisierung der sozialen Situationen, die im Laufe der individuellen Sozialisations- und Entwicklungsgeschichte durchlebt wurden. Damasio unterstreicht, dass die wichtigen sozialen und persönlichen Erfahrungen in besonderem Maße von Emotionen begleitet werden. Akteure begegnen bestimmten Komponenten einer sozialen Situation mit einer Reihe von (sozialen) Emotionen, die zum Teil angeborene automatische Reaktionen sind oder durch vorangegangene Erfahrungen (Belohnung bzw. Bestrafung) induziert wurden. Mit der Zeit lernt das Individuum, die erlebten Situationen unter dem Einfluss der sozialen Emotionen (wie Sympathie, Scham, Stolz, Empörung usw.) nach der internen Situationsstruktur, den Situationskomponenten und der Signifikanz für die gerade vorherrschenden eigenen Ziele und Überzeugungen zu kategorisieren. Die resultierenden konzeptuellen Kategorien werden auf physischer und der korrespondierenden neuronalen Ebene mit den Mechanismen verbunden, die für die Emotionsgenerierung verantwortlich sind. Die Situationsstruktur, die Handlungsoptionen, die Zukunftsalternativen werden zunehmend mit korrespondierenden Emotionen assoziiert, so dass bei einem erneuten Eintritt einer Situation aus der gleichen Kategorie die entsprechenden emotionalen Zustände automatisch

6.2. Neurobiologische Grundlagen von Emotion

aufgerufen werden. “This arrangement allows us to connect categories of social knowledge - whether acquired or refined through individual experience - with the innate, gene-given apparatus of social emotions and their subsequent feelings“ ([Dam03], S. 147).

Ferner wird in [Dam03] gezeigt, wie soziale Normen und Verhaltensregeln durch geeignete Positionierung der akteursinternen Emotionen implementiert werden können ([Dam03], 156 ff.). Der Autor klassifiziert soziale Emotionen anhand des sie auslösenden ECS (Emotion Competent Stimulus) und der Funktion der Emotion in einem sozialen System. Demnach werden Emotionen der *Verlegenheit*, *Scham* und *Schuld* durch Fehler in der eigenen Person bzw. Verhalten ausgelöst, das nach eigenem Ermessen des Akteurs bestimmten (internen oder externen) Standards nicht genügt. Diese Emotionen haben zum Ziel, die Bestrafung durch die andere Akteure vorwegzunehmen, die Balance im Selbst, den anderen oder der Gruppe gegenüber aufrechtzuerhalten bzw. wiederherzustellen und auf diese Weise die sozialen Konventionen und Regeln zu festigen. Analog werden Emotionen wie *Verachtung* und *Empörung* ausgelöst, wenn ein anderer Akteur die sozialen Normen der Gruppe bricht. Diese Emotionen haben zum Ziel, die Missachtung der Gruppenkonventionen und Normen zu bestrafen sowie ihre Bedeutung zu betonen und zu intensivieren. Somit entstehen Emotionen als negative interne Konsequenzen der Normmissachtung und stützen damit soziale Normen zusätzlich zu ggf. externen Bestrafungssanktionen. Im Petrinetz-Modell in Abb. 6.28 wird Damasio's Ansatz zur emotionsgeleiteten Internalisierung von sozialen Normen repräsentiert.

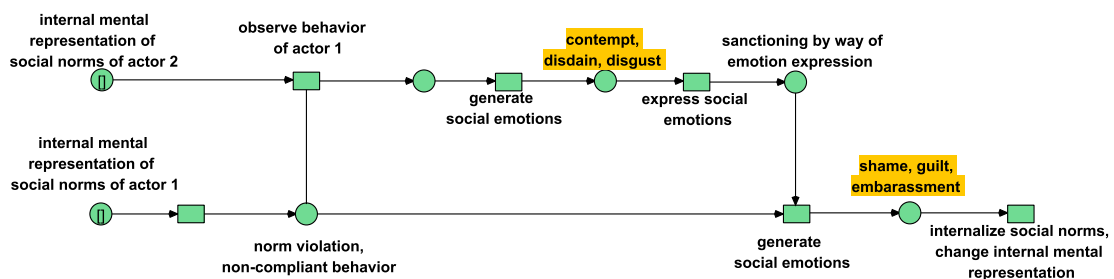


Abbildung 6.28.: Sanktionierung des normabweichenden Verhaltens und Internalisierung der sozialen Normen nach [Dam03].

Das Netz zeigt eine Interaktionssequenz zwischen einem sich nicht konform verhaltenden Akteur und einem anderen Akteur, der die Normverletzung wahrnimmt. Der untere Netzpfad repräsentiert die Aktionsabfolge des Normverletzers (Akteur 1), der obere – die des Sanktionierers (Akteur 2). Hier hätte man anstelle eines Petrinetzes auch ein Agenteninteraktionsprotokoll (AIP, vgl. Kapitel 4.3) als Modellierungstechnik benutzen können, da mit dem Netz eine Interaktion von zwei Agenten wiedergegeben werden soll. Zugunsten der einheitlichen Modellierung der theoretischen Grundlagen wird jedoch eine Petrinetz-Darstellung auch hier vorgezogen

Initial haben beide Akteure ihre eigenen internen mentalen Repräsentationen der sozialen Normen und Verhaltensstandards (Stelle *internal mental representation of social norms*).

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Diese sind nicht unbedingt gleich und entsprechen anfänglich nicht unbedingt den in der Gesellschaft akzeptierten, geteilten Vorstellungen.

Akteur 1 verhält sich nicht normkonform und bricht eine soziale Norm (Stelle **norm violation / non-compliant behavior**). Akteur 2 beobachtet das Verhalten von Akteur 1 (Transition **observe behavior of actor 1**) und nimmt die Normverletzung wahr. Daraufhin werden soziale Emotionen generiert (Transition **generate social emotions**), woraufhin Akteur 2 Verachtung und Empörung empfindet (Stelle **contempt, disdain, disgust**). Indem er diesen Emotionen direkt durch Worte und Handlungen oder indirekt durch Körpersprache, Mimik etc. Ausdruck gibt (Transition **express social emotions**), löst er negative soziale Emotionen bei dem Normbrecher aus (Transition **generate social emotions**) und induziert ihm den emotionalen Zustand der Scham, Schuld oder Verlegenheit (Stelle **shame, guilt, embarrassment**) (dieser Schritt ist bei Damasio nicht explizit beschrieben, wird aber von mehreren soziologischen Autoren (z.B. [Els96], [Els99]) als wichtige Komponente im Prozess der Norminternalisierung angesehen; vgl. Kapitel 6.3.3). Auf diese Weise wird durch die sozialen Emotionen der beteiligten Akteure eine Normverletzung sanktioniert. Die negativen Emotionen (Verachtung und Empörung) der Beobachter werden dem Normbrecher direkt oder indirekt mitgeteilt und lösen bei ihm Verlegenheit, Scham und Schuld aus.

Das Streben nach emotionaler Belohnung wird für eine der Hauptmotivation des menschlichen Verhaltens gehalten ([Col84]). So werden Akteure stets bemüht sein, Interaktionen zu erzwingen, die mit positiven Emotionen einhergehen, und Interaktionen zu vermeiden, die negative Emotionen mit sich bringen. Durch die von Damasio nachgewiesene Verknüpfung von bestimmten Situationskomponenten und Emotionen werden Akteure in die Lage versetzt, soziale Situationen in Anlehnung an die frühere emotionale Erfahrung zu kategorisieren und auf dieser Kategorisierung basierend in den Entscheidungs- bzw. Handlungswahl-Situationen solche Verhaltensalternativen auszuwählen, die in der Vergangenheit mit positiven emotionalen Episoden einhergingen. Dagegen werden Verhaltensweisen vermieden, die in der Situation gleicher Kategorie in negativen emotionalen Zuständen resultierten. Dadurch, dass Missachtung sozialer Normen mit negativen Emotionen assoziiert wird, wird das nichtkonforme Verhalten bei jeder neuen Entscheidungsepisode als weniger vorteilhaft eingestuft als Normkonformität. Der Akteur wird demzufolge die normkonformen Verhaltensweisen kontinuierlich vorziehen (es sei denn, seine aktuellen Ziele (BDIs) stehen im Widerspruch mit den sozialen Normen).³

Demnach kann das Netz aus Figur 6.28 ergänzt werden um eine zusätzliche Transition **internalize social norms, change internal mental representation**. Mit dieser Transition wird der von Damasio vorgestellte Prozess der emotionsbasierten Situationskategorisierung abstrakt dargestellt. Die vom Akteur 1 generierten negativen Emotionen der Scham, Schuld oder Verlegenheit werden mit der vorangegangenen Situation des Normbruchs bzw. dem nichtkonformen Verhalten assoziiert. Nach mehrmaligem wiederholten Auftreten der Situationen des nichtkonformen Verhaltens, das von Beobachtern durch Expression ihrer

³Allgemein betrachten wir hier nicht die Rolle der aktuellen BDIs bei der Entscheidung für bzw. gegen normkonformes Verhalten). Offensichtlich bestimmt in einem solchen Fall die Gewichtung der aktuellen individuellen Ziele, inwieweit der Akteur bereit sein wird, die erwarteten negativen Emotionen in Kauf zu nehmen.

6.2. Neurobiologische Grundlagen von Emotion

negativer sozialen Emotionen sanktioniert wird, was bei dem Sanktionierten wiederum negative Emotionen auslöst, wird diese Situationskategorie bei dem Normbrechenden mit einem automatischen Generieren der assoziierten negativen Emotionen einhergehen. Auch in Abwesenheit der (materiellen und immateriellen) Strafsanktionen, z.B. falls das nicht konforme Verhalten durch andere Akteure nicht beobachtet oder nicht bemerkt wurde, greifen negative Emotionen als interne Bestrafungsmaßnahme.

Die mentale Repräsentation sozialer Normen - bzw. die Verknüpfung bestimmter Verhaltensweisen und assoziierter Emotionen - muss ggf. aktualisiert werden, z.B. wenn diese Repräsentation nicht (mehr) der wahren Norm entspricht, sich die in der Gesellschaft akzeptierten Normen ändern oder der Akteur sich an eine Norm einer anderen Gesellschaft / Gruppe anpassen muss. Die Aktualisierung mentaler Normrepräsentation erfolgt nach dem gleichen Mechanismus wie die initiale Norminternalisierung, z.B. wird eine Verhaltensweise des Akteurs, die mit seiner internen Norm konform ist, wiederholt von den anderen Gesellschaftsmitgliedern sanktioniert. Diese Handlungsweise wird dann allmählich nicht mehr mit positiven sozialen Emotionen, die normalerweise auf Belohnung des normkonformen Verhaltens durch die Gesellschaft folgen, sondern mit negativen sozialen Emotionen, die Sanktionierung nach sich ziehen, assoziiert. So werden die mit dem fraglichen Verhalten verknüpften Emotionen geändert und damit die interne mentale Repräsentationen der Normen aktualisiert.

Entsprechend stellt das Modell in Abb. 6.29 die Internalisierung der normkonformen Verhaltensweisen durch positive Emotionen dar (vgl. [Dam03], [Els96], [Els99]). Die Tendenzen zur Kooperation innerhalb der Gruppe werden demnach durch die positiven Emotionen *Bewunderung*, *Dankbarkeit* und *Anerkennung* von Seiten der anderen Akteure bzw. durch *Stolz* auf die eigene Person belohnt und dadurch werden die kooperativen Verhaltensweisen in der Gruppe weiter angeregt und stimuliert. Norm-konformes Verhalten des Akteurs 1 (Transition conform to social norms, keep behavioral rules / conventions) löst positive soziale Emotionen wie Lob, Anerkennung und Bewunderung (Stelle admiration, appreciation, approval) in den Beobachtern aus (Transition generate social emotions). Der absichtliche oder unbewusste Ausdruck dieser Emotionen (Transition express social emotions) gegenüber dem normkonformen Akteur löst bei ihm Emotionen wie Selbstzufriedenheit, Stolz und Befriedigung aus (Stelle content, pride, self-satisfaction). Diese positiven sozialen Emotionen werden nach mehrmaliger Wiederholung der entsprechenden Situation mit Normbefolgung bzw. Kooperation assoziiert. Beim erneuten Auftreten von Situationen der gleichen Kategorie werden die assoziierten positiven Emotionen die Handlungswahl des Akteurs beeinflussen und sie hin zur normkonformen Verhalten steuern, um eine positive interne emotionale Rückmeldung herbeizuführen. So werden soziale Normen und Verhaltensregeln von dem Akteur internalisiert (internalize social norms, change internal mental representation) und sein Verhalten fortan zur Normkonformität und Kooperation hin steuern.

Als indirekten Beweis seiner Theorie der emotionsbasierten Internalisierung von sozialen Normen und Verhaltensregeln zeigt Damasio anhand von klinischen Beobachtungen, dass eine mangelhafte Entwicklung von sozialen Emotionen und eine fehlende Verbindung von sozialen Verhaltensweisen und emotionalen Belohnungen bzw. Bestrafungen in Form von positiven bzw. negativen Emotionen dazu führt, dass das Individuum die gesellschaftlichen

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

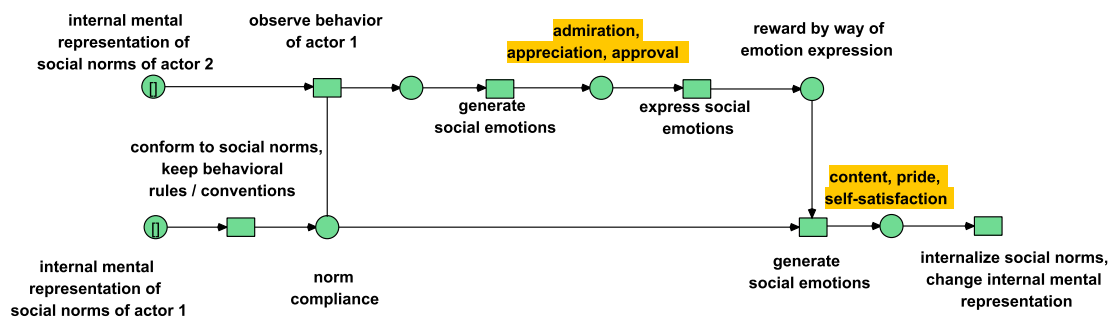


Abbildung 6.29.: Internalisierung des norm-konformen Verhaltens nach [Dam03].

Normen, Gruppenkonventionen und soziale Verhaltensregeln nicht verinnerlichen kann. Eine Gemeinschaft von mehreren Akteuren mit solchen defekten Emotionsmechanismen ist nicht funktionsfähig, da in einer solchen Gesellschaft weder soziale noch moralische und ethische Normen etabliert werden können ([Dam03], S. 162).

6.3. Soziologische Erklärungsansätze zur Emotion

Nachdem Emotion als Forschungsgegenstand der Soziologie lange Zeit eine Residualkategorie blieb, zeigt sie seit etwa drei Jahrzehnten eine rasche Weiterentwicklung. Nach den wegweisenden und initiiierenden Publikationen vor allem von US-amerikanischen Soziologen wie Randall Collins [Col75b], Theodore Kemper [Kem78] und Arlie Hochschild [Hoc79] wurde in der Soziologie zunehmend erkannt, dass Emotionen ein fundamentaler Bestandteil soziologischer Prozesse sind. Emotionen wurden nicht mehr als "irrationale Störfaktoren" rationaler und objektiver Erklärungsbestrebungen gesehen, sondern als grundlegende Mechanismen, die für das "Funktionieren" des Individuums als Teil einer Gruppe und auch für die interne Gruppendynamik von zentraler Bedeutung sind.

Als einer von wenigen Emotionssoziologen hat sich Kemper mit der Frage nach der Legitimation einer Soziologie der Emotionen beschäftigt, insbesondere mit der Abgrenzung der Emotionssoziologie zur Emotionspsychologie und den dort bereits vorhandenen Ansätzen [Kem84]. Kemper stellt fest, dass psychologische Emotionstheorien in erster Linie die Einflüsse externer - unter anderem auch sozialer - Stimuli auf ein Individuum untersuchen und sich mit der Frage beschäftigen, welche Einflüsse diese Stimuli auf die Ziele, Ansichten oder Absichten, die Persönlichkeit oder das allgemeine psychische oder physiologische Empfinden eines Akteurs haben und welche Emotionen daraus resultieren.

Die soziologischen Ansätze sollten nach Kempner explizit die *sozialen* Umgebungsstimuli untersuchen, die für die Emotionsentstehung eine Rolle spielen. Die sozialen Einflussfaktoren sind nach Kemper "social class, bureaucracy, normative order, division of labor, hierarchy, power, status, and the like" ([Kem84], S. 370); diese sollen von der Emotionssoziologie definiert, kategorisiert, konzeptualisiert und systematisiert werden. Davon ausgehend formuliert Kemper generelle soziologischen Forschungsfragen bei der Erforschung der Emotion: "[T]he sociological position is secure: either we investigate the social cues that trigger innate emotions, or we investigate the socialization conditions that instill emotions." ([Kem78], S. 31).

Für die Informatik bedeutet diese Sichtweise, dass soziologische Ansätze eine wichtige Ergänzung zu psychologische Emotionstheorien bilden können und deshalb bei der Umsetzung der Funktionalität des natürlichen Emotionsmechanismus in technische Systeme berücksichtigt werden sollten. Speziell betrifft dies den Aspekt der emotionalen Ausstattung der Agenten in MAS zur sozialwissenschaftlichen Theorienmodellierung: Wenn z.B. psychologische Emotionstheorien auf der Agentenebene und soziologische Theorien auf der Ebene der MAS implementiert werden, können die Zusammenhänge und Wechselwirkungen zwischen den beiden Ebenen untersucht werden. Diese Herangehensweise bietet die Möglichkeit der Integration der bislang nebeneinander stehenden psychologischen und soziologischen Emotionstheorien. Sie ist insbesondere für den psychologischen bzw. soziologischen Theorieentwurf und dessen Validierung interessant, könnte aber auch im Hinblick auf funktionale Systemoptimierung neue Ergebnisse liefern. Z.B. lässt die gleichzeitige Modellierung der emotionspsychologischen und -soziologischen Theorieansätze in einem Multiagentensystem neue Erkenntnisse in Bezug auf das Problem der Kontrollierbarkeit bzw. das Mikro-Makro-Problem erwarten ([vSM04]; siehe auch Kapitel 2.2).

Nachdem in Kapitel 6.1 die psychologischen Ansätze dargestellt wurden, welche zur

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Emotionsspezifikation auf der Ebene des Akteurs / der Agenten (Mikroebene) angewendet werden können, und in Kapitel 6.2 die neurobiologischen Mechanismen geschildert wurden, welche für die emotionalen Prozesse der Mikroebene eine physiologische Grundlage bilden, sollen in diesem Abschnitt ausgewählte soziologische Ansätze dargestellt werden, die die Rolle der Emotion auf der Ebene der Gesellschaft / MAS (Makroebene) aufdecken sowie die möglichen Wechselwirkungen zwischen den beiden Ebenen spezifizieren.

Die Analyse der psychologischen Ansätze in Kapitel 6.1 hat einige wesentliche Schwachstellen der psychologischen Einschätzungstheorien gezeigt, welche besonders häufig in den informatischen Emotionsmodellen umgesetzt werden. So weisen die meisten Theorien bezüglich der Berücksichtigung sozialer Faktoren, speziell im Einschätzungsprozess, erhebliche Insuffizienzen auf. In diesem Zusammenhang können die soziologischen Theorien der Emotion eine wichtige Ergänzung zu den einschätzungstheoretischen Ansätzen bilden (vgl. [vSM04], [vS04], [Fix04]).

In Abgrenzung zu psychologischen Theorien definiert Kemper das Potenzial soziologischer Emotionsforschung wie folgt: “They [sociologists; Anm. d. Verf.] can provide a conceptual model of the social settings that cue the specific appraisals and comparisons which, according to cognitive theorists, precede emotions“ ([Kem78], S. 31).

Ferner hat die Betrachtung psychologischer Theorien gezeigt, dass die meisten Einschätzungstheorien nur den Emotionsgenerierungsprozess spezifizieren, jedoch keine Aussagen über die Auswirkungen der generierten Emotion auf kognitive und exekutive Funktionen (z.B. auf das Entscheiden, Schlussfolgern, Planen, Handlungswahl, Motivation usw.) liefern. In Kapitel 6.1.3 haben wir zusammenfassend festgestellt, dass die meisten Einschätzungstheorien ihren Erklärungsanspruch auf interne psychologische Prozesse, die den Emotionen zugrunde liegen, beschränken, ohne jedoch soziale Determinanten und Funktionen der Emotionen zu berücksichtigen. So kann mit Hilfe der Einschätzungstheorien lediglich die Generierung der Emotionen unter Einfluss der Umweltfaktoren spezifiziert werden, nicht jedoch der Einfluss der Emotion auf Strukturen der Makroebene wie soziale Normen und auch nicht umgekehrt den Einfluss dieser Strukturen auf Emotion bzw. Emotionsgenerierung.

Eine wichtige Frage in Bezug auf die Optimierung der (Multi-)Agentensysteme ist, inwieweit und wie Emotionen Handlungswahl und Entscheidungsfindung der Agenten beeinflussen und auf diese Weise zur Strukturation der Agentengesellschaft beitragen (siehe Kapitel 2.2). Von Bedeutung sind für diese Fragestellung Ansätze, die eine soziologische Spezifikation für den Prozess der (emotionsgesteuerten) Handlungswahl liefern und den Zusammenhang von emotionalen Handlungsdeterminanten auf der Mikro- und Makroebene beleuchten. Da Norm als eine wichtige soziologische Handlungsdeterminante der Makroebene zu betrachten ist, ist die Frage nach dem Zusammenhang von Emotion und sozialer Norm im Hinblick auf die sozionalen Probleme der Strukturation und Kontrollierbarkeit der Multiagentengesellschaften wichtig.

Für die Emotionsmodellierung in MAS sind also insbesondere solche theoretischen Erklärungsansätze von Interesse, die soziale Determinanten und Funktionen der Emotion beschreiben. In diesem Abschnitt soll deshalb eine Reihe von ausgewählten soziologischen Theorien der Emotion zusammengestellt werden, welche soziale Einflussgrößen bei der Emotionsgenerierung und soziale Funktionen der Emotion näher spezifizieren.

6.3. Soziologische Erklärungsansätze zur Emotion

Bei der Auswahl der Theorien wurde ein besonderer Schwerpunkt auf solche Ansätze gelegt, die die Frage nach den Wechselwirkungen der Emotion und der sozialen Norm explizit betrachten, da diese für die Probleme der Strukturation und Kontrollierbarkeit der Multiagentengesellschaften besonders grundlegend und interessant ist. Leider ist die Frage, wie Genese und Aufrechterhaltung von sozialen Normen im Zusammenhang mit Emotionen stehen, in der Soziologie noch wenig erforscht und die existierenden Ansätze bedürfen systematischer Aufarbeitung.

In Kapitel 6.2.2 wurde bereits die Hypothese von Damasio [Dam94] vorgestellt, welche einige wichtige Indizien für die Annahme der Schlüsselrolle der Emotionen bei der akteur-internen Verankerung der sozialen Normen [vSM04] liefert. Nach dieser Annahme können soziale Normen und Verhaltensregeln durch geeignete Positionierung der akteursinternen Emotionen implementiert werden.

Damasio zeigt, wie der Mechanismus der emotionsbasierten Situationskategorisierung dafür sorgt, dass die mit negativen sozialen Emotionen (z.B. Scham, Schuld, Verlegenheit) assoziierten Handlungen unter bestimmten situationalen Gegebenheiten nicht ausgeführt werden, während mit positiven sozialen Emotionen (z.B. Selbstzufriedenheit, Stolz, Anerkennung) verknüpfte Handlungsoptionen von den Akteuren vorgezogen werden. So kommen Regelmäßigkeiten im Handeln der Akteure zustande, die als soziale Normen bezeichnet werden. Die Norm, die als eine wichtige soziologische Handlungsdeterminante der Makroebene zu betrachten ist, wird auf diese Weise unter Mitwirkung der sozialen Emotionen auf der Mikroebene fortdauernd (re-)internalisiert. In diesem Abschnitt sollen nun weitere Ansätze vorgestellt werden, welche die von Damasio aus der neurowissenschaftlichen Perspektive beschriebenen Mechanismen aus Sicht der Soziologie beleuchten.

Somit hat die nachfolgende Theorienübersicht zum Ziel, die in Bezug auf die eben erläuterten Fragestellungen relevanten soziologischen Ansätze zu identifizieren und skizzenhaft zu beschreiben. Die Anknüpfungspunkte an die Agentenkonzepte sowie die Integrationsmöglichkeit mit psychologischen Emotionstheorien werden bei der Auswertung der Ansätze angedeutet.⁴

6.3.1. Emotionen bei soziologischen Klassikern: Simmel, Durkheim, Weber, Parsons

Obwohl Emotion als soziologischer Forschungsgegenstand in der klassischen Soziologie und der kognitiv-normativ orientierten Soziologie der Nachkriegszeit größtenteils regelrecht ignoriert wurde ([Fla03], S. 114), lassen sich bereits bei den soziologischen “Klassikern” Ansätze ausmachen, die sich mit der Rolle der Emotion für die sozialen Entwicklungen in der Gesellschaft beschäftigen, u.a. auch mit ihrer Bedeutung für die Gesellschaftsstrukturation und Normkonstituierung.

Georg Simmel hat den Emotionen eine zentrale Rolle bei der Entwicklung von dauerhaften Institutionen zugeschrieben und sie als wichtige Inhalte des Handelns, der Triebe oder bei Eigeninteressen eingestuft. Treue und Dankbarkeit fungieren nach Simmel als “Bindemittel

⁴An dieser Stelle sei auf die Arbeiten von Moldt und von Scheve hingewiesen, die unter einer ähnlichen Zielsetzung mehrere soziologische Theorien untersuchen.

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

der Gesellschaft“ ([Fla03], S. 15), die Individuen miteinander und mit gesellschaftlichen Institutionen verbinden und so soziale Beziehungen und die Gesellschaftsstruktur konstituieren [Sim01]. Scham sei nach Simmel “mit der Verletzung irgendeiner Norm (sachlichen, sittlichen, konventionellen, personellen) verbunden“ ([Fla03], S. 21) und habe zum Ziel, das Individuum dazu zu bringen, sich wieder “moralisch“ zu benehmen. Damit griff Simmel Parsons Definition der Scham als “allmächtigem Instrument der sozialen Kontrolle“ (zit. nach [Fla03]) vor.

Auch Goffman sah die Vermeidung des äußerst schmerzhaften Auftauchens der Emotion Scham bzw. Verlegenheit (*embarrassment*) als Hauptmotiv der Individuen in ihren sozialen Interaktionen [Gof56].

Emile Durkheim hat ebenfalls die Zusammenhänge von Emotion und Verstößen gegen Normen bzw. deren Wiederherstellung untersucht. Er behauptet, dass eine Handlung kriminell ist, wenn sie starke und bestimmte Zustände des Kollektivbewusstseins verletzt und mit einer Strafe, die “in erster Linie aus einer leidenschaftlichen Reaktion“ besteht, belegt wird ([Dur95]; zit. nach [Fla03], S. 63). Die “kollektiven Gefühle“ können Verletzungen bzw. Angriffe besonders intensiv spüren und ein Individuum reagiert auf sie mit sehr intensiven Gefühlsreaktionen und Handlungen, die sich gegen den Angreifer richten. In einer späteren Arbeit werden die “Kollektivgefühle“ von Durkheim expliziter als *soziale Tatsachen* definiert, die Zwänge auf das Individuum ausüben und so das individuelle Handeln beeinflussen [Dur95]. Diese *sozialen Tatsachen* können unabhängig vom Einzelnen existieren und eine Eigendynamik entwickeln. Sie sind in der Lage, Individuen zu kollektivem Handeln zu bewegen. Wären die Wechselwirkungsmechanismen zwischen Durkheims “Kollektivgefühlen“ und kollektivem Handeln genauer spezifiziert, ließen sich diese Konzepte auf die MAS übertragen und z.B. zur Koordination bzw. Kooperation von Agenten anwenden.

Auch Max Weber hat in seinen Schriften Emotionen einen zentralen Stellenwert eingeräumt. U.a. stellte er fest, dass die Anomie, also Regellosigkeit, die bei einem Systemzusammenbruch eintreten kann, wenn die alten Normen und Regeln unwirksam werden, von hoher Emotionalität begleitet wird. In einem normativen Vakuum oder Chaos steigern sich die Angst und Verunsicherung der Akteure bis zu Hysterie und kompletter Handlungsunfähigkeit [Web22].

Talcott Parsons stellte in seiner *Handlungstheorie* die These auf, dass soziale Integration durch gemeinsame Wertorientierung bzw. gemeinsame Moral garantiert wird. Er nahm die Existenz von “moralischen Gefühlen“ an, die sich bei Verstößen gegen Normen offenbaren und die Aufrechterhaltung der gemeinsamen Wertorientierungen bzw. der gemeinsamen Moral ermöglichen ([Par94], S. 170f.). Der Akteur reagiert mit Empörung auf die Normverstöße der anderen und empfindet Schuldgefühle, Beschämung und Gewissensbisse, wenn er selbst die Normen verletzt hat. Sogar die unterdrückten Schuldgefühle, die von der Psychoanalyse entdeckt wurden, zählt Parsons zu den Reaktionen auf eigene Normverletzung oder dem Wunsch nach Verletzung von Normen, denen sich das Individuum bewusst überhaupt nicht moralisch verpflichtet fühlt ([Par94], S. 171).

Moralische Achtung und moralische Empörung stellen nach Parsons die affektiven Korrelate zu normativen Handlungsmustern dar. Aus diesen Emotionen (Respekt und Empörung) leitet sich der sog. *Wert-Komplex* ab - ein Komplex, der “in der Struktur sowohl der indivi-

duellen Persönlichkeit wie auch der sozialen Systeme eine äußerst wichtige Stellung hat“ und für soziale Integration sorgt ([Par94], S. 83). Im Unterschied zu Durkheim sind es bei Parsons also nicht die Normen, die Emotionen auslösen, sondern im Gegenteil konstituieren die “moralischen Gefühle“ die Normen mit.

In seiner Handlungstheorie postuliert Parsons drei *Handlungsorientierungen*: “die primären, funktional bedeutsamen Arten der Beziehung zwischen dem Akteur, seiner Situation und den relevanten normativen Mustern“ ([Par51], S. 69). Die Handlungssituation wird bei Parsons so beschrieben, dass die Anknüpfung sowohl an die einschätzungstheoretische Situationsauswertung mit anschließender Emotionsgenerierung als auch an die BDI-Konzepte gut realisierbar erscheint: “Im Kontext normativer Orientierung steht der Akteur als aktiver Agent in Beziehung zur Situation. Die Situation besteht für ihn in einer Ansammlung von Mitteln und Bedingungen oder in Hindernissen für die Verwirklichung seiner Zwecke“ ([Par51], S. 73f.).

In dieser Situation basiert das Handeln letztendlich immer auf Normen, die ihrerseits auf Emotionen und Werthaltungen zurückgehen [Fla03].

Durch *pattern variables* versucht Parsons ferner, analytisch zu erfassen, wie der Akteur die Situation definiert und sein Handeln bestimmt. Hier interessiert es ihn vor allem, wie dieser die Werte wählt, nach denen er handelt. Die Handlungssituation wird also als eine Entscheidungssituation verstanden, in der der Akteur zwischen alternativen normativen Werten wählen muss.

Parsons arbeitet *Orientierungen* des Handelns heraus, die der Akteur verwendet, um seine Beziehung zu einer Situation zu definieren. Zu den *pattern variables* gehören die Entscheidungen zwischen Alternativen: Emotionalität vs. emotionale Neutralität, Selbstorientierung vs. Kollektiv-Orientierung, Universalismus vs. Partikularismus usw. Diese Wahlentscheidungen stellen die einzelnen Selektionsschritte bei der Handlungswahl dar.

Parsons Thesen scheinen mit den einschätzungstheoretischen Konzepten durchaus kombinierbar. Da die meisten Einschätzungstheorien nur den Emotionsgenerierungsprozess spezifizieren, jedoch keine Aussagen über die Auswirkungen der generierten Emotion liefern, könnten sie möglicherweise um die Spezifizierung der Handlungswahl nach Parsons ergänzt werden. Dieser Gedanke soll hier jedoch nicht weiter vertieft werden.

Andererseits fehlt bei Parsons eine genaue Spezifizierung, wie die sozialen Bewertungen im Form von “moralischen Gefühlen“ der Akteure entstehen. Hier könnte eine Einschätzungstheorie, etwa OCC-Theorie, eine Abhilfe schaffen, weil sie die Modalitäten zur Einschätzung der anderen Akteure bzw. ihrer Handlungen bereits implementiert hat (vgl. Kapitel 6.1.2).

6.3.2. Emotionen in der modernen soziologischen Forschung

Der Beginn der Soziologie der Emotionen wird auf die Mitte der 1970er Jahre datiert ([Fla03], S. 117), als nach einigen wegweisenden Publikationen ([Col75b], [Kem78], [Hoc79]) die Emotion als expliziter Forschungsgegenstand soziologischer Untersuchung wiederentdeckt wurde. In den in dieser Zeit erscheinenden Arbeiten werden Fragestellungen aufgeworfen zum Zusammenspiel von Emotionen, Physiologie und Kognition und zur Rolle der Emotion als intervenierende Variable, die die Interaktionen auf Mikroebene mit Makrostrukturen verbindet.

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Innerhalb der aktuellen soziologischen Emotionsforschung lassen sich zwei wesentliche Strömungen unterscheiden [Fla03]: *konstruktivistische* - und als Unterkategorie *symbolisch-interaktionistische* - Theorien sowie *positivistische* Ansätze.

Konstruktivistische / symbolisch-interaktionistische Ansätze: Hochschild, Shott, Averill

Konstruktivistische Theorien betonen die soziale Konstruktion der Emotionen und vertreten die These, dass Emotionen das Produkt sozialkultureller Deutungs- und Interpretationsmuster sind ([Sho79], [Hoc79], [Tho89]). Sie sind von kulturspezifischen Deutungsmustern, Gefühlsvokabularen und Normen abhängig und weisen nicht notwendigerweise ein spezifisches biologisch-physiologisches Korrelat auf. Soziale Normen und Regeln, Habitualisierungen und Sozialisation bestimmen darüber, wie Individuen Emotionen erfahren, sie reflektieren und managen und sie kommunizieren.

Shott [Sho79] und Averill [Ave80] betonen, dass Emotionen im Laufe von Sozialisationsprozessen erlernt werden. Da aber die physiologischen Manifestationen von verschiedenen Emotionen oft sehr ähnlich und schwer zu unterscheiden sind, sind es letztendlich die interagierenden Individuen, die ihre Emotionen im Laufe ihrer sozialen Interaktion konstruieren. "Sie definieren, ob das, was sie empfinden, Angst oder Trauer, diffuse Aufregung oder Vorfreude ist. Soziale Strukturen und Normen stellen bloß die Rahmen dieser ständigen Konstruktionen dar" ([Fla03], S. 128).

Streng konstruktivistische Ansätze betrachten Emotionen als sozial konstruierte Konzepte, die so gut wie keine Verbindungen zu physiologischen Stimuli aufweisen. Da diese Sichtweise sich kaum an emotionspsychologische Erkenntnisse anschließen lässt und auch innerhalb der Soziologie relativ isoliert dasteht, wird sie hier nicht weiter betrachtet.

Symbolisch-interaktionistische Ansätze ([Hoc79], [Sho79], [Ave80]) sind eine Teilmenge der konstruktivistischen Ansätze, welche die Notwendigkeit einer kognitiven Interpretation der physiologischen Stimuli durch den Akteur zur Identifizierung und Wahrnehmung einer Emotion annehmen. Die Ausdifferenzierung der Emotionen in "physiologische Stimuli" und aufgrund von kognitiven Auswertungen "akut gefühlter Emotionen" [Hoc79] passt gut zu Damasio's Emotionsdefinition (siehe Kapitel 5.5), die entsprechend metabolische Prozesse und primäre Emotionen unterscheidet.

Shott behauptet, dass sich soziale Kontrolle auf Selbstkontrolle reduzieren lässt. In jeder Interaktionssituation entstehen zwei Typen von Emotionen: *selbstreflexive* Emotionen (Scham, Schuld, Stolz, Eitelkeit) und *empfindende* Emotionen, die entstehen, wenn man sich in die Rolle des anderen hineinversetzt und dessen Emotionen nachempfindet [Sho79]. Beide Emotionstypen motivieren moralische Handlungen und ermöglichen soziale Kontrolle akteurintern, indem sie signalisieren, "dass die Moral, das Selbstbild oder das reflektierte Selbst des Individuums reparaturbedürftig sind und Selbstkorrekturen bewirken." ([Fla03], S. 129).

Übereinstimmend mit Goffman, Simmel und Parsons zeigt Shott, dass unter den selbstreflexiven Emotionen Schuldgefühle, Scham und Peinlichkeitsgefühle eine kritische Rolle spielen: "[S]ie bringen Individuen dazu, sich normkonform zu verhalten" (ebd.).

Hochschild [Hoc79] nimmt an, dass Individuen sich nicht nur normwidrig fühlen, sondern auch in der Lage sind zu bewerten, ob ihre Gefühle nicht normkonform und korrektur-

6.3. Soziologische Erklärungsansätze zur Emotion

bedürftig sind. Das Gefühlsmanagement erfolgt nach *Gefühlsregeln* (*feeling rules*, *display rules*), also Normen, die spezifizieren, welche Gefühle wann, wie und mit welcher Intensität ausgedrückt werden dürfen. Gefühlsregeln stellen ein zentrales Konzept symbolisch-interaktionistischer Emotionstheorien dar und verweisen auf die generelle Bedeutung sozialer Normen und Regeln sowohl bei der Emotionsmanifestation als auch bei der Emotionsgenese (vgl. [Hoc79], [Sho79], [Tho89]). Hochschild führt ein weiteres wichtiges Konzept ein: die *Interpretationsrahmen* (*Frames*) [Hoc79]. Anhand der Interpretationsrahmen können verschiedene Situationen unterschiedlich interpretiert werden, so dass jeweils andere Gefühlsregeln gelten und verschiedene Emotionen legitim sind. Ein Akteur kann auf diese Weise in der gleichen Situation je nach Interpretationsrahmen diametral verschiedene Emotionen empfinden.

Mit der Bewusstwerdung der Nonkonformität seiner Emotion muss der Akteur *Emotionsmanagement* bzw. *Gefühlsarbeit* leisten, um die Qualität und Intensität seiner Empfindungen zu ändern [Hoc79]. kann er versuchen, den Interpretationsrahmen der Situation zu ändern, die die Emotion begleitenden physischen Symptome zu kontrollieren bzw. zu unterdrücken oder jene Verhaltensweisen zu simulieren, die eine normkonforme Emotion begleiten, “in der Hoffnung, dass sich diese Emotion dann auch einstellt“ ([Fla03], S. 130).

Zusammenfassung und Bewertung

Die symbolisch-interaktionistischen Ansätze konzentrieren sich in erster Linie auf Individuen und ihre Interaktionen und sind deshalb für die informatische Zielsetzung der Verbesserung von Benutzerinteraktion und Benutzer-Modellierung besonders interessant.

Diesen Ideen zufolge besteht die Möglichkeit, die sozialen Kontrollmechanismen (nämlich die Emotionen anderer) auf der Ebene des Agenten als selbstreflexive und einfühlende Emotionen zu implementieren. Indem der Agent diese Emotionen selbst ableitet, kann er Rückmeldung über die Normadäquatheit seines Verhaltens bekommen, ohne mit anderen Agenten interagieren zu müssen, und ggf. sein Verhalten korrigieren.

Ferner könnte evtl. die bessere Kontrollierbarkeit und Koordination in MAS über gemeinsame Interpretationsrahmen für alle MAS-Agenten implementiert werden. Nach [Hoc79] wird die Mikroebene mit der Makroebene über die Interpretationsrahmen und Gefühlsregeln verbunden, die durch soziale Institutionen der Makroebene formuliert und von Akteuren bedingt verinnerlicht werden. Entsprechend könnten die Agenten in einem MAS über gemeinsame Interpretationsrahmen zu kooperativem normkonformem Verhalten gebracht werden, indem sie eine bestimmte Situation gleich interpretieren und darauf gleiche Emotionen generieren, die wiederum zu gleichartigen kooperativen Handlungen führen.

Positivistische Ansätze: Kemper, Collins

Positivistische Ansätze ([Col75b], [Kem78]) gehen von der Annahme aus, dass Emotionen notwendigerweise eindeutige physiologische Korrelate aufweisen und als automatische Reaktion des Individuums auf externe Auslösefaktoren entstehen. Es wird dabei eine relationale Beziehung zwischen physiologischen Körperreaktionen, Emotionen und expressiven Verhaltensweisen vorausgesetzt. Ferner gehen Positivisten von der Existenz einer beschränkten

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Menge von Basisemotionen aus (siehe Kapitel 8.6), die relativ unabhängig von kulturellen Mustern in die Subjekte eingebettet sind, Situationen mit Sinn und Bedeutung versorgen und somit eines der wenigen Symbol- und Verständigungssysteme darstellen, das nicht notwendigerweise kulturell verankert sein muss, um in der Interaktion wirksam zu werden. Ziel der positivistischen Vorgehensweise ist es dementsprechend zu analysieren, welche sozialstrukturellen Situationen und Konstellationen spezifische Emotionen auslösen.

Kempers “Social interactional Theory of Emotions“ zufolge lässt sich jede soziale Beziehung anhand der Dimensionen Status (*voluntary*) und Macht (*coerced*) analysieren. Macht wird als Fähigkeit definiert, eigene Ziele auch gegen den Widerstand des anderen durchzusetzen. Status meint freiwillige Gefolgschaft, Anerkennung, Respekt, Unterstützung usw., die Akteure einander gewähren [Kem84].

Soziale Strukturen definiert Kemper als eine hierarchische Gliederung von Akteuren auf der Macht-Status-Achse. Er unterscheidet drei verschiedene Emotionsklassen, die von sozialen Strukturen und sozialen Beziehungen definiert werden (vgl. [Kem78]):

- Strukturelle Emotionen entstehen durch relativ stabile, sich wiederholende, ritualisierte Sozialbeziehungen, die durch eine bestimmte *Macht-Status-Konfiguration* gekennzeichnet sind. Diese Macht-Status-Konfiguration lässt persistente emotionale Stimmungslagen entstehen, die Kemper “strukturelle Emotionen“ nennt.
- Antizipatorische Emotionen entstehen immer dann, wenn Akteure eine Veränderung der Status- und Machtverhältnisse, z.B. als Ausgang bzw. Ergebnis einer Interaktion, gedanklich vorwegnehmen.
- Folgernde Emotionen resultieren aus Interaktionsprozessen, die eine Veränderung des Status-Macht-Gefüges zur Folge haben.

Collins thematisiert explizit Kempers Macht- und Statusdimensionen insbesondere in Bezug auf die Interaktionen sozialer Gruppen. Er führt das Konzept der *emotionalen Energie* ein: Soziale Gruppen tauschen in *Situations- und Machtritualen* emotionale Energie aus; ihre Menge bestimmt letztendlich über die Mächtigkeit und den sozialen Rang der Gruppe ([Fla03], S. 159f.).

Zusammenfassung und Bewertung

Kempers Theorie ist im Wesentlichen auf Interaktionssituationen fokussiert und bietet deshalb viele interessante Möglichkeiten für die Verbesserung von Benutzerinteraktion und Benutzer-Modellierung durch Emotionseinbettung. Offensichtlich erscheint die Möglichkeit der Ergänzung von Multiagentenarchitekturen um die sozialwissenschaftlich fundierte Mechanismen zur Modellierung der Handlungsreaktionen der anderen Agenten in bestimmten Situationen interessant (z.B. Handlungen anderer Agenten durch antizipatorische Emotionen erklären und vorhersagen). Bislang wurde sich der Problematik der Verhaltensmodellierung anderer Agenten in MAS eher durch intuitive Modellierungsansätze genähert (vgl. die COO-Komponente des Affective Reasoner in Kapitel 7.3.1)).

Der besondere Vorteil von Kempers Theorie als Ergänzung psychologischer Emotionstheorien um soziale Einflussgrößen der Emotionsgenese liegt darin, dass er soziale Faktoren

der Emotionsgenese betrachtet, die sowohl in psychologischen Emotionstheorien als auch in symbolisch-interaktionistischen Ansätzen weitestgehend vernachlässigt werden. Interessant ist vor allem auch das logisch-kausale Regelmuster, ähnlich der OCC-Theorie (vgl. Kapitel 6.1.2), mit dem Kemper die Emotionsgenese beschreibt. Eventuell ließen sich mit Kempers Modell einige Einschätzungstheorien ergänzen: So wurde die Ergänzung des OCC-Modells in Bezug auf die Einschätzung anderer Akteure / Agenten durch Kempers Status-Macht-Matrix vorgeschlagen ([vS04], S. 100f.).

Für die technische Optimierung der Agentensysteme erscheint Kempers Theorie wegen ihres Fokus auf soziale Interaktion eher weniger geeignet: “[T]he interaction model proposed here is concerned neither with specific motives nor with mechanisms to induce compliance, but with the *intentional behavior* of actors toward each other. The emphasis is on the link *between* actors“ ([Kem78], S. 29; Hervorh. i.O.; zit. nach [vS04], S. 94). Kemper macht zwar formale Ausführungen über die kausale Rolle der sozialen Strukturen bei der Entstehung von Normen [Kem84], diese werden aber nicht präzisiert und wurden hier deshalb auch nicht näher betrachtet.

6.3.3. Die Rolle der sozialen Emotionen bei der Etablierung und Aufrechterhaltung sozialer Normen: Modellierung und Integration der Ansätze

Der Überblick über die soziologischen und neurowissenschaftlichen Emotionstheorien in den Kapiteln 6.3 und 6.2 ließ bereits einen engen Zusammenhang zwischen den internen Emotionen der Akteure und der sozialen Kontrollstrukturen der Makroebene erkennen. Wir konnten feststellen, dass in einer Reihe der soziologischen Theorien eine essenzielle Rolle der Emotion bei der Ausbildung der sozialen Normen in einer Gesellschaft postuliert wird. Auf diese Ergebnisse zurückgreifend, werden wir in diesem Abschnitt die Möglichkeit der Emotionsimplementierung zur flexiblen Kontrolle und Koordination eines Multiagentensystems diskutieren, wie sie in [SP00] vorgeschlagen wurde (vgl. Kapitel 2.2).

Staller und Petta betonen: “[T]he computational study of social norms can profit by theories, which suggest that emotions play an important role in sustaining norms“ ([SP00], S. 1). Wir wollen diese These aufgreifen und zeigen, wie mit Hilfe der Emotionstheorien verschiedener Disziplinen eine Wechselbeziehung zwischen Emotionen und sozialen Normen spezifiziert werden kann.

Aus den betrachteten sozialwissenschaftlichen und neurowissenschaftlichen Theorien lässt sich die Aussage ableiten, dass Emotion eine Schlüsselrolle bei der Etablierung und Aufrechterhaltung sozialer Normen spielt, indem sie für die Internalisierung von sozialen Normen durch einzelne Akteure eine notwendige neurobiologische Basis bildet. Durch die Übertragung dieser Mechanismen auf eine Multiagentengesellschaft soll die Steuerung und Koordination des Gesamtsystems durch die Strukturen der Mikroebene (hier Emotionen) erfolgen: Die emotionsbasierte Verankerung sozialer Normen in einem Agenten eines Multiagentensystems und die dadurch bedingte Erzwingung normkonformen Verhaltens macht Emotion zu einem Schlüsselmechanismus für die Etablierung der Kontrollstrukturen auf der Makroebene. Zur Unterstützung dieser These stellen wir eine Reihe von theoretischen

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Grundlagen zusammen, welche für die Existenz einer wechselseitigen Beziehung zwischen Emotion und sozialen Normen argumentieren.

Im Gegensatz zu psychologischen und neurowissenschaftlichen Theorien der Emotion fokussieren die soziologischen Theorien im Wesentlichen auf die Makro-Funktionen der Emotion, wie z.B. die Rolle der Emotion bei der sozialen Strukturation, während sie z.B. die Emotionsgenerierung und weitere emotionale Prozesse der Mikro-Ebene in einem einzelnen Akteur i.d.R. außer Acht lassen. Die Aufgabe besteht also in diesem Abschnitt darin, zusätzlich zur Modellierung von soziologischen und neurowissenschaftlichen Erklärungsansätzen verschiedene Perspektiven auf Emotion zu untersuchen und zu integrieren, um ein generelles Modell der Wechselwirkungen der Emotionen und sozialen Normen als einen speziellen Fall der Mikro- und Makro-Strukturen zu erstellen.

Im vorigen Abschnitt wurden soziologische Ansätze vorgestellt, die für eine Schlüsselrolle der Emotion beim Aufbau und Aufrechterhaltung der allgemeingültigen Verhaltensnormen in einer Gesellschaft argumentieren. Ein Fazit aus den betrachteten Ansätzen kann wie folgt zusammenfasst werden: Durch bestimmte Emotionen (z.B. Verachtung, Abscheu) wird ein Akteur für das Nichtbefolgen gemeinsamer Verhaltenskonventionen bestraft. Andererseits ist das Nichtbefolgen der gesellschaftlichen Verhaltensnormen für den betroffenen Akteur mit den Emotionen der Scham und Schuld verbunden, oft sogar wenn kein anderer Akteur das nichtkonforme Verhalten beobachtet hat. Auf diese Weise kann durch Emotionen eine Internalisierung der gemeinsamen Verhaltensnormen durch jedes einzelne Gesellschaftsmitglied stattfinden und folglich die Ausbildung und Aufrechterhaltung der sozialen Normen in einer Gesellschaft erfolgen.

Nachstehend wird diese Hypothese auf Basis von weiteren theoretischen Ansätzen ausgeführt, wobei die grundlegenden theoretischen Aussagen schrittweise in Petrinetzmodelle übersetzt werden, so dass am Ende die Kernaussagen der Ansätze in einer Reihe von Petrinetzen anschaulich dargestellt werden. Bei der Modellierung der Theorieansätze geht es insbesondere auch darum, einen Vergleich der Ansätze möglich zu machen, so dass die Gemeinsamkeiten in den Kernaussagen deutlich werden und über die Kompatibilität der Theorien diskutiert werden kann. Die Operationalisierbarkeit der theoretischen Ansätze dient von Anfang an als Ausschlusskriterium für die Petrinetz-Modellierung: In diesem Abschnitt werden vor allem solche Ansätze für die Modellierung ausgewählt, die nicht nur allgemein der Emotion eine wichtige Rolle bei der Etablierung sozialer Normen zuschreiben, sondern auch explizit zeigen, wie eine solche Normetablierung und -aufrechterhaltung durch Emotion auf der Mikroebene (d.h. akteursintern) funktioniert. Dies sind insbesondere die soziologischen Ansätze von Elster und Elias, die zu den neurowissenschaftlichen Untersuchungsergebnissen von Damasio (siehe Kapitel 6.2.2) in Beziehung gesetzt werden.

Wir gehen folgendermaßen vor: Zuerst spezifizieren wir die Definition der sozialen Norm und fassen die sozialen Funktionen von Emotionen auf einer allgemeinen Ebene zusammen. Ferner zeigen wir, welche Funktionen Emotionen bei der Etablierung der sozialen Strukturen und sozialen Normen in natürlichen Gesellschaften übernehmen und veranschaulichen die beschriebenen theoretischen Ansätze mit Petrinetz-Modellen. Schließlich wird ein allgemeines Modell der sozialen Kontrolle hergeleitet, das auf der Relation zwischen sozialen Normen und Emotionen basiert.

Soziale Normen

In [Cas01] werden grundlegende theoretische Fragen nach der Möglichkeit unbewusster, ungeplanter und emergenter Formen der Kooperation, Organisation und Intelligenz zwischen deliberativen, planenden BDI-Agenten formuliert. In [vSMFvL06] haben wir diskutiert, dass diese Frage nicht ohne Betrachtung des Einflusses der Emotion auf normkonformes Verhalten und auf soziale Strukturdynamiken beantwortet werden kann.

Für den hier verfolgten Ansatz, die Wechselwirkungen zwischen Normen und Emotion darzustellen, ist es insbesondere Erkenntnis versprechend, soziale Normen als “mental objects“ [CC95a] anzusehen und damit vor allem auch ihre mentalen, für kognitive und emotionale Prozesse ausschlaggebenden Komponenten zu betrachten. Als mentale Objekte sind soziale Normen einerseits als Strukturelemente der Makroebene anzusehen, da sie sozial ver- und geteilt sind, gleichzeitig sind sie jedoch auch - aufgrund ihres Einflusses auf die Entscheidungsfindung und Handlungswahlprozesse - der Mikroebene zuzurechnen (vgl. [CN89]).

Diese Sichtweise wird vergleichbar auch in [CC95b] vertreten. Die Autoren definieren Normen als “behavioural regularities which emerge from the strategic agent’s choices to conform“ [CC95b] und bringen sie mit BDI-Konzepten in Verbindung. In ihrer Eigenschaft als Verhaltensvorschriften, die als *Beliefs* des Agenten repräsentiert werden, bestimmen Normen maßgeblich zukünftiges Handeln, indem sie bei der Generierung neuer Ziele mitwirken: “[T]hey represent a powerful mechanism for inducing new goals in people’s minds in a cognitive way“ ([CC95a], S. 189). Die zentralen Fragen jedoch - “[H]ow and why does a normative belief come to interfere with x’s decisions? What is it that makes her [an actor] responsive to norms concerning her? What is it that makes a normative belief turn into a normative goal?“ ([CC95a], S. 192) - lassen die Autoren offen.

Gerade die Beantwortung dieser Fragen ist jedoch von herausragender Bedeutung, wenn man sich dem in [Cas01] formulierten Problem der ungeplanten emergenten Formen der Kooperation in MAS auf der theoretischen Ebene annähern will. Diverse empirische und theoretische Untersuchungen zur Dualität von Kognition und Emotion (z.B. [Fri94], [LeD96b]) zeigen, dass die Betrachtung von Emotion für das Verständnis der Mechanismen der aktuerinternen Normetablierung unabdingbar ist.

Nach Elster lassen sich soziale Normen als Vorschriften für das Verhalten mit den folgenden Eigenschaften erfassen ([Els99], S. 145f.):

1. Normen sind nicht unmittelbar am Nutzen orientiert (“non-outcome-oriented“), d. h. sie beziehen sich auf die Aktionen als solche und sind im einfachsten Fall von der Art “Do X“ or “Do not do X“. Dagegen wird die Handlungswahl eines rationalen Akteurs, die nicht durch Normen beeinflusst wird, im Allgemeinen dem Grundsatz folgen: “If you want to achieve Y, do X“. Wenn der von einer sozialen Norm ausgedrückte Imperativ konditional ist, ist er nicht zukunftsorientiert, sondern z.B. von der Art “If others do Y, then do X“.
2. Soziale Normen werden mit anderen Mitgliedern einer Gesellschaft bzw. einer sozialen Einheit geteilt. Einige Normen werden von allen Mitgliedern einer Gesellschaft geteilt, andere sind für bestimmte Gesellschaftsgruppen spezifisch.

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

3. Da Normen gemeinhin sozial geteilt werden, können sie von Mitgliedern einer sozialen Einheit mittels Sanktionierung erzwungen werden. Präsenz (oder mentale Repräsentation) anderer Akteure ist für die Sanktionierung abweichenden Verhaltens notwendig, d.h. Normen werden aufrechterhalten, indem die Normmissachtung durch die bestrafenden Sanktionen anderer Akteure geahndet wird.

Im Folgenden werden wir zeigen, dass Emotion nicht nur maßgeblich am Erhalt sozialer Normen beteiligt ist, sondern umgekehrt auch am vermeintlich rationalen Handeln der Akteure. Dabei unterliegen sie dem Einfluss der sozialen Umwelt, in der sich Akteure befinden, die sich wiederum durch das Vorhandensein bestimmter Normen auszeichnet. Emotion und soziale Normen stehen also in einem reziproken Verhältnis zueinander, so dass Emotion nicht ohne soziale Normen und soziale Normen nicht ohne Emotion vollständig analysiert und erklärt werden können. Unter Einbeziehung von Forschungsergebnissen aus [Dam03], [KH99], [Els96], [Els99], [Eli76], [Sch03] und [SP00] könnte man argumentieren, dass weder externe Bestrafung noch rationale Interessen allein dazu führen, dass der Akteur sich den sozialen Normen unterordnet, sondern dies wird vielmehr verursacht durch die Antizipation der negativen Emotionen, die infolge der Normbrechung akteursintern generiert werden.

Soziale Funktionen der Emotion

In Kapitel 5.4 wurden die funktionalen Emotionsdefinitionen diskutiert. Es wurde gezeigt, dass eine grundlegende Funktion von Emotion in der Steuerung des Verhaltens im Hinblick auf eine optimale Anpassung an die Umweltsituation besteht (vgl. auch exemplarisch [KH99], S. 468). Hier konzentrieren wir uns vor allem auf soziale Aspekte von Emotion, die sich aus der sozialen Situiertheit des Akteurs ergeben bzw. in diese hineinwirken. In diesem Zusammenhang ist eine Spezifizierung der Emotionsdefinition und auch des zu Grunde gelegten Funktionsverständnisses notwendig. Emotionen werden als “functional, organized responses to environmental demands that prepare and motivate the person to cope with the adaptational implications of those demands“ [SP81] verstanden, wobei die angesprochenen Anforderungen (*demands*) sowohl physischer als auch sozialer Natur sein können.

Bei der Analyse der sozialen Emotionsfunktionen werden meist mehrere Analyseebenen eingeführt, wobei die meisten vorgeschlagenen Unterteilungen sich nicht grundsätzlich, sondern lediglich in ihrem Abstraktions- bzw. Reduktionsgrad unterscheiden (vgl. auch [Tur02], S. 30-41). Beispielsweise verortet Averill die sozialen Emotionsfunktionen auf drei Analyseebenen (biologisch, psychologisch, sozial) [Ave80], Keltner und Haidt auf vier (“individual“, “dyadic“, “group“, “cultural“) [KH99], Gerhards ebenfalls auf vier Ebenen (Organismus, Persönlichkeit, soziale Struktur und Kultur) und v. Scheve und Moldt wiederum auf drei Ebenen (mikro, meso, makro) [vSM04]. Des Weiteren lassen sich intraindividuelle ([Lev99], [Lev03]), ontogenetische [AI99] und phylogenetische [Tur99] soziale Emotionsfunktionen unterscheiden.

Akteursintern kommen Emotionen vor allem zwei Funktionen zu. Zum einen die, den Akteur über diejenigen Ereignisse in der (sozialen) Umwelt zu informieren, die eine Reaktion erfordern ([CSC94], [Sch90]). Beispielsweise informiert Ärger über die empfundene Fairness

6.3. Soziologische Erklärungsansätze zur Emotion

einer Handlung, Liebe über den Grad an Zuneigung und Verbundenheit (*commitment*), Scham und Verlegenheit über die Normkonformität einer Handlung usw. (vgl. [KH99]). Zum anderen bereiten Emotionen den Akteur darauf vor, auf Anforderungen in sozialen Situationen adäquat zu reagieren, beispielsweise durch physiologische Veränderungen ([CSC94], [Sch90]). Ganz ähnlich werden auch in [OJ96] die intraindividuellen Funktionen von Emotion zum einen in der Vorbereitung der Ausführung von Handlungen und zum anderen in der Strukturierung des kognitiven Systems in bestimmte Operationsmodi gesehen.

Auf der intraindividuellen, d.h. der expressiv-kommunikativen Ebene sind vor allem drei soziale Funktionen der Emotion von Bedeutung:

- Erstens erlaubt die Expression von Emotion eine kontingente Zuschreibung von emotionalen Zuständen, Ansichten, Intentionen und korrespondierenden Situations-einschätzungen.
- Zweitens evozieren Emotionsexpressionen der anderen komplementäre oder reziproke Emotionen im Alter Ego und tragen so zu einer verbesserten bilateralen Situations-einschätzung bei.
- Drittens befördern oder behindern Emotionsmanifestationen das Verhalten von Interaktionspartnern, indem sie entweder motivierenden oder sanktionierenden Charakter aufweisen (vgl. [KH99]).

Die soziale Funktion von Emotion in größeren sozialen Einheiten zeigt sich an ihrem Beitrag zur Identifikation von sozialen Gruppen und Gruppenmitgliedern [Dur95], an der Zuschreibung von Status- und Machtressourcen [Kem78] sowie an der Internalisierung und Aufrechterhaltung von sozialen Normen, Machtstrukturen, Moralvorstellungen und Ideologiekonzepten ([Els99], [Hoc79]).

Zusammenfassend lassen sich folgende soziale Funktionen der Emotion hervorheben, die in Bezug auf Etablierung sozialer Normen und sozialen Strukturierung von besonderer Bedeutung sind (vgl. [vSMFvL06]):

1. Emotion informiert einen Agenten über die Ereignisse in seiner sozialen Umwelt, die oftmals einer unmittelbaren Reaktion und / oder Anpassung des Verhaltens bedürfen [CSC94]. Sie bereitet so den Agenten darauf vor, auf die wahrgenommenen situationalen Begebenheiten adequat zu reagieren, indem zum Beispiel die physiologische Reaktionen evoziert und das kognitive System in den entsprechenden operationalen Modus versetzt wird ([CBA⁺02], [CSC94], [OJ96], [Fri86], [SP81], [Laz91]).
2. Emotion fungiert als Instrument sozialer Kontrolle. Die sog. “soziale“ Emotionen ([Dam94]; vgl. auch Kapitel 6.2.2) wie Scham, Verachtung, Verlegenheit oder Schuld werden unter anderem dadurch ausgelöst, dass der Akteur eine oder mehrere soziale Normen / Konventionen verletzt (vgl. [Sch84], [Fri86], [Dam94]). Die Auslösung erfolgt automatisch und unbewusst, so dass der Akteur normalerweise keinen willentlichen Einfluss darauf ausüben kann. In diesem Sinne wird z.B. Scham als primäre soziale Emotion gesehen, die eine entscheidende Rolle bei der Etablierung sozialer Bindungen

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

spielt, indem sie eine Verletzung sozialer Normen aufzeigt, die diesen Bindungen zugrunde liegen [Sch03].

In [KH99] wird die Rolle der Emotion (speziell Verlegenheit) für die Aufrechterhaltung sozialer Normen betrachtet. Es wird gezeigt, wie Verlegenheit als “appeasement function“ wirkt: Indem sie soziale Beziehungen, die sich nach der Verletzung einer sozialen Norm in Ungleichgewicht befinden, wieder in Einklang bringt, hilft sie, die beteiligten Akteure auszusöhnen und davor zu bewahren, die gleiche soziale Norm (wieder) zu brechen [KH99]. In der Theorie von Elster ([Els96], [Els99]) wird beschrieben, wie durch Scham und Verachtung soziale Normen etabliert und aufrechterhalten werden (vgl. Kapitel 6.3.3).

Mehrere weitere soziologische Ansätze verfolgen die These, dass Emotionen eine entscheidende Rolle bei der Aufrechterhaltung der sozialen Strukturen und der sozialen Normen spielen ([Fla03], S. 149). Den soziologischen Klassikern Simmel, Parsons und Goffman folgend, definieren auch Kemper, Scheff, Honneth und Neckel die Emotion Scham als “Hauptinstrument der Selbst- und der sozialen Kontrolle“ ([Fla03], S. 150). “Scham wird empfunden, wenn sich das Selbst im Vergleich mit den geltenden Normen oder, noch wichtiger, mit dem Selbstideal als mangelhaft empfindet. Scham, so Parsons, stellt einen subjektiven Kontrollfaktor dar, der für Selbstkorrekturen sorgt. Scham, so Neckel, bedeutet Angst vor dem Verlust sozialer Achtung“ ([Fla03], S. 156).

3. Emotionsgenerierung und -expression wird durch soziale Normen gesteuert. Die sozial-konstruktivistische Perspektive in der Soziologie der Emotion (z.B. [Hoc79], [Tho89]) sieht Emotionen als sozial konstruierte Phänomene, die der bewussten / absichtlichen Lenkung, Anpassung, Aufrechterhaltung oder Unterdrückung durch den Akteur zugänglich sind (z.B. [Hoc79]).

Die soziale Konstruktion der Emotion wird vor allem durch kulturell definierte, sozial geteilte Normen ermöglicht, die den Akteur darüber informieren, ob eine Emotion bzw. eine Emotionsexpression in einer gegebenen sozialen Situation adäquat bzw. angebracht ist. Hochschild definiert in diesem Kontext das Konzept der “feeling rules“ und “expression rules“ - kulturell bedingte Vorgaben für angemessene emotionale Wahrnehmung und emotionale Reaktion ([Hoc79]; vgl. auch Kapitel 6.3.2).

In [EF75] wird eingehend das Konzept der “display rules“ diskutiert, die die kulturabhängigen Standards für Emotionsausdruck festlegen. Die Autoren unterscheiden vier Strategien zur Umsetzung der “display rules“: Durch “minimization“ wird der exponierte emotionale Ausdruck minimiert, d.h. die emotionale Reaktion wird zurückgehalten oder versteckt; durch “maximization“ wird diese Reaktion übertrieben; “masking“ ist der Versuch, Neutralität bzw. Gleichgültigkeit vorzutäuschen; “substitution“ heißt, dass eine andere Emotion anstelle der Empfundenen vorgetäuscht wird.

Ebenso unterliegt die *Generierung* von Emotionen dem Einfluss sozialer Normen. So definieren mehrere Einschätzungstheorien das Einschätzungskriterium “Legitimität“, bei dem bewertet wird, inwiefern eine fremde oder eigene Handlung mit den gesell-

schaftlichen Normen und Konventionen in Einklang steht (z.B. [Fri86], siehe auch Kapitel 6.1.2).

Nach [Sch93] bewertet die Kompatibilität (*compatibility check*) ein Ereignisses bzw. eine Handlung hinsichtlich der Normen der Moral, Legitimität und Gerechtigkeit (siehe Kapitel 6.1.2). Überprüft wird hierbei einerseits, ob eine eigene oder fremde Handlung sozialen Normen sowie kulturellen Konventionen oder Regeln entspricht (*external standards subcheck*), andererseits, inwieweit eine Handlung mit den eigenen Normen und Standards als Teil des Selbstkonzepts oder Selbstideals übereinstimmt (*internal standards subcheck*). Diese Einschätzungsdimension erlangt Bedeutung vor allem im Zusammenhang mit "sozialen" Emotionen wie Verachtung, Scham, Schuld und Ärger.

4. Der verbale und nonverbale Emotionsausdruck ermöglicht den interagierenden Akteuren die gegenseitige Zuschreibung der meisten durch den Interaktionskontext (mit-)bestimmten Variablen, wie emotionaler Zustand, Situationseinschätzung und Absichten.

Voraussetzung für diese gegenseitige Zuschreibung ist jedoch der gleiche kulturelle Hintergrund und Sozialisationskontext: Die gleiche soziale Prägung bzw. Emotionssozialisation ermöglicht es den Agenten, sich gegenseitig relativ zuverlässig Ansichten, Wünsche und Intentionen zuzuschreiben, die sie aufgrund der gleichartig geprägten Emotionsexpressionen im Laufe der Interaktion erschließen können. "Auf diese Weise wirkt einerseits die durch ein soziales Umfeld geprägte emotionale Handlungs- und Verhaltenskontrolle direkt in die unmittelbare soziale Umgebung zurück, andererseits stellt die Rezeption einer Emotionsexpression durch andere Akteure in Aussicht, in eben diesen Akteuren "reaktive" Emotionen hervorzurufen, die eine entsprechend konforme oder konträre Handlungsregulation zur Konsequenz haben." ([vSM04], S. 207).

Wenn man annimmt, dass Akteure, deren Sozialisation innerhalb eines bestimmten sozialen Aggregats stattgefunden hat, über ähnliche Ausprägungen mentaler Objekte wie soziales Wissen über Normen und Konventionen verfügen, sind diese im Akteur verankerten sozialen Strukturen diejenigen Instrumente, die die Normkonformität des Akteurs und u.a. seinen Umgang mit den eigenen Emotionen und Emotionen anderer Akteure regulieren. "Das explizite Wissen über bestimmte Qualitäten eines sozialen Aggregats, hier insbesondere Normen, Regeln und Konventionen, veranlasst Akteure zu einem konsequenten Monitoring des eigenen emotionalen Zustands sowie der emotionalen Zustände von Interaktionspartnern." ([vSM04], S. 209).

Des Weiteren ermöglicht ein sozial geteiltes System der Emotionsausdrucksregeln einem Akteur, das Verhalten seines Interaktionspartners als emotional bzw. als Emotionsausdruck einzuordnen und daraus Rückschlüsse auf die dieser Verhaltensweise zugrunde liegenden Emotionen zu ziehen. So kann ein in den Kontext eingebundener Beobachter seiner ggf. emotionalen Reaktion auf das beobachtete Verhalten Ausdruck geben mit der Aussicht, dass der Akteur, der das Verhalten produziert hat, von dem Emotionsausdruck auf den emotionalen Zustand des Beobachters schließen kann.

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

Auf diese Weise bilden die sozial geteilten Ausdrucksregeln für Emotionen eine notwendige Basis für die Interaktionssequenzen, die die emotionsgeleitete Sanktionierung und Verstärkung von Verhaltensweisen konstituieren.

Elster ([Els96], [Els99]) argumentiert in diesem Sinne, dass Emotionen für die Sanktionierung nicht konformen Verhaltens essenziell sind. Seine theoretischen Annahmen werden im Folgenden dargestellt und modelliert.

Emotionsbasierte Etablierung und Aufrechterhaltung von sozialen Normen

Damit die zuvor beschriebenen sozialen Funktionen der Emotion strukturdynamische Effekte auf der Makroebene hervorbringen können, bedarf es der Vermittlung und Wirksamkeit derselben in der sozialen Interaktion, was in diesem Abschnitt näher untersucht wird.

Um das Verhältnis von Emotion und sozialen Normen im Hinblick auf das Strukturationsproblem künstlicher und natürlicher Gesellschaften näher zu beleuchten, muss zunächst klar sein, in welchem Verhältnis soziale Normen zum Handeln und Verhalten der Akteure stehen. Geht man davon aus, dass Normen das Handeln der Akteure restringieren, indem sie bestimmte Handlungsoptionen und -alternativen als inadäquat kennzeichnen, dann lassen sich Strukturierungseffekte derart vermuten, dass bestimmte Handlungen unter bestimmten situationalen Gegebenheiten nicht ausgeführt werden, andere Optionen von den Akteuren vorgezogen werden, und so Regelmäßigkeiten im Handeln der Akteure zustande kommen. Wir suchen hier jedoch in erster Linie nach akteursinternen Mechanismen, die diese Regelmäßigkeiten im Handeln der Akteure entstehen lassen. Mit anderen Worten wollen wir klären, wie normkonforme bzw. sozial akzeptierte Handlungsmuster bzw. Verhaltensweisen überhaupt entstehen und warum die den abweichenden Verhaltensweisen von dem Akteur wiederholt vorgezogen werden.

Ausschlaggebend für den hier vertretenen Ansatz ist die Art der Sanktionierung im Falle einer Nichtbeachtung sozialer Normen. Insbesondere in der ökonomischen Theorie wird die Sanktionierung abweichenden Verhaltens durch den Entzug materieller Ressourcen beschrieben ([Axe86], [Col91], [Els96]). Materielle Ressourcen jedoch sind keinesfalls als maßgebende Objekte der Sanktionierung anzusehen. Viel entscheidender in diesem Zusammenhang ist, dass sich abweichend verhaltende Akteure materielle Sanktionen immer auch als Vehikel für den Ausdruck der Verachtung, Geringschätzung, Ekel oder Abscheu sehen und in der Konsequenz Scham oder Schuld empfinden. Scham kann für den Sanktionierten als wesentlich schwerwiegender angesehen werden, da Scham - im Gegensatz zu Schuld - für Ego auch die Perspektive Alter Egos wesentlich deutlicher in Rechnung stellt und Ego eine akute Gefährdung sozialer Bindungen anzeigt [Sch03].

Elster führt dazu aus, dass der zentrale materielle Aspekt von Sanktionen lediglich in der Frage liegt, wie viel es den Sanktionierenden kostet und nicht, wie viel die Sanktionen den Sanktionierten kosten bzw. wie belastend diese Sanktionen sind ([Els99], S. 146).

Je höher die Kosten sind, die ein Sanktionierer bereit ist in Kauf zu nehmen, um eine beabsichtigte Strafsanktion umzusetzen, desto eindringlicher nimmt der Sanktionierte die darin liegende Verachtung wahr und desto deutlicher kann er sich das Ausmaß der durch seine Aktionen ausgelösten negativen Emotionen vorstellen, die die entsprechenden Sanktionen verursacht haben und die mit Sanktionen verbundenen Aufwendungen rechtfertigen. Folglich

6.3. Soziologische Erklärungsansätze zur Emotion

wird sich der Bestrafte desto beschämter und schuldbewusster fühlen, je mehr negative Emotionen er bei dem Beobachter seiner normbrechenden Aktionen vermutet. Die Höhe der Kosten für den Sanktionierer signalisiert demzufolge dem Sanktionierten die Schwere seines abweichenden Verhaltens und die Ernsthaftigkeit seines Normbruchs.

Nach Elias sind Scham und Peinlichkeitsempfindungen verantwortlich dafür, dass Personen ihr Verhalten an gesellschaftliche Standards binden. Verstößt ein Akteur gegen die Haltung der “anderen“, erzeugt das in ihm Scham: “[D]as ist die Furcht von sozialer Degradierung und Überlegenheitsgesten anderer“ [Eli76]; verstoßen andere dagegen, erzeugt das Verhalten Peinlichkeit. Die bewusste oder unbewusste Kenntnis der sozialen Normen versetzt den Akteur wiederum in die Lage, erwartungskonform / an Normen orientiert zu handeln. Die normkonformen Verhaltensweisen werden von der Gesellschaft mit Lob / Anerkennung honoriert, wodurch dem Akteur Emotionen des Stolzes und der Freude induziert werden.

Einen ähnlichen Blick auf die Wirkung sozialer Normen in Verbindung mit Emotion nimmt auch Frijda ein, der soziale Zurückweisung in Form emotionaler Sanktionierung mittels Scham oder Verachtung als “severe punishment, [...] most likely not merely because of its more remote adverse consequences“ beschreibt ([Fri86], S. 351).

Diese Annahmen konnten in jüngster Vergangenheit durch die sozialen Neurowissenschaften bestätigt werden. [EL04] zufolge liegen sozialem Schmerz, ausgelöst beispielsweise durch Verachtung und Zurückweisung, dieselben neuronalen Korrelate zugrunde wie physischem Schmerz. Welche Konsequenz haben nun diese Möglichkeiten der Sanktionierung durch Emotionen auf Seiten der Sanktionierer und der Sanktionierten für das Handeln der Akteure?

Das Streben nach emotionaler Gratifikation, d.h. die Motivation, positive emotionale Zustände anzustreben und negative zu vermeiden, wird von vielen Autoren als grundlegender Motivationsmechanismus des Menschen angesehen. So geht Turner davon aus, dass Ängstlichkeit (“anxiety“) als einer von sechs primären Motivatoren angesehen werden kann, wobei diese als “need to avoid a sense of disequilibrium with the environment“ definiert wird ([Tur02], S. 21). Explizit können Emotionen - besonders Besorgnis - auch in Relation gesetzt werden zu zukünftigem Handeln, da sie maßgeblich in deren Planung eingreifen. So steht Sorge um einen Verlust ontologischer Sicherheit beispielsweise auch in Giddens Sozialtheorie an zentraler Stelle (Giddens 1991). Dort ist es vor allem die Furcht vor dem Verlust ontologischer Sicherheit und Faktizität, die Akteuren als Motivation für ihr Handeln dient.

Elias (1987) veranschlagt als elementare Bedingung der menschlichen Existenz ein Bedürfnis nach “Geben und Nehmen in emotionaler Beziehung, nach Sicherheit und Beständigkeit einer spontanen und emotional warmen Bejahung der eigenen Person“ ([vLMV03], S. 57). Die Rückseite dieses Bedürfnisses ist die Angst vor dem Verlust an Achtung und Akzeptanz durch andere Akteure.

Andere Autoren wie beispielsweise Collins (1984) oder Hammond (1991), die emotionale Gratifikation als Handlungsmotivator auch unmittelbar für skalierungsfähig und damit aussagekräftig in Bezug auf soziale Aggregation halten, unterstellen Akteuren ein Bedürfnis nach positiven emotionalen Austauschprozessen, die sich - so zumindest Collins - zu

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

“interaction ritual chains“ verfestigen können und damit strukturprägend wirken ([Col75b], [Col84]).

Aufgrund der hier erläuterten Ergebnisse kann festgehalten werden, dass besonders die Emotionen Scham und Verachtung als Vehikel für die Aufrechterhaltung sozialer Normen dienen, indem sie nach [CC95b] “n-goals (normative goals)” generieren. Normkonformität als Ziel wird also nicht unbedingt durch die bloße Antizipation des Verlusts materieller Ressourcen durch Sanktionen, sondern vor allem durch die Furcht vor emotionalen Sanktionen (Verachtung) und daraus resultierender Scham gespeist.

Emotionale Sanktionierung und Verstärkung nach Elster Elster stellt ebenfalls fest, dass die Sanktionierung des normabweichenden Verhaltens durch Emotionen der Verachtung und Abscheu geleitet wird. Dabei können “Sanktionen” lediglich darin bestehen, dass diese Emotionen auf subtile Weise zum Ausdruck gebracht werden, z.B. durch Gesten, Gesichtsausdruck, Kommunikationsverweigerung usw. Auch wenn dem normbrechenden Akteur dadurch keine materiellen Nachteile entstehen, wird die Effektivität der Sanktionen durch Emotionen der Scham und Schuld sichergestellt, die der Normbrecher empfindet, wenn er Sanktionen wahrnimmt, und zwar als “vehicle for the emotions of contempt or disgust” ([Els99], S. 146). Noch entscheidender ist in dieser Beziehung die Tatsache, dass sogar materielle Sanktionen üblicherweise von den Sanktionierten als Ausdruck von starken negativen Emotionen, wie Verachtung, Geringschätzung, Abneigung, Ekel usw. interpretiert werden. Die von Elster beschriebenen Mechanismen wurden teilweise empirisch und experimentell validiert ([FG02], [dQFT⁺04], [SRA⁺03], [EL04], [BABD02]).

Das Referenznetz in Abb. 6.30 zeigt die Modellierung der Bestrafung / Sanktionierung des abweichenden Verhaltens nach Elster ([Els96], [Els99]) aus der Perspektive eines Sanktionierten.

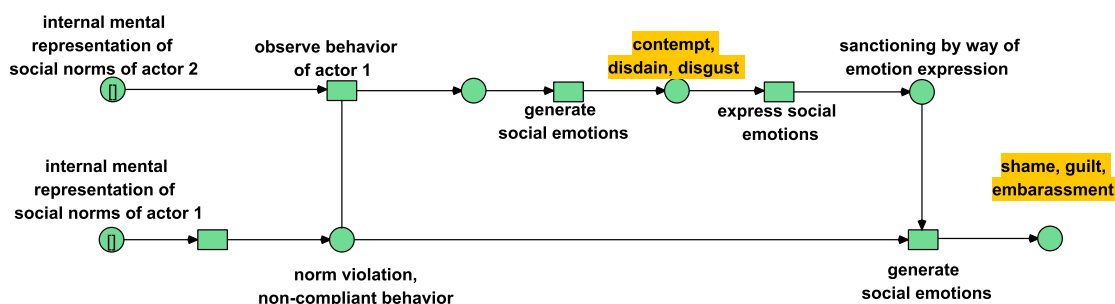


Abbildung 6.30.: Normbruch und Sanktionierung nach [Els99].

Entsprechend stellt das Modell in Abb. 6.31 die Verstärkung der norm-konformen Verhaltensweisen durch positive Emotionen dar ([Els96], [Els99], [Dam03]). Norm-konformes Verhalten des Akteurs 1 (Transition conform to social norms, keep behavioral rules / conventions) kann positive soziale Emotionen, wie Lob, Anerkennung und Bewunderung (Stelle admiration, appreciation, approval) in den Akteuren auslösen, die dieses konforme Verhalten beobachten (Transition generate social emotions). Die positiven Emotionen der Beobachter,

6.3. Soziologische Erklärungsansätze zur Emotion

die absichtlich oder unbewusst zum Ausdruck kommen, (Transition express social emotions), können ihrerseits in dem normkonformen Akteur Emotionen wie Selbstzufriedenheit, Stolz und Befriedigung auslösen (Stelle content, pride, self-satisfaction).

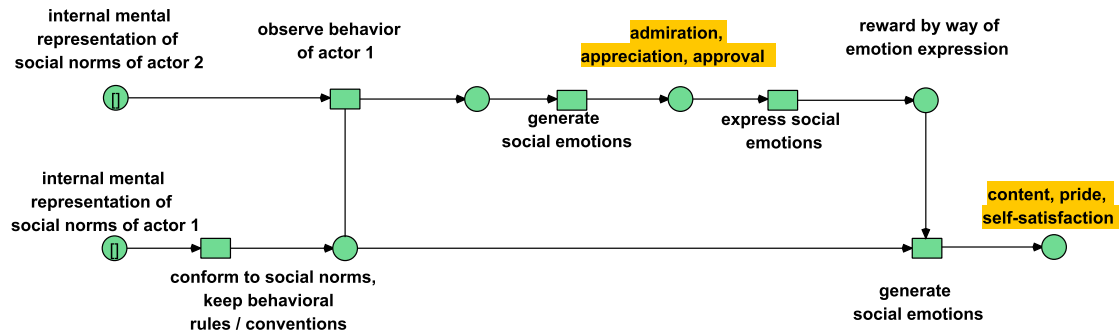


Abbildung 6.31.: Verstärkung der Normbefolgung nach [Els99].

Normetablierung und -internalisierung nach Elias Die Theorie von Elias liefert eine sehr ähnliche Konzeptualisierung für den Prozess der Aufrechterhaltung und Internalisierung von sozialen Normen, jedoch ohne explizite Bezugnahme auf Emotion als Hauptinstrument der Abbildung von sozialen Normen innerhalb eines individuellen Akteurs. In [vLMV03] werden drei Selbstregulationsmechanismen von sozialen Gruppen nach Elias mit Petrinetzen modelliert. Hier betrachten wir eines dieser Modelle, das die Normetablierung in einer sozialen Gruppe durch einen vielschichtigen Internalisierungsprozess von Zwängen darstellt, nach dem das normkonforme Verhalten reproduziert wird. Wir geben das entsprechende Petrinetzmodelle aus [vLMV03] wieder und veranschaulichen daran die für unsere Fragestellung relevanten Aspekte der Theorie von Elias. Für eine ausführliche Darstellung der Gesamtheorie sei an dieser Stelle auf [vLMV03] verwiesen.

Elias setzt voraus, dass Menschen auf eine soziale Umfeld existenziell angewiesen sind. Das Bedürfnis nach Akzeptanz und Anerkennung in einer sozialen Gruppe ist demnach ein Grundbedürfnis menschlicher Existenz. Mit diesem Bedürfnis hängt die Angst der Akteure vor dem Verlust der Achtung und sozialen Akzeptanz in der Gruppe zusammen. Um in die Gruppe aufgenommen und akzeptiert zu werden, müssen Akteure sich an die gruppeninternen Verhaltensnormen und -regeln halten. “Die wichtigsten und schwersten Ängste, die Personen haben können, beziehen sich auf den Verlust von Achtung und deshalb auf Verstöße gegen das, was die “Haltung der anderen“ erwartet (vgl. Abb. 6.33). Und darüber wird die fundamentale Tatsache, dass Menschen den Verlust von Liebe, Achtung, Anerkennung und Prestige fürchten, für Prozesse der Gruppenbildung und Verhaltenssteuerung relevant (vgl. [vLMV03], S. 57).

Um nicht gegen gesellschaftskonformes Verhalten zu verstoßen, muss ein Akteur seine Handlungen zwangsweise auf eine erhebliche Anzahl anderer Handlungen abstimmen. Das erfordert beständige Rück- und Voraussicht auf die Aktionen und Absichten anderer. Zugleich erfordert es, das eigene Verhalten selbst zu regeln und sich selbsttätig zu diszi-

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

plinieren. Diese Selbststeuerung und -regulierung der Individuen wird durch eine innere Instanz gewährleistet, die Akteure ausbilden, um ihr eigenes Verhalten und Empfinden zu beurteilen. Sie beobachten die wertende Haltung, die andere Akteure ihren Handlungen gegenüber einnehmen, und übernehmen deren Haltung, um sich selbst zu bewerten.

Die Instanz, die in [vLMV03] als Über-Ich bezeichnet wird, nimmt also die vermutete Haltung anderer Akteure ein und der Akteur beurteilt sich bald automatisch mit ihrer Hilfe, so dass das eigene Verhalten im Sinne der gesellschaftlichen Verhaltensnormen geregelt und umgeformt wird. Die Selbstzwänge entstehen auf Basis von schweren Ängsten, wie Angst vor sozialem Abstieg, Prestigeverlust, Zurechtweisung und Verachtung durch andere, und zwingen die Akteure zum konformen Verhalten. Verstößt ein Akteur gegen die Haltung der "anderen", erzeugt das in ihm Scham, verstoßen andere dagegen, erzeugt das Verhalten Peinlichkeit. Nach Elias sind also Scham und Peinlichkeitsempfindungen verantwortlich dafür, dass Personen ihr Verhalten an gesellschaftliche Standards binden.

Negative emotionale Erfahrungen bei Normverletzung bewirken die Internalisierung der sozialen Normen durch den Akteur. Diese werden im Unterbewusstsein (Über-Ich) des Akteurs gespeichert bzw. von ihm bewusst gelernt. So wird das Missachten sozialer Normen in dem Sanktionierten auch dann negative Emotionen wie Scham oder Schuld hervorrufen, wenn kein anderer Akteur die normbrechende Aktion beobachtet hat.

Soziale Zwänge (Stelle Fremdzwänge) - werden damit unter Einfluss der Scham bzw. Furcht vor Verlust der Achtung und Prestige in personale Zwänge (Stelle Selbstzwänge) umgewandelt. Im Zuge dieser Umwandlung werden Ängste vor der Übertretung gesellschaftlicher Norm in Form von Scham- und Peinlichkeitsempfindungen internalisiert (siehe Abb. 6.32).

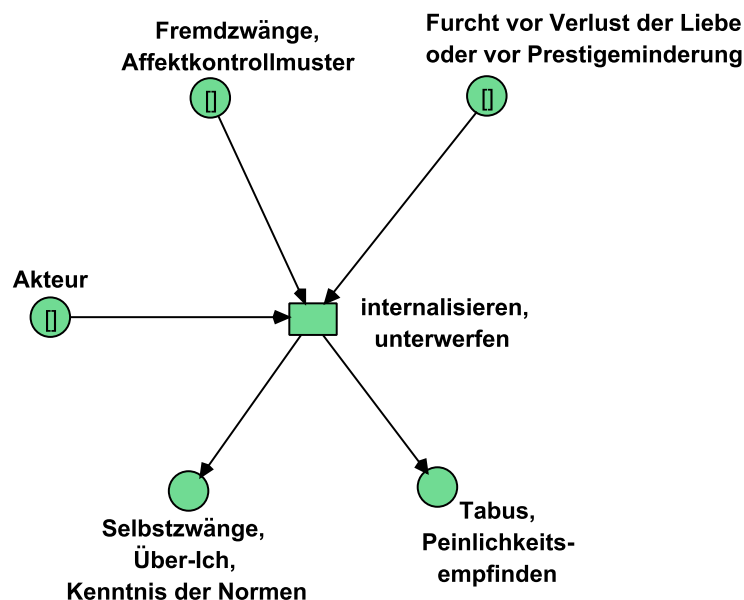


Abbildung 6.32.: Internalisierung der sozialen Zwänge (aus [vLMV03], S. 58)

6.3. Soziologische Erklärungsansätze zur Emotion

Die bewusste oder unbewusste Kenntnis der sozialen Normen versetzt den Akteur wiederum in die Lage, erwartungskonform / an Normen orientiert zu handeln. Die normkonformen Verhaltensweisen werden von der Gesellschaft mit Lob / Anerkennung honoriert, wodurch im Akteur Emotionen des Stolzes und der Freude aufkommen. Entsprechend wird das normkonforme Verhalten mit positiven Emotionen assoziiert und internalisiert; dies führt zu der Zunahme der normkonformen Verhaltensweise.

Ein Petrinetz in Abb. 6.33 macht den Mechanismus der Normetablierung in einer Gruppe deutlich: Akteure, die von einer Gruppenzugehörigkeit profitieren wollen, müssen sich an Gruppennormen anpassen und erwartungskonform / an Normen orientiert handeln. Dazu müssen solche Akteure einerseits diese Gruppennormen (bewusst oder unterbewusst) kennen und andererseits Selbstregulierungsmechanismen (Selbstzwänge) ausbilden, die nicht konforme Handlungsweisen (unmittelbare Affekte, Neigungen, Triebe der Akteure) unterdrücken. Unter der Voraussetzung, dass alle der Gruppe zugehörigen Akteure über Selbstregulierungsmechanismen und Normkenntnisse verfügen, können sie an Gruppennormen orientiert handeln, wodurch die Gruppennorm etabliert wird.

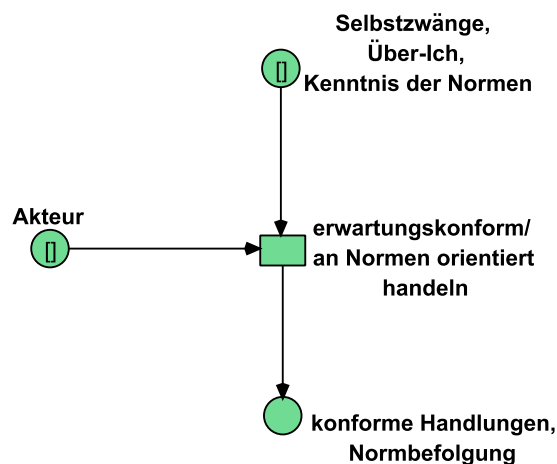


Abbildung 6.33.: Konformes Handeln nach Elias (aus [vLMV03], S. 58).

Die Angst vor dem Verlust oder der Minderung von Ansehen, Achtung und Prestige innerhalb der eigenen Gruppe (Scham) motiviert die Gruppenmitglieder dazu, sich selbst und die anderen hinsichtlich der Verhaltenskonformität unablässig zu beobachten (Abb. 6.34).

Die Befolgung der Gruppennormen erhöht und festigt die Anerkennung anderer Gruppenmitglieder für den Normbefolger und verschafft ihm Achtung, Lob und Prestigesteigerung. Indem die Gruppenmitglieder sich wiederholt zu den gruppenspezifischen Normen, Verhaltensregeln, Affektkontrollmuster etc. bekennen bzw. ihr Verhalten diesen Normen unterwerfen, festigen sie ihre Stellung innerhalb der Gruppe und erlangen dadurch eine Erhöhung der Achtung durch andere, aber auch Selbstachtung.

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

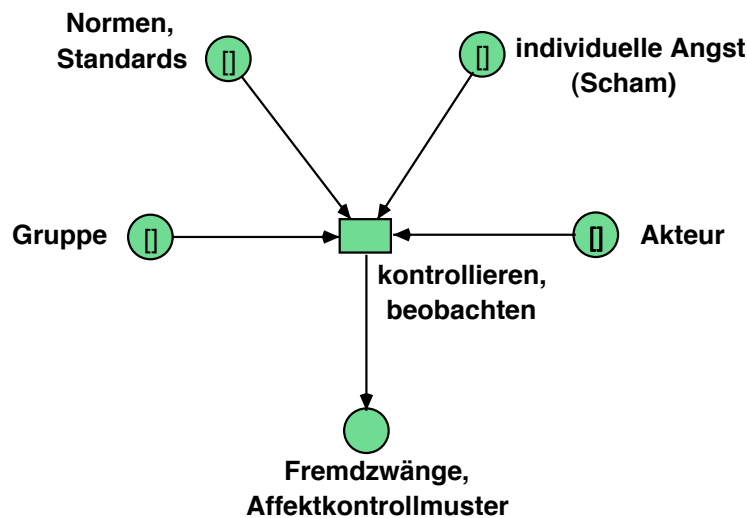


Abbildung 6.34.: Gegenseitiges Beobachten von Gruppenmitglieder (aus [vLMV03], S. 63)

Die Abweichung von den Gruppennormen und -forderungen wird hingegen von der Gruppe sanktioniert: der Verstoßende selbst schämt sich, den anderen Akteuren ist dieser Verstoß peinlich. Akteure, die die Gruppennormen dauerhaft nicht sehr streng befolgen, werden von den Normbefolgern mit Verachtung und Überheblichkeit als unterlegen und sozial minderwertig stigmatisiert ([ES90], S. 243).

Es bildet sich eine negative Gruppenmeinung über einen solchen Akteur, die sich in Ablehnung, Diskriminierung, Geringschätzung bis hin zur negativen emotionalen Wahrnehmung der Verstoßenden äußern kann. Diese Gruppenmeinung wirkt als Regulativ für das Empfinden und Verhalten der Gruppenmitglieder: Sie kontrolliert die emotionale Wahrnehmung der anderen. Auf Basis dieser emotionalen Wahrnehmung wird der Beobachter das Verhalten anderer entweder als Normverstoß einstufen und sanktionieren (durch Abwertung und Geringschätzung des verstoßenden Akteurs) oder als normkonformes Verhalten bewerten und honorieren (durch Anerkennung und Achtung). Auf diese Weise werden die Gruppennormen gesetzt, bekräftigt oder modifiziert⁵ (vgl. Abb. 6.35). So besteht soziale Kontrolle aus der gegenseitigen Beobachtung, Bewertung und Sanktionierung der einzelnen Akteure. Indem die Gruppenmitglieder sich selbst und andere andauernd hinsichtlich ihres Verhaltens überwachen und kontrollieren, „bringen sie die Gruppennormen zur Geltung und gewährleisten, dass sich das Verhalten und Empfinden der Gruppenmitglieder an diesen Normen orientiert“ (vgl. den Reproduktionszyklus in Abb. 6.35).

⁵Dieser Prozess illustriert das allgemeine Postulat von Elias, dass die emotionale Lage des Akteurs maßgeblich durch den Aufbau bzw. die Sozialstruktur der Gesellschaft bestimmt wird. Die Art des gewohnten Denkens, Verstehens und Fühlens eines Akteurs ist gesellschaftlich modelliert und damit abhängig von der gesellschaftlichen Lage. Die Veränderungen der Sozialstruktur ziehen Veränderungen des Wahrnehmens, Verhaltens und der psychischen Selbststeuerung nach sich.

6.3. Soziologische Erklärungsansätze zur Emotion

Elias postuliert zwar, dass Emotionen (Scham und Peinlichkeit) für die Anpassung des Verhaltens der Akteure an die gesellschaftliche Normen verantwortlich sind, spezifiziert jedoch nicht, auf welche Weise Emotion das Verhalten von Akteuren steuert.

In Kapitel 6.2.2 wurden neurowissenschaftliche Ergebnisse präsentiert, die eine wichtige Ergänzung zu eliaschen Beschreibung in Bezug auf die Spezifizierung der internen Mechanismen der emotionsbasierten Norminternalisierung bilden könnten. Es wurde der in [Dam03] beschriebene Prozess der Norminternalisierung dargestellt und mit Petrinetzen modelliert. Dabei wurde deutlich, dass negative soziale Emotionen, insbesondere Scham und Verlegenheit, als interne negative Konsequenzen der Normverletzung entstehen, mögliche externe Sanktionen seitens anderer Akteure antizipieren und so zur Internalisierung der sozialen Normen durch den Akteur beitragen.

In Ergänzung zu den Sichtweisen von Parsons, Kemper, Scheff und Elster liefert Damasio einen neurobiologischen Nachweis dafür, dass die menschliche Tendenz zum sozialen Einvernehmen durch den regulatorischen Apparat der Emotionalität zumindest zum Teil biologisch verankert ist. Der Ansatz von Damasio bietet somit eine detaillierte neurobiologische Erklärung für die Funktionsweise der von Elias beschriebenen emotionalen Mechanismen, die der Internalisierung von sozialen Normen zugrunde liegen (vgl. Kapitel 6.2.2 bzw. Abb. 6.28).

Kooperation als soziale Norm Zurückkommend auf das Problem der Kontrollierbarkeit und Kooperation in MAS, werden wir den Mechanismus der Normetablierung für den Fall explizieren, dass Kooperation eine akzeptierte soziale Verhaltensnorm in einer Multiagentengesellschaft ist. Hierbei nehmen wir Bezug auf die Darstellung in [Dam03], wo der Mechanismus der Etablierung von Kooperationstendenzen in einer Gesellschaft aus mehreren Akteuren geschildert wird. Danach werden Tendenzen zur Kooperation in einer solchen Gesellschaft durch emotionsbasierte Belohnung der Kooperierenden verstärkt und demzufolge die kooperativen Verhaltensweisen in der Gruppe weiter angeregt und stimuliert.

Eine Reihe weiterer Theorien aus der Soziologie vertritt ebenfalls die Auffassung, dass bestimmte Emotionen (z.B. Ärger) als Mittel bei der Bestrafung nicht kooperierender Gesellschaftsmitglieder dienen. Falls Kooperation und Reziprozität gesellschaftlich akzeptierte Verhaltensnormen darstellen, wird die Sanktionierung der Normmissachtung durch Emotionen realisiert. Gouldner (1960) behauptet, dass Reziprozität eine universale Norm darstellt, d.h. dass sie in allen menschlichen Gesellschaften existiert.

In seiner Theorie des reziproken Altruismus benennt Trivers (1979) eine Reihe von Emotionen, die er für notwendige Instrumente bei der Etablierung und Regulation des altruistischen Verhaltens hält. Z.B. dient "moralistic aggression" als Mittel der Bestrafung von Individuen, die die auf sie gerichteten altruistischen Handlungen nicht erwidern und keine Gegenleistungen erbringen. Die Emotion kann z.B. zur Einstellung aller zukünftigen altruistischen Aktivitäten gegenüber dem nicht kooperierenden Individuum ("cheater") führen. Schuld dient dazu, den Normbrecher zu Wiedergutmachungsmaßnahmen zu bewegen und dadurch die reziproke Beziehung bzw. Kooperation fortzusetzen (bzw. wieder aufzunehmen).

Diese Sichtweisen wurden im Petrinetz in 6.29 zusammengefasst. Entsprechend kann kooperatives Verhalten innerhalb einer Agentengesellschaft aus mehreren emotionalen Agenten durch positive Emotionen (wie Bewunderung, Dankbarkeit und Anerkennung) belohnt

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

werden. Diese Emotionen lösen ihrerseits positive Emotionen (wie Stolz und Selbstzufriedenheit) in dem kooperierenden Agenten aus. Auf diese Weise werden Kooperationstendenzen auf der Mikroebene verstärkt. Dadurch, dass in einem Agent kooperatives Verhalten mit assoziierten positiven Emotionen verknüpft wird und er dementsprechend im Normalfall die Kooperation der Verweigerung der Kooperation vorzieht, wird die soziale Norm der Kooperation auf der Mikroebene verankert. Dadurch, dass in jedem Agenten eines MAS kooperatives Verhalten zur Verhaltensnorm wird, wird Kooperation auch zu einer sozialen Norm der Makroebene.

6.3.4. Auswertung der soziologischen Emotionstheorien und offene Forschungsfragen

Die Soziologie der Emotionen erweist sich insgesamt als heterogenes Forschungsfeld: Zwar existiert eine Vielzahl von Arbeiten, die sich mit unterschiedlichen Aspekten von Emotionen und ihrer Bedeutung für soziologische Gegenstandsbereiche befasst - eine umfassende, systematische und verschiedene Perspektiven integrierende Analyse zu den Wechselwirkungen von Emotionen und sozialen Strukturen liegt bislang jedoch nicht vor.

Vor diesem Hintergrund besteht die Notwendigkeit, vorhandene soziologische Emotionstheorien mit Blick auf ihre Erklärungskraft für die Entstehung und Reproduktion sozialer Ordnung zu systematisieren sowie gemeinsame Schnittstellen zwischen den skizzierten Ansätzen zu identifizieren und zu einem eigenständigen Theorieentwurf zusammenzuführen. Um die angesprochenen Probleme und blinden Flecken bisheriger soziologischer Emotionstheorien zu vermeiden, sollen auch Emotionstheorien aus anderen Disziplinen in die Analyse mit einbezogen werden.

In diesem Kapitel wurde u.a. exemplarisch gezeigt, wie sozionische Modellierungstechnik und insbesondere Petrinetz-Formalismen zur Modellierung und Integration der interdisziplinären Theorien der Emotion verwendet werden können.

Bei den meisten behandelten soziologischen Theorien bildet Emotion nicht den zentralen Forschungsgegenstand, sondern wird im Rahmen des komplexen Erklärungsmodells eines bestimmten Aspekts der Sozialität mitbehandelt. Hierbei werden meist auch nur spezielle, in dem jeweiligen sozialen Zusammenhang relevante Facetten des Emotionsphänomens thematisiert. Da jedoch der Modellierungsfokus bei jeder zur Modellierung ausgewählten soziologischen Theorie auf einem speziellen Aspekt lag - nämlich auf dem Erklärungsansatz der Generierung und Funktionsweise der Emotion in natürlichen (sozialen) Systemen aus der jeweiligen theoretischen Perspektive -, bestand das Ziel nicht darin, die betrachteten Theorien in ihrer gesamten Komplexität zu modellieren, sondern vielmehr eben diese speziellen Aspekte der Wechselwirkungen von Emotionen und anderen sozialen Entitäten / Konzepten zu thematisieren und nachzubilden.

Aus Sicht der Wechselwirkungen von Emotion, Norm und normkonformen Verhalten kann zusammenfassend festgestellt werden, dass insbesondere Scham und Verachtung als Instrumente der Normetablierung dienen, indem sie einerseits das Verhalten hin zur Normkonformität steuern und andererseits die Vermeidung der negativen Konsequenzen nicht konformer Handlungen bedingen. Die Konformität zu einer sozialen Norm ist somit nicht notwendigerweise durch die Vermeidung materieller Sanktionen bedingt. Vielmehr ist sie das

6.3. Soziologische Erklärungsansätze zur Emotion

Resultat der automatischen internalisierten Prävention der emotional gesteuerten Reaktionen der anderen Gruppenmitglieder (durch Demonstration negativer sozialer Emotionen wie Verachtung, Abscheu, Missbilligung), die zu negativen Emotionen (Scham, Schuld, Verlegenheit) beim Normbrecher führen. Entsprechend führt emotionsgesteuerte Verstärkung (durch Anerkennung und Bewunderung von Seiten des Beobachters des normkonformen Verhaltens) zur Generierung von positiven Emotionen wie Zufriedenheit und Stolz. Somit unterstützt die unwillkürliche Antizipierung der positiven sozialen Emotionen die Internalisierung der sozialen Norm.

Anfänglich gingen wir von der Annahme aus, dass Normen das Handeln der Akteure restringieren, indem sie bestimmte Handlungsoptionen und -alternativen als inadäquat kennzeichnen. Mit Hilfe der Ergebnisse aus Damasio (2003) konnten wir zeigen, dass dank der emotionsbasierten Situationskategorisierung die mit negativen Emotionen assoziierten Handlungen unter bestimmten situationalen Gegebenheiten nicht ausgeführt werden, während mit positiven Emotionen verknüpfte Handlungsoptionen von den Akteuren vorgezogen werden. So kommen Regelmäßigkeiten im Handeln der Akteure zustande, die wir als soziale Normen bezeichnet haben. Mit der Theorie von Elias konnten wir erläutern, wie diese unter Mitwirkung der sozialen Emotionen auf der Mikroebene fortdauernd (re-)internalisiert werden.

6. Psychologische, soziologische und neurowissenschaftliche Emotionstheorien

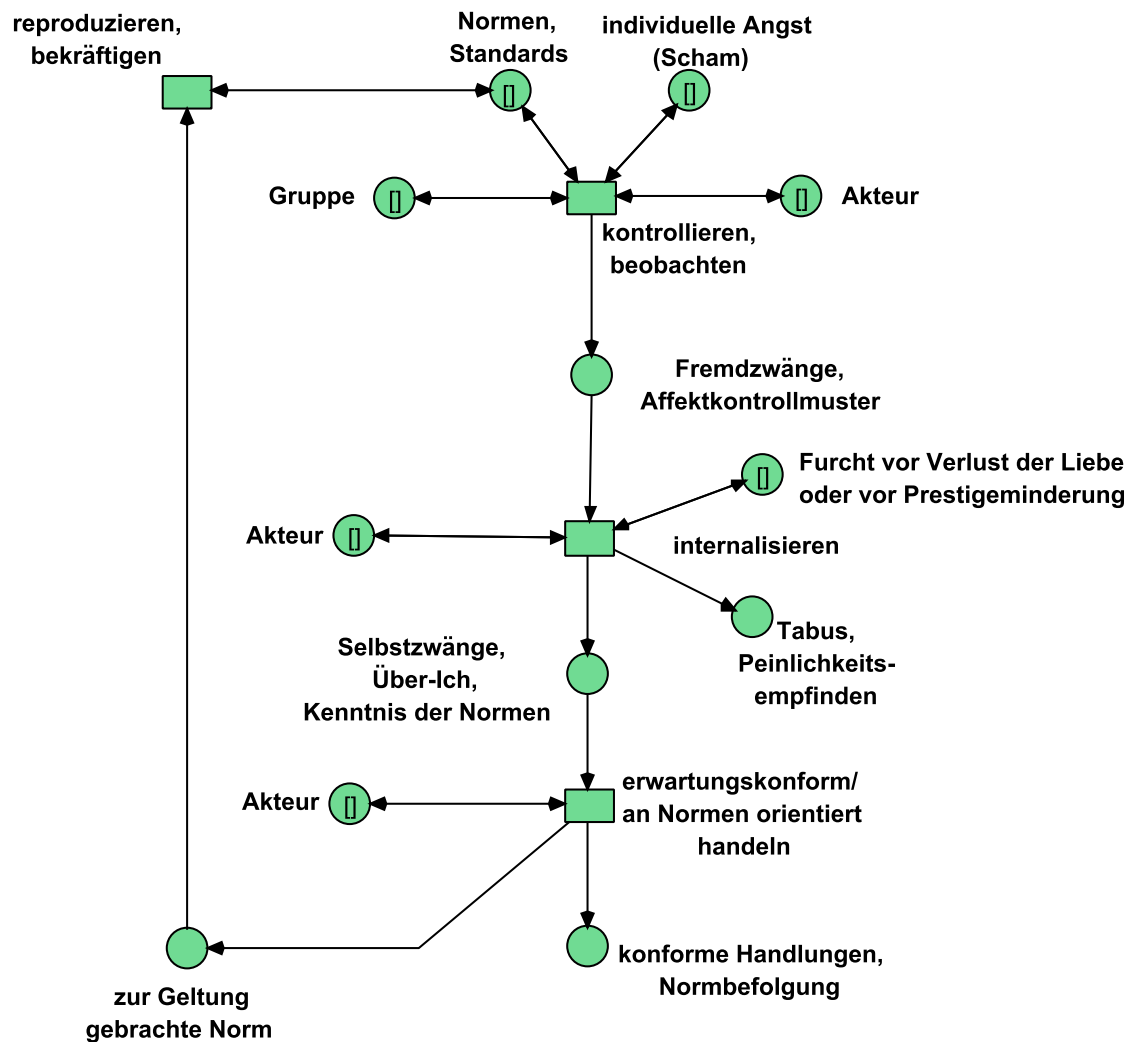


Abbildung 6.35.: Reproduktionskreislauf: Internalisierung von Fremdwängen (aus [vLMV03], S. 66).

7. Architekturen emotionaler Agenten

Nachdem im vorangegangenen Kapitel die emotionstheoretischen Grundlagen vorgesehlt wurden, können nun die aktuellen Emotionsmodelle der Informatik analysiert und u.a. ihre theoretische Fundierung beurteilt werden. In Kapitel 2 wurden bereits die Forschungsrichtungen innerhalb der informatischen Emotionsforschung und in Kapitel 3 einige wichtige KI- und VKI-Modellierungsansätze beschrieben und grundlegende Agentenarchitekturen vorgestellt.

In diesem Kapitel nehmen wir eine Reihe von Agentenarchitekturen, die die Entwicklung auf dem Gebiet der informatischen Emotionsforschung maßgeblich beeinflusst haben, in den Blick. Hierbei soll die Architekturbetrachtung nicht dazu dienen, die jeweilige Agentenarchitektur detailliert vorzustellen, sondern vielmehr das zugrunde liegenden Emotionsmodelle zu analysieren und ihre emotionstheoretische Fundierung zu beurteilen.¹ Für die Analyse der Modelle wenden wir erneut den sozionischen Modellierungsansatz an (vgl. Kapitel 4.4), nun in Bezug auf die informatische Architekturmodelle. Nachdem wir die grundlegenden Komponenten der jeweiligen Agentenarchitektur kurz vorgestellt haben, wird ein Petrinetz-Modell der Architektur bzw. des zugrunde liegenden Emotionsmodells erstellt und in Beziehung zu den in Kapitel 6 formulierten theoretischen Grundlagen gesetzt.

Für die nachfolgende Betrachtung der Ansätze zur Emotionsmodellierung werden wir in erster Linie zwischen Ansätzen unterscheiden, die Emotionen als Mittel zur Performanz- und Effizienzsteigerung isoliert agierender Entitäten im Kontext der KI-Systeme implementieren (z.B. [Pet03], [Vel97], [Vel98], [Vel99], [Can03], [Slo01]), und solchen, die die Bedeutung der Emotion für sozial interagierende Agenten in Multiagentensystemen, verteilten Systemen und künstlichen Gesellschaften erschließen wollen (z.B. [AS96], [Aub98], [Aub05], [Aub09], [Ell92], [EO92], [RB92], [BLR92b]).

Die KI- und die VKI-Ansätze werden gesondert betrachtet, da für die Schwerpunktsetzung bei der Modellbildung entscheidend ist, ob die Emotionen der isolierten Entitäten (autonome Agenten) oder der sozial situierten Entitäten (Agenten in Agentengemeinschaften) modelliert werden müssen. Insbesondere kann man von den VKI-Modellen eine stärkere Berücksichtigung der sozialen Faktoren erwarten, während bei den KI-Modellen offensichtlich andere funktionale Gesichtspunkte im Mittelpunkt der Modellierung stehen.

Zuerst stellen wir verschiedene Klassifizierungsmöglichkeiten für Emotionsarchitekturen vor. Anschließend werden die KI-Modelle dargestellt, und zwar die Ansätze von Staller / Petta, Velásquez und Dörner / Bach. Allen diesen Ansätzen gemeinsam ist die Kontrollfunktion der Emotionen innerhalb eines Gesamtsystems, das sich an eine dynamische unsichere Umwelt anpassen muss. Die Emotion ist in der Architektur des Systems ver-

¹Für die Spezifikation und Diskussion der Begriffe “Agentenarchitektur“, “Emotionsarchitektur“ und “Emotionsmodell“ siehe Kapitel 1.5.

7. Architekturen emotionaler Agenten

ankert und entsteht aus dem Zusammenspiel der verschiedenen Architekturkomponenten. Als Quintessenz dieser Auffassung wird im Anschluss die Emotionsarchitektur von Aaron Sloman vorgestellt, die Emotion in den Kontext eines global angelegten Rahmenwerkes zur Beschreibung natürlicher und künstlicher intelligenter Systeme stellt.

Im nächsten Teilabschnitt wird auf die VKI-Emotionsmodelle von Elliot, Bates / Reilly und Aubé / Senteni eingegangen, die Motivationen und Vorgehensweisen illustrieren, wie Agenten in einem Multiagentensystem mit dem Faktor Emotionalität ausgestattet werden können.

Das Kapitel schließt mit einer Zusammenfassung und Auswertung der betrachteten Ansätze im Hinblick auf ihre psychologische, sozialwissenschaftliche bzw. neurowissenschaftliche Fundierung, die verwendete Terminologie, Validität etc.

7.1. Klassifizierung der Emotionsarchitekturen

Bereits Ende der 1980er Jahre erschienen Übersichtsartikel über existierende KI-Emotionsmodelle (vgl. [Dye87], [PN85], [Pfe88]). Heute gibt es eine Vielzahl von informatischen Emotionsmodellen, sowohl in Form von Architekturvorschlägen als auch in Form von implementierten Systemprototypen. Es wurden mehrere Klassifizierungsmöglichkeiten für Emotionsarchitekturen nach verschiedenen Gesichtspunkten vorgeschlagen, die sich z.B. aus der Modellstruktur, der Herangehensweise bei der Modellierung oder auch aus der allgemeinen Motivation für die Modellierung der Emotion in einem (Agenten-)System herleiten.

In [Pfe88] wurde eine Klassifizierung der Ansätze zur Modellierung der Emotion in zwei Kategorien vorgeschlagen: Die erste Kategorie bilden *Augmented Cognitive Models*, die nicht die Emotion in den Mittelpunkt stellen, sondern sich vorwiegend mit der Modellierung kognitiver Prozesse befassen, wobei Emotion lediglich als Zusatzfaktor eine ergänzende Rolle spielt (z.B. in der Architektur von Sloman [SC81]). Die zweite Kategorie der *AI Models of Emotion* beinhaltet Modelle, die sich die Emotionsmodellierung als Hauptziel setzen (z.B. die Modelle von Frijda [MFP93], [FS87]).

In einer weiteren Arbeit [PN85] hat Pfeifer zusätzlich zwischen Modellen unterschieden, die auf spezifischen Taxonomien von Emotionen beruhen und sich zum Ziel setzen, diese zu operationalisieren, um unterschiedliche Emotionskategorien zu erzeugen (z.B. BORIS [Dye87] oder Affective Reasoner [Ell92]). Solche Modelle nennt Pfeifer *Reasoners* und stellt sie den *psychologischen Modellen* gegenüber, deren Hauptinteresse darin besteht, psychologische Emotionstheorien im Sinne eines sozialwissenschaftlichen und psychologischen Erkenntnisgewinns zu modellieren [PN85]. Dazu kann zum Beispiel das Herausarbeiten einer Emotionstaxonomie, die Ermittlung von Evaluationskriterien zur Emotionsidentifikation [Sch93] oder die Überprüfung einer existierenden Emotionstheorie anhand eines Computermodells (z.B. [FS87], [DLD⁺01]) gehören.

Die psychologischen Modelle unterteilt Pfeifer [Pfe88] weiter in solche, die Emotionen allein wegen Unzulänglichkeiten einer rein kognitiven Lösung in die Modellierung einbeziehen (z.B. PARRY [Col75a]), und solche, deren Ziel in der Erfassung der emotionalen Prozesse an

sich liegt (z.B. Pfeifers FEELER [Pfe82], Frijda und Swagermans ACRES [FS87], Dörners PSI [DLD⁺01]).

In [Weh98] werden zwei Klassen der informatischen Emotionsmodelle aufgrund ihrer Zielsetzungen und Herangehensweisen bei der Modellierung unterschieden: *Black-Box-Modelle* und *prozessbasierte Modelle*.²

- Bei der *Black-Box-Modellierung* wird versucht, die Funktionalität der natürlichen intelligenten Systeme nachzubilden, ohne auf die Details der zugrunde liegenden natürlichen Mechanismen einzugehen. Solche Modelle versuchen unter Annahme der Existenz eines emotionalen Zustandes, seine begleitenden Phänomene wie physiologische und kognitive Ursachen und vor allem die Auswirkungen der Emotion auf die Deliberationsprozesse Entscheiden, Verhalten usw. nachzubilden, ohne auf die zugrunde liegenden natürlichen Mechanismen (z.B. neurobiologische und metabolische Prozesse, Wechselwirkungen zwischen verschiedenen Subsystemen usw.) einzugehen. Für diese Modellierungsart ist vor allem die Input / Output-Relation von Interesse, deshalb ist dieser Ansatz vor allem für rein entwicklungstechnisch motivierte Modelle geeignet, die in erster Linie funktionale Systemoptimierung oder Effizienzgewinn bezwecken. Black-Box-Modellierung bringt keine emotionstheoretischen Erkenntnisse über die natürlichen Mechanismen, die die Funktionalität der Emotion in natürlichen Systemen hervorbringen. Zu dieser Modellierungsart könnte man die weiter unten vorgestellten Ansätze von Staller / Petta [SP98], Aubé / Senteni [AS96], Elliott [Ell92] und Bates / Reilly [BLR92a] zählen.
- Umgekehrt zielt die *Prozessmodellierung* darauf ab, die natürlichen kognitiven Prozesse (z.B. Entscheidungsprozesse) durch Erforschung der zugrunde liegenden u.a. emotionalen Mechanismen zu verstehen und zu simulieren. Diese Modellierungsart bietet sich insbesondere an, wenn weder die Umgebungseigenschaften noch die Art der Agenten bzw. der Systeminteraktion mit dieser Umgebung zum Zeitpunkt der Modellierung bekannt sind bzw. wenn eines der Modellierungsziele darin besteht, die Wechselwirkungen zwischen Umgebung, Agenten und Emotionen zu erforschen. Ein typisches Beispiel für Prozessmodellierung ist die architekturbasierte Herangehensweise von Aaron Sloman (z.B. [Slo01]). Diese Prozessmodellierung ist wesentlich schwieriger umzusetzen, jedoch ermöglicht sie dank ihrer explorativen Natur tiefere Einsichten in den Emotionsprozess an sich und ist besonders geeignet für die Entwicklung der (agentenbasierten) Testumgebungen zur Überprüfung der Emotionstheorien. Es existieren nur einige wenige Architekturen, die Emotionalität auf dem Niveau der biologischen Mechanismen nachzubilden versuchen (z.B. Armony [ASSCL95], Velásquez [Vel97], Canamero [Can03]).

Frijda und Moffat unterscheiden ferner *imitierende* und *funktionale* Emotionsmodelle ([MF00], S. 59). Unter imitierenden Modellen verstehen sie Agentenarchitekturen aus

²Eine sinngemäß gleiche Aufteilung, lediglich mit abweichenden Klassenbezeichnungen, findet man auch bei [Slo93] (vgl. auch [Can03], S. 124). Sloman trennt die Modelle in *phänomenbasierte* (entspricht der Black-Box-Modellierung) und *designbasierte* (entspricht der prozessbasierten Modellierung).

7. Architekturen emotionaler Agenten

dem MCI-Bereich, die z.B. zur Modellierung der Benutzeremotionen oder zur emotionalen Ausstattung der Benutzerinteraktion eingesetzt werden, sodass der Interaktionsagent anscheinend emotionale Reaktionen zeigt. Funktionale Emotionsmodelle dagegen haben den Zweck, entweder die Untersuchung und Weiterentwicklung von psychologischen und kognitiven Emotionstheorien durch Demonstration der adaptiven Funktion von Emotionen voranzutreiben oder durch Einsatz der Emotionen bessere KI-Systeme zu konstruieren, wobei die Nachbildung von natürlichen adaptiven Emotionsfunktionen im System eine zentrale Rolle spielt.

Dörner unterscheidet anhand der Realisierung der Emotionsmodelle die *regelbasierten Produktionssysteme* und *neuronale Netze* [DLD⁺01]. Unter Produktionssystemarchitekturen werden Modelle zusammengefasst, die semantische Regeln für symbolische Informationen formulieren. Unter dem Begriff “neuronale Netze” werden Modelle verstanden, die Informationen auf subsymbolischer Ebene durch Aktivitätsausbreitung verarbeiten. Als regelbasierte Produktionssysteme können die im Folgenden vorgestellten Modelle von Bates und Reilly [BLR92a] sowie Elliott [Ell92] betrachtet werden. Ein psychologisches Modell von Scherer [Sch93] arbeitet nicht regelbasiert, sondern benutzt einen mehrdimensionalen Vektorraum, um einen Deliberationsmechanismus zu realisieren. Auch in den Modellen von Dörner [DSD01] und Bach [Bac03] wird eine spezielle Art neuronaler Netze als Grundelement der Informationsverarbeitung vorgeschlagen. Alle Arten von Informationsverarbeitungs-, Speicherungs- und Gedächtnisprozessen werden neuronal realisiert.

Gratch und Marsella trennen schließlich zwischen *kommunikationsbasierten* (“communication-driven”) und *simulationsbasierten* (“simulation-driven”) Emotionsmodellen, wobei die Ersteren auf der Emotionsmanifestation und dem Potenzial der Emotion für die Optimierung der MCI fokussieren, während die Letzten der Emotion zugrunde liegende kognitive Mechanismen modellieren und das Potenzial der Emotion für die Steuerung der kognitiven Prozesse untersuchen [GM05].

In [MGP12] findet sich eine Übersicht der aus der Sicht der Autoren bis dato einflussreichsten Emotionsmodelle, die in Abb. 7.1 wiedergegeben wird.

7.2. KI

In diesem Abschnitt werden Ansätze vorgestellt, die sich mit der Emotionsmodellierung vor dem Hintergrund der KI-Forschung befassen. Die Auswahl der dargestellten Modelle zielt auf die Widerspiegelung der Forschungsansätze, die die Emotionsforschung im KI-Bereich besonders beeinflusst haben, aber auch solcher Arbeiten, die viel Polemik in die wissenschaftliche Emotionsdiskussion gebracht haben.

Zuerst wird das Emotionsmodell von Staller und Petta dargestellt. Es folgen die Ansätze von Velásquez und Dörner / Bach. Am Ende des Kapitels werden die Modelle von Sloman vorgestellt und analysiert.

Den KI-Ansätzen gemeinsam ist die Annahme einer Steuerungsfunktion der Emotionen innerhalb des Gesamtsystems. Im Unterschied zu den in Kapitel 2.2 behandelten Emotionsmodellen liegt der Schwerpunkt nicht darauf, den emotionalen Zustand eines Agenten zu *repräsentieren*, sondern vielmehr darauf “arriving at an understanding of emotion that

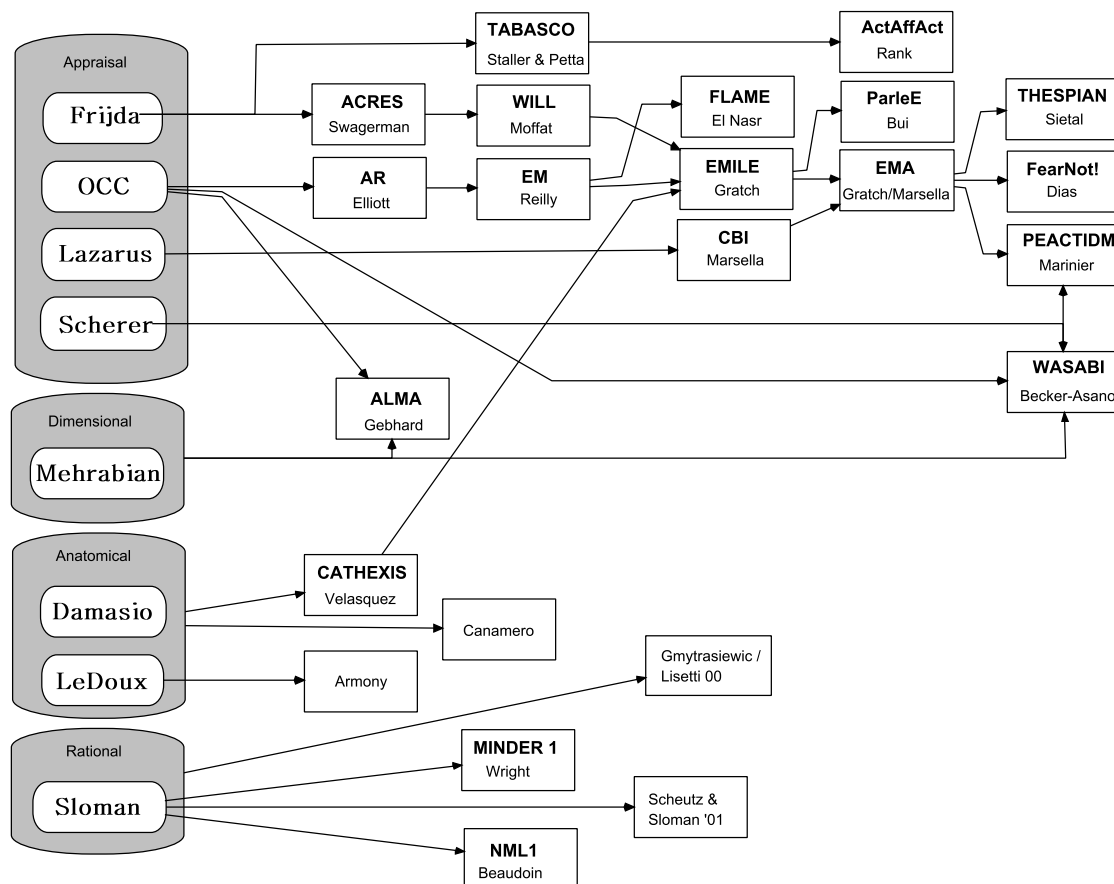


Abbildung 7.1.: Eine Übersicht über die historische Entwicklung der Emotionsmodelle (nach [MGP12]).

allows the software agents to undergo emotional states“ ([Bac02], S. 2). Die Herausforderung besteht also darin, den Agenten einen emotionalen Zustand *einnehmen* zu lassen. Dazu ist die theoretische Ausarbeitung und praktische Implementierung der Funktion der Emotion innerhalb der Agentenarchitektur erforderlich.

Ferner ist den Modellen die Zielsetzung gemeinsam, die Emotionalität auf dem Niveau der biologischen Mechanismen nachzubilden. Die Emotionen sind in der Architektur des Systems verankert und entstehen aus dem Zusammenspiel der verschiedenen Architekturkomponenten, wobei eine explizite Definition bzw. Repräsentation des emotionalen Zustandes nicht unbedingt Bestandteil der Modellierung ist (vgl. die Modelle von Dörner ([DLD⁺01], [DSD01]) bzw. Bach ([Bac02], [Bac03])).

7.2.1. Staller und Petta: Die TABASCO-Architektur

Die TABASCO-Architektur (“Tractable Appraisal-Based Architecture for Situated Cognizers“) von Staller und Petta (1998) wurde als Rahmenwerk für situierte Softwareagenten

7. Architekturen emotionaler Agenten

in virtuellen Umgebungen konzipiert. Eine grundlegende Zielsetzung bei der Architekturentwicklung bestand in der Integration der Emotion in die Architektur eines situierten Softwareagenten. Die Autoren streben nicht die Abbildung verschiedener emotionaler Zustände im System an, sondern vielmehr die Nachbildung der Funktionalität der Emotion bei der Interaktion des Agenten mit seiner Umgebung, insbesondere bei der Steuerung der Aktionsgenerierungs- und Planungsprozesse: “We model emotions not as reified entities, but as adaptive processes related to the interaction of an agent with its environment.” ([SP98], S. 8).

Das Architekturmodell stützt sich im Wesentlichen auf die Emotionstheorien von Scherer [Sch88], Lazarus [Laz91] und Frijda [Fri86] (siehe Kapitel 6 für die Beschreibung und Modellierung der entsprechenden Theorien).

In der TABASCO-Architektur werden alle Schritte des Emotionsprozesses nach [Fri86] modelliert (vgl. Abb. 5.1 bzw. 5.2), nämlich Emotionseinschätzung, Generierung der Aktionstendenzen ([Fri86]) und Bewältigungsaktivitäten (*coping*) ([Laz91], [Sch88]).

Der Einschätzungsprozess wird in der TABASCO-Architektur auf der Grundlage des Einschätzungsmodells von Leventhal und Scherer [LS87] aufgebaut, das Perzeption als Teil der Einschätzung betrachtet. In [LS87] wird ein hierarchisches Verarbeitungssystem mit drei Ebenen - sensomotorisch, schematisch und konzeptionell - definiert: “The sensory-motor level is based on innate hard-wired feature detectors which can give rise to emotional reaction directly. The schematic level is based on schema matching. The conceptual level involves reasoning and inference processes that are abstract, active, and reflective.” ([SP00], S. 2).

Eine grundlegende Erweiterung dieser emotionstheoretischen Grundlage in TABASCO besteht darin, dass die drei in [LS87] postulierten Verarbeitungsebenen nicht nur in Bezug auf Einschätzungen Verwendung finden, sondern auch auf die Aktionsgenerierung übertragen werden.³ Die beiden Hauptkomponenten der Architektur, *Perception and Appraisal* und *Action* (siehe Abb. 7.2), sind als Hierarchie mit drei Ebenen angelegt.

Im Folgenden werden die einzelnen Komponenten der TABASCO-Architektur etwas detaillierter dargestellt.

Perception and Appraisal Diese Architekturkomponente verarbeitet die Umgebungsstimuli und modelliert den Einschätzungsprozess basierend auf den Komponenten der Einschätzungstheorien von Scherer und Frijda (vgl. Kapitel 6.1.2). Die Einschätzungsdimensionen entsprechen im Wesentlichen denen von Scherer, jedoch greifen die Autoren die Argumentation aus [FM93] auf, dass es nicht nötig sei, alle von Scherer definierten SECs (*stimulus evaluation checks*) explizit zu modellieren. Vielmehr kann ein großer Teil der Einschätzungen als impliziter Bestandteil des Perzeptionsprozesses angesehen werden. In [Pet03] wird z.B. behauptet, es sei nicht notwendig den *novelty check* explizit zu modellieren, da bereits im Rahmen des Perzeptionsprozesses bewertet wird, ob ein Stimulus mit einem der bestehenden Schemata übereinstimmt und entsprechend als neu oder “bekannt” einzuschätzen sei.

Entgegen der Konzeption von [LS87] hält Petta die Evaluation des Inputs auf jeder Verarbeitungsebene entlang aller Einschätzungsdimensionen für nicht notwendig. Er argumentiert,

³wie von Leventhal in seiner “perceptual-motor theory of emotion“ vorgeschlagen (Leventhal 1984).

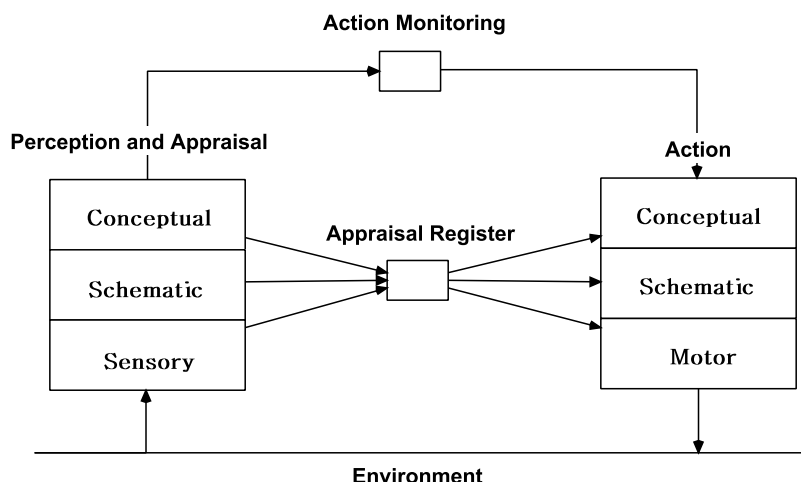


Abbildung 7.2.: TABASCO-Architektur von Staller / Petta (nach [SP98], S. 9).

dass einige Einschätzungskriterien auf bestimmten Ebenen nicht relevant sind: So sei z.B. der *compatibility check*, der das Ereignis mit inneren und äußeren Standards und Normen vergleicht, auf der sensorischen Verarbeitungsebene nach Petta nicht relevant ([Pet03]).

Die Prozesse auf den Verarbeitungsebenen der *Perception and Appraisal*-Komponente können wie folgt zusammengefasst werden (nach [SP98] S. 9, [Pet03] S. 251-287):

- Auf der sensorischen Ebene befinden sich die Eigenschaftsdetektoren (feature detectors), z.B. für die Registrierung von plötzlichen intensiven Stimuli (vgl. Scherers *novelty check*) bzw. der Qualität eines Stimulus (*intrinsic pleasantness check*).
- Auf der schematischen Verarbeitungsebene wird der Input mit den Schemata, insbesondere den sozialen und selbstdarstellenden Schemata verglichen. Petta schlägt die Implementierung der schematischen Ebene in Form eines assoziativen Netzwerks mit sich verbreitender Aktivierung nach der Anregung durch von [SP81] vor.
- Auf der konzeptuellen Ebene sind abstraktes Denken und Inferenzmechanismus auf der Grundlage des Wissens und der Überzeugungen (Beliefs) des Agenten angesiedelt. Die Schlussfolgerungsregeln für die Ursachenzuschreibung sowie die Auswertung der eigenen Aktionen des Agenten in Bezug auf seine internen Normen und Selbstbild (*norm / self compatibility check*) werden hier modelliert. Deshalb sind Modelle von dem Selbst und von anderen für diese Verarbeitungsebene relevant (vgl. Elliotts “Concerns of Others“ (COOs) in der AR-Architektur (siehe Kapitel 7.3.1) bzw. Fergusons Anwendung der BDI-Architektur auf dem “Modelling Layer“ der TouringMachines-Architektur [Fer92]).

Action. Die *Action*-Komponente hat ebenso wie die *Perception and Appraisal*-Komponente drei Verarbeitungsebenen:

7. Architekturen emotionaler Agenten

- Die motorische Ebene enthält motorische Kommandos und ist für die Generierung von automatischen Reaktionen zuständig. Eine wichtige Klasse von motorischen Mechanismen bilden expressive Verhaltensreaktionen wie Gesichts- oder Körperausdruck. Solche expressiven Verhaltensweisen werden auf der motorischen Ebene von *Action*-Komponente generiert.
- Auf der schematischen Ebene werden Aktionstendenzen in Form von “flexible programs” ([Fri86], S. 83) generiert, d.h. verschiedene Handlungsmöglichkeiten, bei denen Abweichungen in den Situationsumständen und Feedbacks von den bereits ausgeführten (äquivalenten) Handlungen berücksichtigt werden können.
- Auf der konzeptionellen Ebene werden (langfristige) Planungsprozesse modelliert. Während die Aktionstendenzen angeborene biologische Impulse sind, ist “coping” nach der Definition von [Laz91] ein komplexer und häufig planmäßiger psychologischer Prozess, der stark auf Einschätzungen ausgerichtet ist wie “Was ist möglich?” oder “Was ist effektiv?” in einer bestimmten Situation und “Was ist kompatibel mit sozialen und persönlichen Verhaltensstandards?”. Die konzeptionelle Ebene besteht deshalb in erster Linie aus einem Planer und ist für die Implementierung des “coping” zuständig.

Appraisal Register Zwischen der *Action*-Komponente und der *Perception and Appraisal*-Komponente moderiert das *Appraisal Register*, das auf einen Vorschlag von Smith et al. (1996) zurückgeht ([Pet03], S. 268). Es registriert und kombiniert die Einschätzungsergebnisse der drei Ebenen der *Perception and Appraisal*-Komponente und steuert auf der Grundlage des eingeschätzten Zustands die *Action*-Komponente. Der Einfluss der *Perception and Appraisal*-Komponente auf die *Action*-Komponente kann verschiedene Formen annehmen. Zum einen kann dies etwa bei Angst- oder Orientierungsreaktionen zu einer unmittelbaren Ausführung eines motorischen Reaktionsprogramms führen. Zum anderen kann eine Aktionstendenz generiert werden, wobei das *RAP-System* in den Zustand der Handlungsbereitschaft überführt wird.⁴ Schließlich kann ein Planungsprozess auf der konzeptionellen Ebene initiiert werden, welcher die Anpassung an das betreffende einzuschätzende Ereignis ableiten soll.

Action Monitoring Die *Action Monitoring*-Komponente überwacht die Planungs- und Ausführungsprozesse der *Action*-Komponente und übermittelt die Ergebnisse an die *Perception and Appraisal*-Komponente, wo sie in den Einschätzungsprozess integriert werden. Diese Komponente soll die von Frijda spezifizierten Regulierungsprozesse implementieren. Staller und Petta zeigen in [SP00] prinzipielle Möglichkeiten der Integration auf: So könnte z.B. die Impulssteuerung modelliert werden, wenn dem Planungsprozess auf der konzeptionellen Ebene der Zugriff auf die Prozesse der schematischen Ebene gewährt wird, so dass die Ausführung bestimmter Aktionstendenzen verhindert wird. Außerdem besteht die Möglichkeit, die Ergebnisse der *Action Monitoring*-Komponente, die an die *Perception and Appraisal*-Komponente weitergeleitet werden, eine erneute Einschätzung der Situation

⁴Das RAP-System (“reactive action package”) von Firby [Fir89] wird in TABASCO zur Implementierung der Aktionstendenzen auf der schematischen Ebene der *Action*-Komponente eingesetzt.

auslösen zu lassen. Z.B. kann auf diese Weise eine bestimmte Einschätzungsdimensionen neu ausgewertet werden, wenn die Ausführung einer zuvor durch die *Action*-Komponente berechneten Aktion erfolglos blieb.

Bisherige Implementierungen TABASCO wurde als Steuerungsarchitektur für die Steuerung eines autonomen Interaktionsagenten mit der Bezeichnung *Invisible Person* in einer interaktiven virtuellen Umgebung implementiert und erprobt. Bei der Implementierung wurde die *Action*-Komponente von TABASCO durch die 3T-Architektur von Bonasso et al. [BFG⁺97] realisiert. Die drei Schichten der 3T-Architektur (*“reactive skills“*, *“sequencing“* und *“deliberation“*) wurden zu den TABASCO-Verarbeitungsebenen in Beziehung gesetzt ([Pet03], S. 261).

Die unterste Schicht der 3T-Architektur (*“reactive skills“*) korrespondiert mit der sensorischen Verarbeitung in TABASCO. Die mittlere Schicht der 3T (*“sequencing“*) basiert auf dem RAP-System (*“reactive action package“*) von Firby [Fir89] und korrespondiert zu der schematischen Verarbeitung: Hier werden Aktionstendenzen generiert und gesteuert. Die von Frijda im Zusammenhang mit Aktionstendenzen definierte *“flexible programs“* [Fri86] sind alternativen Abfolgen von Aktionen, die folgendermaßen definiert werden: *“Flexible programs are those that are composed of alternative courses of action, that allow for variations in circumstances and for feedback from action executed“* ([Fri86], S. 83). Die Funktionalität von *“flexible programs“* kann mit RAP auf einfacher Weise umgesetzt werden, denn RAP *“is simply a description of how to accomplish a task in the world under a variety of circumstances using discrete steps“* ([BFG⁺97], S. 240). Die Berücksichtigung von Feedback-Informationen über die Ergebnisse von bereits ausgeführten Aktionen wird ebenfalls unterstützt. Die konzeptuelle bzw. deliberative Ebene der 3T-Architektur besteht im Wesentlichen aus einem Planer [ES94], in Übereinstimmung mit Lazarus' Charakterisierung des *“coping“* als Planungsprozess [Laz91].

Petrinetz-Modell der TABASCO-Architektur Abb. 7.3 zeigt die bisher beschriebenen Komponenten der Tabasco-Architektur als ein Petrinetz.

Der Verarbeitungsprozess wird durch einen Impuls aus der Umwelt initiiert (Transition input) und läuft auf allen Verarbeitungsebenen der *Perception and Appraisal*-Komponente parallel ab.

Auf der sensorischen Ebene findet die automatische Registrierung bestimmter Eigenschaften des Stimulus durch Eigenschaftsdetektoren statt (Transition **feature detection**). Ggf. wird bereits an dieser Stelle eine Emotion generiert und löst eine Aktion aus (Transition **action generation** auf der motorischen Ebene der *Action*-Komponente wird aktiviert). Dies ist bei bestimmten automatischen Reaktionen der Fall. Die Autoren erläutern, dass in vielen Fällen darauf verzichtet werden kann, auf die Einschätzungsergebnisse *aller* Ebenen der *Appraisal*-Komponente zu warten. Ein Reaktion auf Schmerz z.B. erfolgt sofort, nachdem der Input auf der sensorischen Ebene bearbeitet wurde, und bedarf keiner Verarbeitung auf der konzeptionellen Ebene. Entsprechend sind die Einschätzungsergebnisse von höheren Verarbeitungsebenen der *Appraisal*-Komponente für die Generierung der Aktionen

7. Architekturen emotionaler Agenten

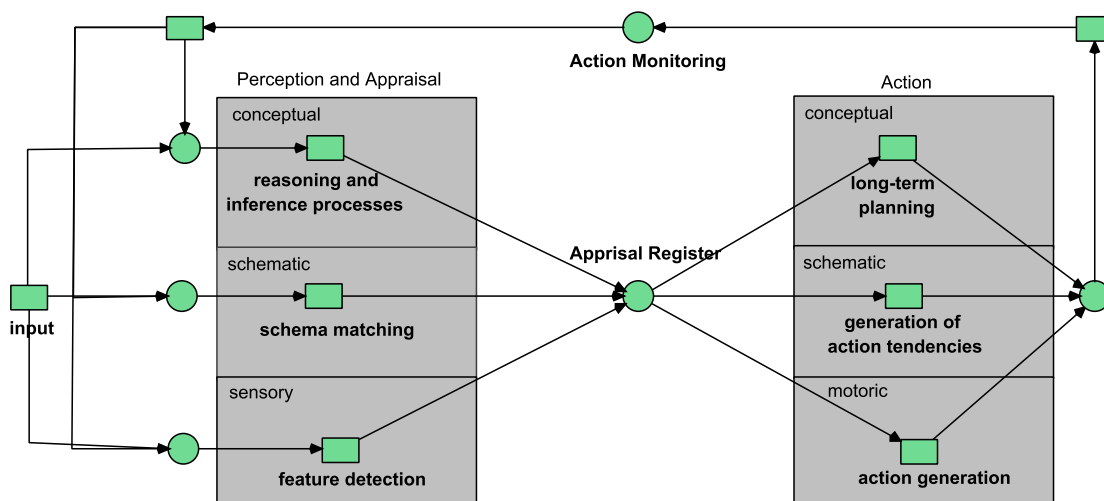


Abbildung 7.3.: TABASCO-Architektur von Staller / Petta als Petrinetz.

in diesem Fall nicht relevant; vielmehr erfolgt die Aktionsgenerierung allein aufgrund der Input-Verarbeitung auf der sensorischen Ebene.

Auf der schematischen Verarbeitungsebene wird Input mit den vorgegebenen Schemata, insbesondere den sozialen und selbstdarstellenden Schemata, verglichen (Transition **schema matching**). Auf der konzeptuellen Ebene erfolgt die Auswertung des Inputs in Bezug zur Kompatibilität mit den internen Normen und dem Selbstbild des Agenten. Hier sind komplexe Inferenzprozesse angesiedelt (Transition **reasoning and inference processes**).

Entsprechend der grundlegenden Designentscheidung von Staller und Petta [SP98] läuft der Emotionsprozess in TABASCO nicht sequenziell ab. Vielmehr sollen Agenten bereits aktiv auf Umweltzustände reagieren können, bevor die Verarbeitungsprozesse auf allen Ebenen der *Perception and Appraisal*-Komponente vollständig abgeschlossen sind. Die Autoren distanzieren sich von der Modellierung eines strikt sequentiellen Emotionsprozesses im Sinne von Frijda (vgl. Kapitel 6.1.2), um eine effizientere Verarbeitung zu ermöglichen und Reaktionszeiten zu reduzieren: “Our layered architecture has the advantage that the agent can respond reactively to events in the environment without having to execute a sequential process with action generation as the last step“ ([SP00], S. 5). Daher wird die Stelle **Appraisal Register** auch beim Schalten nur einer der drei Transitionen **feature detection**, **schema matching** oder **reasoning and inference processes** markiert. Die Funktion von Appraisal Register besteht in der Registrierung und Fusionierung der Einschätzungsergebnisse und der Initiierung der Verarbeitungsprozesse der *Action*-Komponente: Planungsprozesse (Transition **long-term planning**), Aktionstendenzen (Transition **generation of action tendencies**) oder Aktionen (Transition **action generation**).

In [SP98] diskutieren die Autoren über Notwendigkeit der Appraisal Register und weisen auf die Möglichkeit hin, die Verarbeitungsebenen der *Perception and Appraisal*-Komponente mit entsprechenden Ebenen der *Action*-Komponente direkt zu verbinden: “It may be possible to design a three-level architecture in which at each level of processing the appraisal process

and the influence on action are directly coupled“ [SP98], S. 11). Diese Möglichkeit wurde in einem Petrinetz aus Abb. 7.4 dargestellt. Dennoch entschieden sich die Autoren für die Beibehaltung der Appraisal Register, da sich dadurch die Koordination und Kontrolle des Zwischenspiels von “coping“, Aktionstendenzen und expressiven Verhaltensweisen wesentlich flexibler und einfacher gestaltet ([SP98], S. 11).

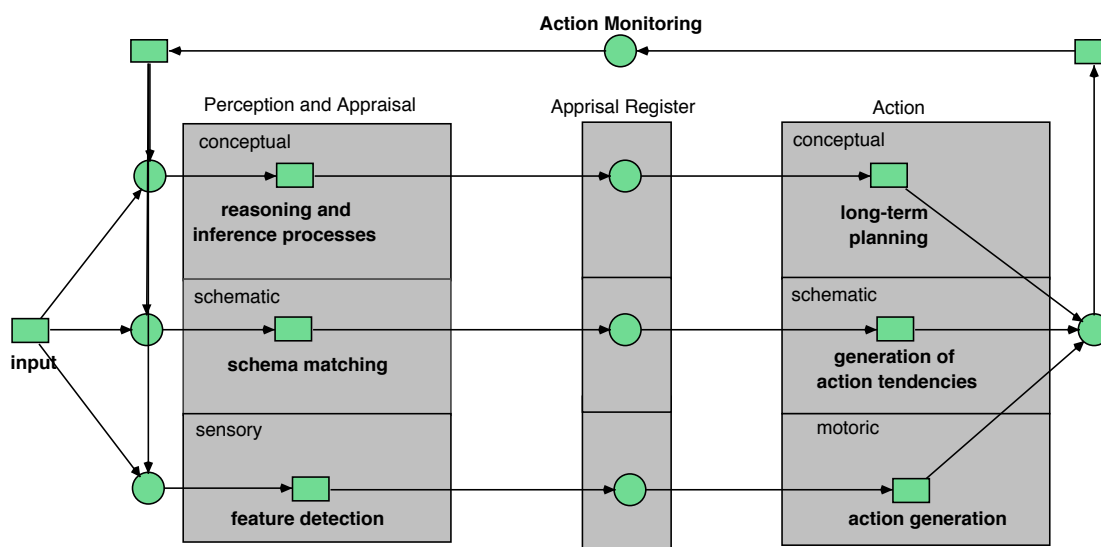


Abbildung 7.4.: Direkte Verbindung der Architekturebenen in TABASCO.

Die *Action Monitoring*-Komponente wurde durch die gleichnamige Transition dargestellt, die die Verarbeitungsergebnisse der *Action*-Komponente an die *Perception and Appraisal*-Komponente übermittelt. Hier können sie in die Einschätzungsprozesse integriert werden.

Der Einschätzungsprozess in TABASCO wird spezifiziert als Evaluierung der Input entsprechend der Einschätzungsdimensionen der Theorie von Scherer ([SP98], S. 9). Diese Einschätzungsdimensionen werden aus van Reekum and Scherer [vRS97] adoptiert und beinhalten “the perception of a change in the environment that captures the subject’s attention (novelty and expectancy), the perceived pleasantness or unpleasantness of the stimulus or event (valence), the importance of the stimulus or event to one’s goals or concerns (relevance and goal conduciveness or motive consistency), the notion of who or what caused the event (agency or responsibility), the estimated ability to deal with the event and its consequences (perceived control, power or coping potential), and the evaluation of one’s own actions in relation to moral standards or social norms (legitimacy), and one’s self-ideal“ ([vRS97], S. 259–260).

Nach dieser Darstellung entsprechen die TABASCO-Einschätzungsdimensionen den von Scherer spezifizierten SECs, die in Abb. 6.15 als ein Netzmodell dargestellt wurden. Nach Leventhal and Scherer [LS87] sollte ein Stimulus auf jeder Einschätzungsebene entlang aller Einschätzungsdimensionen ausgewertet werden. Da jedoch Leventhal and Scherer keine empirische Evidenz für diese Behauptung liefern, entscheiden sich Staller und Petta

7. Architekturen emotionaler Agenten

für die Verteilung der SECs auf verschiedene Einschätzungsebenen, wobei die Zuordnung der Einschätzungsdimensionen zu den Ebenen der *Perception and Appraisal*-Komponente nicht genau spezifiziert wird ([SP98], S. 9). Aus der Beschreibung der Funktion und Wirkungsweisen einzelner Verarbeitungsebenen in [SP98] lässt sich die Verteilung der SECs auf die Einschätzungsebenen ableiten (Abb. 7.5).

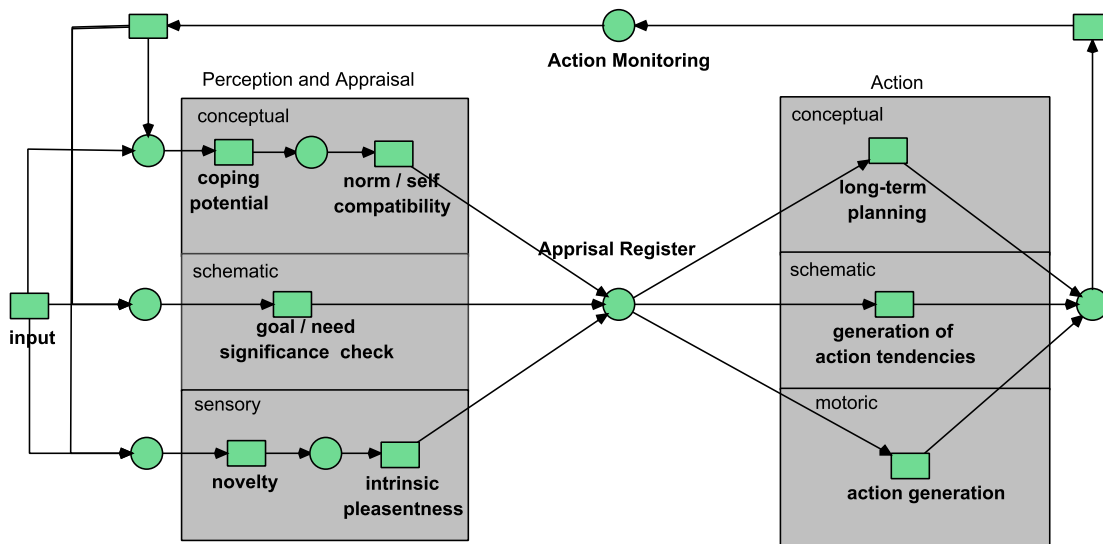


Abbildung 7.5.: Stimulus Evaluation Checks in TABASCO.

7.2.2. Velásquez: Emotion bei der Entscheidungsfindung

Verwendet man die Klassifizierung von Frijda und Moffat ([MF00], vgl. Kapitel 7.1), so gehört der Modellierungsansatz von Velásquez zu den funktionalen Emotionsmodellen. Bei den funktionalen Emotionsmodellen werden natürliche adaptive Emotionsfunktionen in KI-Systemen zwecks Optimierung bestimmter Systemeigenschaften nachgebildet, wobei die *funktionale* Rolle der Emotionen innerhalb der Systemarchitektur einen zentralen Stellenwert einnimmt.

In den Modellen von Velásquez ([Vel97], [Vel98], [Vel99]) werden die Steuerungsfunktionen der Emotionen bei der Wahrnehmung, der Motivation, dem Verhalten und der Motorik eines autonomen Agenten nachgebildet. Der Autor schlägt ein Architektur-Framework für eine emotionsbasierte Steuerung von autonomen Agenten vor, die sowohl in virtuellen Agenten als auch in autonomen Robotern eingesetzt werden kann und bereits in mehreren Implementierungen evaluiert wurde.

Einen besonderen Schwerpunkt in dem Modellierungsansatz von Velásquez bildet die Modellierung des Einflusses von emotionalen Mechanismen auf das Entscheiden und die Aktionswahl: "A particular emphasis is placed on using some of the mechanisms of emotions as building blocks for the acquisition of emotional memories that serve as biasing signals

during the process of making decisions and selecting actions.“ ([Vel98], S. 1). Das von Velásquez vorgeschlagene *Modell des emotionsbasierten Entscheidens* (Model for Emotion-Based Decision-Making) bezieht seine Anregung aus neueren Erkenntnissen der neurobiologischen Forschung ([Dam94], [Ado99], [BDTD97]). In [Vel98] wird z.B. die emotionsbasierte Steuerung der Entscheidungsfindungs- und Inferenzmechanismen in einem autonomen Agenten modelliert, die vor allem auf den Arbeiten von Damasio zur Somatic-Marker-Hypothese (siehe Kapitel 6.2.1) basiert.

Das Modell der emotionsbasierten Steuerung Velásquez definiert ein Framework für die so genannte emotionsbasierte Steuerung (Emotion-Based Control, EBC): “control of autonomous agents that relies on, and arises from, emotional processing“ ([Vel98], S. 2). Das Framework wurde als Erweiterung des ersten Modells von Velásquez - des Cathexis-Modells - entwickelt [Vel97] und für seine späteren, auf emotionalen Steuerungsmechanismen basierenden Agentenarchitekturen benutzt, die sowohl in virtuellen Agenten als auch in autonomen Robotern implementiert wurden. Das Framework für EBC besteht aus fünf Systemen: dem perzeptuellen, dem motorischen, dem motivationalen, dem emotionalen und dem verhaltensgenerierenden System. Dabei übt das emotionale System direkten Einfluss auf perzeptuelle, motivationale und verhaltensgenerierende Systeme aus und es beeinflusst indirekt auch das motorische System (siehe Abb. 7.6).

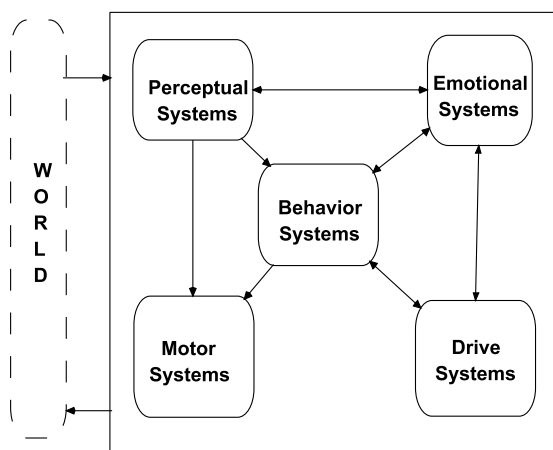


Abbildung 7.6.: Framework for Emotion-Based Control (nach [Vel98], S. 2).

Das perzeptuelle System (*perceptual system*) bekommt Informationen aus der Umgebung und leitet die wahrgenommenen Stimuli an die verhaltensgenerierende (*behavior*) und emotionale (*emotional*) Systeme weiter. Des Weiteren erhalten diese beiden Fehlersignale (*error signals*) vom motivationalen System (*drive system*). Das emotionale System bewertet die emotionale Signifikanz von Stimuli und steuert die emotionale Reaktion bzw. das Verhalten des Agenten sowie die nachfolgende Perzeption in Abhängigkeit von dieser Bewertung. Das verhaltensgenerierende System ist für die Erzeugung von motorischen

7. Architekturen emotionaler Agenten

Aktivitäten zuständig, die durch das motorische System (*motor system*) zur Ausführung gebracht werden.

Dieser Ablauf der Informationsverarbeitung wird in Abb. 7.7 anhand eines Petrinetzes illustriert: Nach der Wahrnehmung des Stimulus aus der Umgebung (Transition **wahrnehmen**) leitet das perzeptuelle System die gewonnenen Informationen an das emotionale und verhaltensregulierende Systeme weiter (Transition **weiterleiten**). Diese werden von dem emotionalen System bewertet (Transition **bewerten**). In Abhängigkeit von den Ergebnissen der Bewertung folgt ggf. die Generierung von Emotionen, welche die anschließende Verhaltensgenerierung, Perzeption und Motivation beeinflussen. Über die Transition **entscheiden, Aktionswahl** werden die Ergebnisse der (emotionsgesteuerten) Entscheidungsprozesse an das motorische System übertragen und die entsprechenden motorischen Aktionen ausgeführt (Transition **ausführen**).

Das Netz aus Abb. 7.7 leitet sich von der allgemeinen Modellbeschreibung in Velásquez (1998) ab (vgl. [Vel98], S. 2). Im Folgenden werden die einzelnen Systeme des Frameworks näher beschrieben, wozu insbesondere die Architekturbeschreibungen in [Vel97], [Vel98] und [Vel99] herangezogen werden. Zusätzlich zu der allgemeinen Modellbeschreibung liefert Velásquez in diesen Arbeiten eine detailliertere Beschreibung der Architekturkomponenten sowie ihrer konkreten Umsetzung. Aus dieser Konkretisierung ergeben sich Ergänzungen und Verfeinerungen für das vorgestellte Petrinetz-Modell, die weiter unten diskutiert werden.

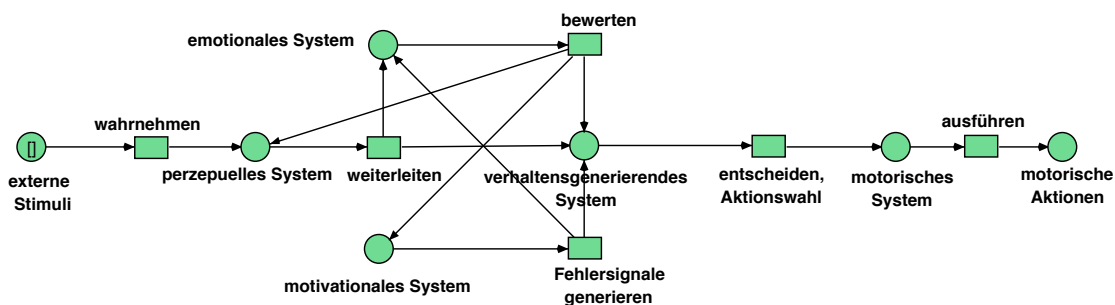


Abbildung 7.7.: Petrinetz-Modell des Frameworks zur emotionsbasierten Steuerung nach [Vel98], S. 2.

Alle fünf Systeme des Frameworks werden durch eine Reihe von nicht linearen Berechnungseinheiten (*computational units*) repräsentiert. Jedes System besteht aus jeweils drei Komponenten ([Vel99], S. 3f.): einer Menge von Inputs, einem Einschätzungsmechanismus und einer Menge von Outputs. Eine wichtige Komponente des Einschätzungsmechanismus sind die so genannten Auslöser (*releasers*): Eine für ihre Realisierung zuständige Berechnungseinheit filtert die Eingabedaten und identifiziert besondere Bedingungen, aufgrund derer die Berechnungseinheit dann exzitatorische oder inhibitorische Outputsignale an die korrespondierenden Systeme sendet.

Dieser Sachverhalt ist in einem Petrinetz in der Abb. 7.8 festgehalten. Durch die Verfeinerung der Stelle **Einschätzungsmechanismus** wird das Filtern der Eingabedaten durch

die Berechnungseinheit und die Generierung der Auslöser (releasers) als Output an die verschiedenen Verarbeitungssysteme des EBC-Frameworks visualisiert.

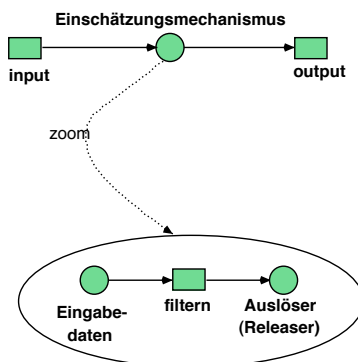


Abbildung 7.8.: Berechnungseinheit (computational unit).

Die allgemeine Form für die Ausgabe bzw. Aktivierung des jeweiligen Systems ist eine nicht lineare Funktion des Inputs dieses Systems, die wiederum zu dem Output seines korrespondierenden Releasers und ihrer Gewichtung proportional ist. Formel (7.1) beschreibt diesen Zusammenhang. Dabei bezeichnet A_i die Aktivierung des Systems i , R_{ki} den Wert des Auslösers k , W_{ki} seine zugehörige Gewichtung (wobei k die Werte aus der Menge aller Auslöser des Systems i angibt) und f eine Begrenzungsfunktion (z.B. eine logistische oder Rampenfunktion).

$$A_i = f\left(\sum_k (R_{ki} * W_{ki})\right) \quad (7.1)$$

Mit dieser Formel lässt sich die Ausgabe von jedem der fünf Systeme beschreiben - lediglich das emotionale System weist einige Besonderheiten auf, von denen weiter unten die Rede sein wird.

Das motivationale System Das motivationale System repräsentiert physiologische Triebe (*drives*), die in dem Modell von Velásquez als Motivatoren zur Befriedigung von grundlegenden Bedürfnissen des Agenten aufgefasst werden und sein Verhalten maßgeblich steuern.⁵ So trägt z.B. der Trieb *Hunger* zur Steuerung der Verhaltensweisen bei, die die Aufnahme von Nahrung direkt beeinflussen.

Velásquez [Vel98] betont jedoch, dass in seinem Modell das emotionale System die wichtigsten Motivatoren für die Aktivitäten des Agenten liefert und das motivationale System übersteuert: “In fact, within the proposed framework, emotional systems constitute the main motivations for the agent and even drive systems exploit their power to bias

⁵Bei Velásquez wird das motivationale System “*Drive System*“ genannt, es produziert “*drives*“ als Output. Den Begriff “*drive*“ übersetzen wir hier nicht mit “Trieb“, sondern mit “Motivation“, um den Zusammenhang mit der Motivation als Ausgangssignal des motivationalen Systems auf der syntaktischen Ebene aufrechtzuerhalten.

7. Architekturen emotionaler Agenten

specific behaviors. In other words, it is both the error signal produced by a *Hunger* drive, and the *Distress* caused by it, that motivates the agent to obtain food.“([Vel98], S. 2).

Die Auslöser (*releasers*) von Trieben werden durch Kontrollsysteme repräsentiert, die den Wertebereich einer bestimmten Variable überwachen. Liegt der Wert der Variable außerhalb der festgelegten Grenzen, wird eine Fehlermeldung (*error signal*) erzeugt. Dieses Signal wird an das Subsystem des motivationalen Systems weitergeleitet, das für den korrespondierenden Trieb zuständig ist. Hier wird sie ggf. mit anderen Fehlersignalen der weiteren relevanten Kontrollsysteme kombiniert: “For instance, in the case of a mobile robot with two different batteries powering its base and the different actuators in its body, a *Recharging Regulation* drive would combine the signals from two different control systems: one controlling the base battery level and one controlling the body battery level“ ([Vel98], S. 2). Die Aktivierung des motivationalen Systems ist eine nicht lineare Kombination der Ausgaben seiner Kontrollsysteme und wird mit der allgemeinen Aktivierungsformel (7.1) beschrieben.

Die Generierung von Motivationen wird in einem Netz in Abb. 7.9 veranschaulicht. Dieses Netz ist eine Verfeinerung des Modells einer Berechnungseinheit aus Abb. 7.8. Das sensorische Input wird an die Kontrollsysteme (Stellen **Kontrollsystem 1** bis **k**) weitergeleitet, welche jeweils die Eingaben filtern und ggf. ein Fehlersignal (Stelle *Fehlersignal*) erzeugen, falls ein durch das betroffene Kontrollsystem überwachter Parameter außerhalb seines zulässigen Wertebereichs liegt. Die Fehlersignale der einzelnen Kontrollsysteme werden kombiniert und bilden den Auslöser (*releaser*) für eine bestimmte Motivation (Stelle **Auslöser (Releaser)**). Die Ausgaben des motivationalen System (d.h. die Motivationen bzw. Triebe, die in Velásquez’ Modell als *drives* bezeichnet werden) werden an die anderen Verarbeitungssysteme des Frameworks weitergeleitet (Transition **Output**, **Motivationsaktivierung**).

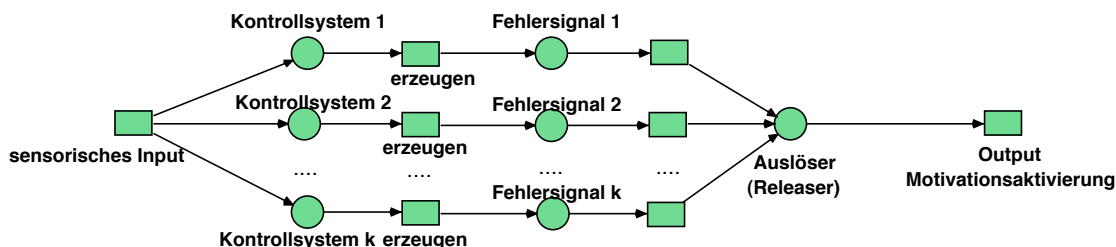


Abbildung 7.9.: Petrinetz-Modell des motivationalen Systems nach [Vel98].

In Abb. 7.10 wird das obige Netz mittels Faltung in ein gefärbtes Netz umgewandelt. Das Netz stellt ebenfalls ein Modell des motivationalen Systems dar, wobei die gleichartigen Elemente des ungefärbten Petrinetz-Modells zu den Stellen **Kontrollsystem**, **Fehlersignal** bzw. Transition **erzeugen** zusammengefasst wurden. Die Variablenbezeichner (k für den Index des Kontrollsystems und f für das jeweilige Fehlersignal) stellen den Bezug zu den jeweiligen Stellen des zugrunde liegenden Netzes in Abb. 7.9 dar.

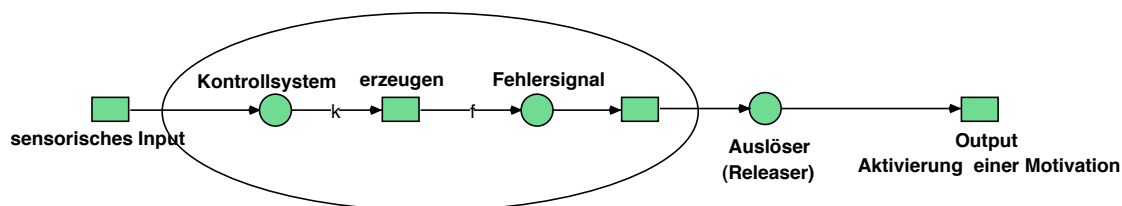


Abbildung 7.10.: Modell des motivationalen Systems nach [Vel98] als Faltung der Abb. 7.9.

Das verhaltensgenerierende System Das verhaltensgenerierende System ist für die Entscheidungsfindung und Aktionswahl des Agenten zuständig. Die Subsysteme des verhaltensgenerierenden Systems repräsentieren verschiedene Verhaltensweisen, die sich gegenseitig behindern oder anregen können. Wie das motivationale und das emotionale System verfügt das verhaltensgenerierende System über Auslöser (releasers), welche den sensorischen Input in Bezug auf bestimmte Eigenschaften filtern, um Faktoren zu identifizieren, die die Aktivierung bestimmter Verhaltensweisen entweder herabsetzen oder steigern. Die Auslöser können sowohl spezielle Eigenschaften repräsentieren (z.B. “battery recharger is present“) als auch Motivationen (z.B. “battery level is low“; “distress is high“), die höchstwahrscheinlich die Aktivierung des *Recharge-Batteries*-Verhaltens steigern werden.

Das Netz in Abb. 7.11 zeigt ein Modell des verhaltensgenerierenden Systems als Verfeinerung des allgemeinen Berechnungseinheiten-Modells aus Abb. 7.8. Die nebenläufigen Verarbeitungspfade repräsentieren Teilsysteme, die für die Generierung verschiedener Verhaltensweisen verantwortlich sind. Diese Teilsysteme werden durch den Input aus der Umgebung oder durch die Ausgabesignale anderer Komponenten des Frameworks (emotionales oder motivationales System; vgl. Abb. 7.7 für eine Gesamtübersicht) aktiviert (Transition Input) und filtern diesen Input, um Auslösefaktoren für die jeweilige Verhaltensweise, für die sie zuständig sind, zu identifizieren (Transition filtern). Wenn eine Auslösebedingung für eine Verhaltensweise identifiziert wird (Stelle Auslöser (releaser)), wird die Aktivierung des jeweiligen Teilsystems erhöht bzw. herabgesetzt und das entsprechende Verhalten verstärkt bzw. verhindert. Der Output des verhaltensgenerierenden Systems wird, entsprechend dem Gesamtmodell (vgl. Abb. 7.7) an das motorische System weitergereicht (Transition Output, Verhaltensgenerierung) und führt dort zur Ausführung der motorischen Aktionen.

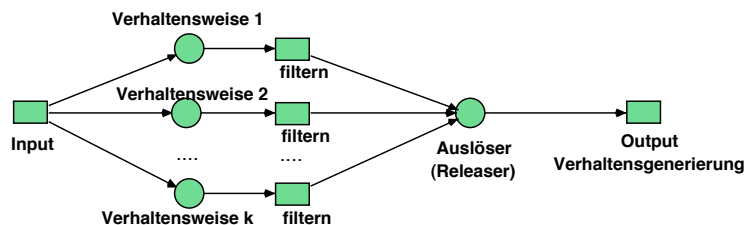


Abbildung 7.11.: Petrinetz-Modell des verhaltensgenerierenden Systems nach [Vel98].

7. Architekturen emotionaler Agenten

In den ersten Implementierungen folgte das verhaltensgenerierende System der “Winner-takes-All“-Strategie, so dass nur eine Verhaltensweise zur gleichen Zeit aktiv sein kann [Vel97]. Dadurch ist jedoch auch die gleichzeitige Ausführung von mehreren nebenläufigen Verhaltensweisen ausgeschlossen, wie z.B. “gehen“ und “sprechen“. Um die gleichzeitige Ausführung von nebenläufigem Verhalten wieder zu ermöglichen, wurde das Verhaltenssystem dahingehend geändert, dass aktive, sich nicht ausschließende Verhaltensweisen die Ausführung von motorischen Aktionen gleichzeitig initiieren können. Das entsprechende Netzmodell für die nebenläufige Generierung von verschiedenen Verhaltensweisen, die anschließend die gleichzeitige Ausführung von motorischen Aktionen initiieren können, zeigt Abb. 7.12.

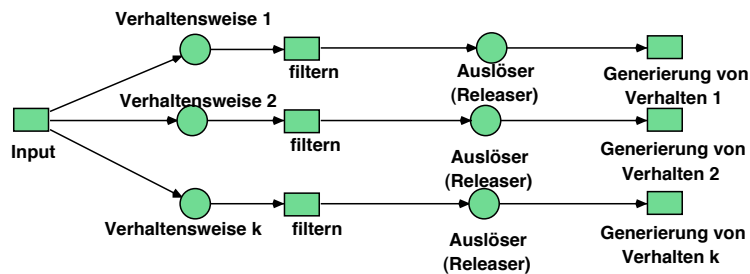


Abbildung 7.12.: Nebenläufige Verhaltensgenerierung nach [Vel98].

Das emotionale System Das emotionale System besteht aus sechs verschiedenen Subsystemen, den sogenannten *Emotionsfamilien*. Auf die Emotionstheorien von Ekman [Ekm92], Izard [Iza91] und Johnson-Laird / Oatley [OJL87] zurückgreifend, postuliert Velásquez sechs verschiedene Emotionstypen ([Vel98], S. 2): *Anger*, *Fear*, *Distress / Sadness*, *Joy / Happiness*, *Disgust* und *Surprise*. Der Wahl dieser Emotionsmenge wird mit der Ekmans Hypothese der universellen Basisemotionen begründet. Damit repräsentiert jedes der sechs Subsysteme eine Basisemotion (bzw. in der Terminologie von Velásquez einen Emotionstyp) und ist für die Generierung dieses bestimmten Emotionstyps zuständig.

Zu jedem dieser Emotionstypen definiert Velásquez eine korrespondierende Emotionsfamilie, bestehend aus einer Gruppe von verwandten affektiven Reaktionen. So bilden Schreck, Angst, Furcht und Panik (*Fright*, *Fear*, *Terror*, *Panic*) die Emotionsfamilie für Angst (*Fear*). Alle Emotionen einer Familie haben gewisse Merkmale und Charakteristiken gemeinsam, z.B. gleiche auslösende Ereignisse, gleichen expressiven Ausdruck, ähnliche Verhaltensantworten und physiologische Reaktionsmuster. Diese Charakteristik variiert zwischen verschiedenen Emotionsfamilien und dient als Unterscheidungskriterium.

Außerdem ist jede Emotionsfamilie durch eine Reihe von veränderbaren Parametern gekennzeichnet: Die Zerfallsfunktion (*decay function*) bestimmt die Dauer der Aktivierung der Emotionen einer Emotionsfamilie; zwei Schwellenwerte legen die Aktivierungsschwelle und die Sättigungsgrenze für alle Emotionen dieser Familie fest. Das Modell von Velásquez

erlaubt die Unterscheidung zwischen drei Arten von affektiven Phänomenen: *Basisemotionen*, *Emotionen höherer Ordnung* und *Stimmungen*. Basisemotionen (*basic emotions*) entstehen im Modell als Aktivierung eines einzigen spezifischen emotionalen Subsystems (Emotionsfamilie), etwa bei der *Disgust* oder der *Fear*-Familie.

Den theoretischen Annahmen aus der Forschung zu Basisemotionen entsprechend ([Plu94], [Iza91]), modelliert Velásquez Emotionen höherer Ordnung (*emotion blends*, *mixed emotions*) als Kombination aus mehreren Basisemotionen. In seinem Modell entstehen *emotion blends* und *mixed emotions*, wenn mehrere emotionale Subsysteme, welche die *basic emotions* hervorbringen, gleichzeitig aktiv sind, ohne dass eines von ihnen die anderen dominiert ([Vel98], S. 3).

Des Weiteren können mit dem Modell von Velásquez Stimmungen (*moods*) dargestellt werden, die sich, entsprechend der neurobiologischen Sichtweisen von Panksepp ([Pan98], [Pan06]), von den Emotionen rein quantitativ unterscheiden, nämlich durch die Höhe des Erregungsniveaus in einem Emotionssystem, das für die Emotionsgenerierung zuständig ist (ebd.).

Schließlich lässt das Modell noch die Implementierung des *Temperaments* eines Agenten zu. Temperamente werden durch die Variation der Emotionsfamilienparameter (Schwellenwerte und Zerfallsrate) innerhalb jedes emotionalen Systems erzeugt. Wenn man zum Beispiel einen grimmigen Agenten modellieren möchte, kann man die Schwellenwerte und die Zerfallsrate für die Emotionsfamilie *Anger* herabsetzen und die gleichen Parameter für die *Joy*-Familie erhöhen.

Im Gegensatz zu den meisten Emotionsmodellen schließt Velásquez nicht nur die kognitive Emotionseinschätzung in sein Modell mit ein, sondern macht auch die metabolischen und regulatorischen Prozesse, welche von einigen Emotionsforschern als Auslöser bzw. Vorstufen von Emotionen angesehen werden (vgl. [Dam03], [Iza91]), zum Modellierungsgegenstand. Diese Prozesse treten im Verlauf der Interaktion mit der Umwelt auf und spiegeln die physikalischen Aspekte dieser Interaktion wider. Sie sind also insbesondere für Agenten relevant, die in einer realen oder simulierten Umgebung agieren, z.B. für Roboter oder synthetische Agenten in virtuellen Umgebungen. Im Falle von nicht eingebetteten Agenten, bei denen die physikalische Interaktionsebene fehlt, erscheint die Modellierung der physikalischen Auslösefaktoren von Emotionen weniger sinnvoll zu sein. Der Klassifizierung von Izard [Iza91] folgend, teilt Velásquez die Auslösefaktoren, die bei der Aktivierung einer Emotion eine Rolle spielen (vgl. *releasers*), in vier Gruppen auf ([Vel98], S. 2):

1. *Neuronale Auslösefaktoren*: Effekte von Neurotransmittern, Gehirntemperatur und andere neuronale Einflussgrößen, die durch Hormone, Schlaf, Ernährung oder Umweltbedingungen bestimmt sind.
2. *Sensomotorische Auslösefaktoren*: Wirkung von sensomotorischen Prozessen wie Gesichtsausdruck, Körperhaltung und Muskelaktivität, die einerseits durch gegenwärtige Emotionen bestimmt werden, andererseits auch selber Emotionen auslösen können.
3. *Motivationale Auslösefaktoren*: Im Modell von Velásquez werden drei mögliche Motivationsgruppen angenommen: Triebe (z.B. Hunger und Durst), Emotionen (z.B.

7. Architekturen emotionaler Agenten

Ärger und Glück) und Schmerzregulation. Als Beispiele für motivationale Auslösefaktoren aus verschiedenen Gruppen nennt Velásquez die angeborene Reaktion auf faulen Geruch oder Geschmack als Motivation für Ekel oder Schmerz oder negative Stimulation als Motivation für Ärger sowie Traurigkeit, die weitere Emotionen (nach Velásquez u.a. Ärger) motivieren kann.

4. *Kognitive Auslösefaktoren*: Dies sind alle Typen von kognitiven Prozessen, die Emotionen auslösen können, z.B. Einschätzung von Ereignissen, Ursachenzuschreibung, Überzeugungen, Wünsche und Erinnerungen.

Velásquez [Vel97] unterscheidet zwischen natürlichen und erlernten Emotionsauslösern: Natürliche Auslöser (*natural releasers*) sind angeborene Mechanismen, die fest in das System eingebaut (*hard-wired*) sind; Erlernte Auslöser (*learned releasers*) repräsentieren Reize, die aufgrund von früheren Erfahrungen des Agenten mit dem Auftreten bestimmter Auslöser assoziiert werden bzw. deren Auftreten voraussagen. Erlernte Auslöser erfordern größere Verarbeitungskapazität und sind komplexer in ihrer Realisierung als natürliche Auslöser, da sie unter anderem auf persönlichen emotionalen Erinnerungen beruhen, die erst einmal aktiviert werden müssen.

In der aktuellen Implementierung des Modells sind alle Auslösefaktoren der ersten drei Kategorien die natürlichen Auslöser (*natural releasers*) und die kognitiven Auslösefaktoren sind die erlernten Emotionsauslöser (*learned releasers*). In den ersten Modellumsetzungen basierten die kognitiven Auslöser auf einer kognitiven Einschätzungstheorie (vgl. [Vel97]), in einer weiteren Version des Architekturmodells sollen die vordefinierten kognitiven Auslösefaktoren nicht mehr existieren, vielmehr sollen neue Emotionsauslöser im Laufe der Interaktion mit der Umgebung erlernt werden - entsprechend dem Mechanismus des somatischen Markers ([Dam94]), dessen Umsetzung im Modell im folgenden Abschnitt näher erläutert wird.

Die Aktivierung des emotionalen Systems unterscheidet sich etwas von der allgemeinen Form der Aktivierung, die mit der Formel (7.1) beschrieben wurde. Die wichtigsten Unterschiede betreffen die zusätzliche Berücksichtigung des exzitatorischen (positiven) und des inhibitorischen (negativen) Inputs aus den anderen emotionalen Subsystemen sowie ihre jeweiligen Zerfallsfunktionen. Formel (7.2) beschreibt die Aktivierungsfunktion für das emotionale System:

$$A_i(t) = f(\psi(A_i(t-1)) + \sum_k (R_{ki} * W_{ki}) + \sum_l \mu_{li} * A_l(t)) \quad (7.2)$$

$A_i(t)$ bezeichnet die Aktivierung des emotionalen Subsystems i zu der Zeit t ; $A_i(t-1)$ ist seine Aktivierung zu einem vorangegangenen Zeitpunkt; $\psi()$ ist eine Funktion, die die Abnahme der Aktivierung des emotionalen Subsystems mit der Zeit beschreibt; R_{ki} ist der Wert des Auslösers k und W_{ki} ist seine zugehörige Gewichtung, wobei k die Werte aus der Menge aller Auslöser des emotionalen Subsystems i einnimmt; $\mu()$ ist die Intensität des exzitatorischen (positiven) und des inhibitorischen (negativen) Inputs seitens eines emotionalen Subsystems l , wobei A der Wert der Aktivierung dieses Subsystems zu der Zeit t und f eine Begrenzungsfunktion (z.B. logische oder Rampenfunktion) ist.

Emotionen höherer Ordnung (*emotion blends*, *mixed emotions*) entstehen als Kombination von mehreren Basisemotionen, d.h. wenn mehrere emotionale Subsysteme gleichzeitig aktiv sind, ohne dass eines von ihnen die anderen dominiert. Die Entstehung der Emotionen höherer Ordnung wird in Abb. 7.13 mit einem Petrinetz dargestellt, in dem die Subsysteme, die für die Generierung der sechs Basisemotionen verantwortlich sind, mittels “oder“-Verknüpfung verbunden werden.

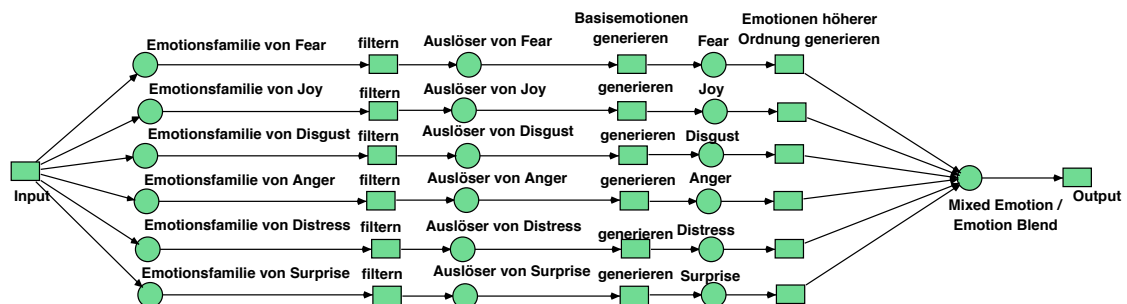


Abbildung 7.13.: Petrinetz-Modell des emotionalen Systems: Emotionen höherer Ordnung

Der Output des emotionalen Systems wird entsprechend dem Gesamtmodell (vgl. Abb. 7.7) an das verhaltensgenerierende System weitergereicht (Transition Output) und beeinflusst somit das Verhalten und den Aktionswahl.

Sekundäre Emotionen als somatischen Marker Die Aktivierung der emotionalen Systeme durch erlernte Emotionsauslöser (*learned releasers*) korrespondiert mit den Konzepten des emotionalen Gedächtnisses [LeD96a] bzw. den sekundären Emotionen ([Dam94], vgl. Kapitel 6.2.1). Die sekundären Emotionen entstehen aufgrund von Assoziationen zwischen bestimmten Objekten bzw. Situationen und im Zusammenhang mit ihnen auftretenden primären Emotionen [Vel98]. Das Besondere am Konzept der *learned releasers* ist, dass sie zur Implementierung von Damasio's Somatic-Marker-Hypothese eingesetzt werden können, nämlich als Heuristiken für die Aktionswahl. Sie erfüllen die Rolle der von Damasio postulierten somatischen Marker, indem die gespeicherte emotionale Reaktion bei der wiederholten Konfrontation des Agenten mit dem gleichen Stimulus reproduziert wird und die Aktionswahl mitbestimmt.

Sekundäre Emotionen erfordern komplexere Berechnungen als primäre Emotionen und benötigen zusätzlichen Speicher für die erlernten Assoziationen. Diese Speicher realisiert Velásquez durch ein assoziatives Netzwerk, in dem einzelne Stimuli (z.B. perzipierte Eigenschaften von Objekten bzw. anderen Agenten) mit den primären Emotionen verbunden werden, die im Laufe der Interaktion des Agenten mit seiner Umwelt im Zusammenhang mit diesen Stimuli aktiviert wurden. Während des emotionalen Lernens können sich die Verbindungen im Netzwerk verändern.

Velásquez illustriert seinen Ansatz durch ein Szenario, bei dem das emotionale Gedächtnis eines Roboters sich durch die Interaktion mit einer Person ausbildet. “In this example,

7. Architekturen emotionaler Agenten

petting and disciplining the robot are natural releasers for the Joy and Fear emotional systems, respectively. When the Fear emotional system becomes active because of the punishing action, the salience of all other present stimuli is determined and the most salient one, in this case the presence of a person, is associated with the Fear emotional system as a new learned releaser [...]. Thus, the next time an agent encounters a marked stimulus such as a person in this scenario, the memory represented in the associative network will be relived, reproducing the emotional responses that occurred previously, and influencing the selection of actions to follow“ ([Vel98], S. 4).

Cathexis Cathexis ist eine Testumgebung für das EBC-Framework - implementiert als Avatar namens Simón the Toddler -, welches das Verhalten und die emotionalen Reaktionen eines Säuglings imitiert. Das Cathexis-Modell verfügt über fünf motivationale Systeme (Hunger, Durst, Temperaturregulation, Ermüdung, Interesse), die sechs oben behandelten Emotionsfamilien und ein Repertoire von 15 Verhaltensalternativen, darunter Schlafen, Essen, Trinken, Lachen, Weinen und mit Spielzeug spielen.

Der Benutzer interagiert mit dem System, indem er zum Beispiel die Parameter von Simóns emotionalen Systemen - sein Temperament - verändert, das Level der Neurotransmitter variiert oder direkt mit Simón kommuniziert, ihn füttert, streichelt usw.

“These interactions provide stimuli to Simón, which, along with the internal stimuli generated by his motivational systems will cause him to react emotionally. For instance, when his desire to play is high, and there are toys available for him to play with, Simón becomes happy. Similarly, when the level of Hunger is high and there is no food around, his level of Distress increases and eventually becomes sad and angry. Once this need is satisfied, his level of Anger and Sadness decays, and he becomes happy.“ ([VM97], S. 5)

Auf dem Bildschirm kann man dann die Reaktionen des “Säuglings“ beobachten, der zu unterschiedlichen emotionalen Ausdrucksweisen und rudimentären Verbalisierungen fähig ist.

Yuppy Yuppy ist ein Roboter, der ein emotionales Haustier darstellt. Er wurde zunächst als virtuelle Simulation entwickelt, bevor er einen Körper erhielt. Sein Wahrnehmungssystem besteht aus zwei Farb-CCD-Kameras als Augen, einem Stereo-Audio-System mit zwei Mikrofonen als Ohren, Infrarotsensoren für die Entdeckung von Hindernissen, einem Luftdrucksensor, um angenehme und schmerzliche Berührungen zu simulieren, sowie einem Pyrosensor, der Veränderungen der Raumtemperatur wahrnimmt, wenn Menschen den Raum betreten.

Yuppys motivationales System umfasst vier verschiedene Motivationen (*drives*): Aufladeregulierung, Temperaturregulierung, Ermüdung und Neugier. Jede dieser Motivationen kontrolliert eine ihr zugeordnete interne Variable, welche den Ladezustand der Batterie, die Höhe der Temperatur, die Menge an Energie oder das Maß des Interesses des Agenten repräsentiert. Yuppy besitzt emotionale Systeme mit fest programmierten natürlichen

Auslösern (*natural releasers*) für jede der sechs oben dargestellten Emotionsfamilien. Velásquez implementierte in Yuppy drei Gruppen von Emotionsauslösern (*releasers*), auf die entsprechend unterschiedliche emotionale Subsysteme ansprechen:

- *Interaktion mit einem motivationalem System*: Aktivierte Motivationen (d.h. unbefriedigte Triebe) erzeugen *Distress* und *Anger*; auch überbefriedigte Triebe produzieren *Distress*, jedoch mit geringer Intensität. Triebbefriedigung erzeugt *Happiness*.
- *Interaktion mit der Umgebung*: Alle Objekte mit rosa Farbe erzeugen *Happiness* in unterschiedlichem Maß - besonders die rosa Knochen; gelbe Objekte produzieren *Disgust*. Dunkelheit induziert Aktivität in das *Fear*-Emotionssystem und laute Geräusche erzeugen *Surprise*.
- *Interaktion mit Menschen*: Menschen können Yuppy tätscheln und disziplinieren. Dies erzeugt entweder Lust oder Schmerz. Lust führt zu *Happiness*; Schmerz produziert *Distress* und *Anger*.

Yuppys Verhaltenssystem besteht aus einem verteilten Netz aus etwa neunzehn unterschiedlichen Verhaltensarten, die vorwiegend die Befriedigung seiner Bedürfnisse und die Interaktion mit Menschen realisieren. Beispiele für solche Verhalten sind *Search-for-Bone*, *Approach-Bone*, *Recharge-Battery*, *Wander*, *Startle*, *Avoid-Obstacle*, *Approach-Person* und *Express-Emotion*.

Der Benutzer kann Yuppys Temperament durch die Manipulation von Emotionsfamilienparametern wie Schwellenwerte, inhibitorische oder exzitatorische Verbindungen usw. kontrollieren. Außerdem kann er dem Roboter innere und äußere Reize darbieten.

“For instance, when the robots *Curiosity drive* is high, Yuppy wanders around, looking for the pink bone. When it encounters one, the activity of the Happiness emotional system increases and specific behaviors, such as Wag-Tail and Approach-Bone become active. On the other hand, as time passes by without finding any bone, the activity of its Distress emotional system rises and appropriate responses, such as Droop-Tail, get executed. Similarly, while wandering around, it may encounter dark places, which will elicit fearful responses in which it backs up and changes direction.” ([Vel98], S. 5).

Außerdem implementiert Yuppy den oben beschriebenen Mechanismus des emotionalen Lernens und der Entscheidungsfindung mit Hilfe von sekundären Emotionen: “Yuppy will create positive or negative emotional memories with respect to people, and future selection of behaviors such as approaching or avoiding them will be influenced.” ([Vel98] S. 6).

7.2.3. Dörners PSI und Bachs MicroPSI

Das PSI-Modell wurde von dem Psychologen Dietrich Dörner ([DLD⁺01], [DSD01]) auf Basis einer integrativen kognitiv-emotional-motivationalen Theorie der Handlungsregulation (PSI-Theorie) entwickelt. Es stellt ein psychologisches Modell im Sinne von Pfeifers Definition dar ([Pfe88], siehe Kapitel 7.1), da es vor allem zur Simulation, Konsistenzprüfung und

7. Architekturen emotionaler Agenten

Validierung der PSI-Theorie gebaut wurde. Das Modell integriert kognitive, motivationale und emotionale Prozesse, die an der Absichten- und Handlungsregulation beteiligt sind ([DLD⁺01], [DSD01]).

Dörners Modell weist eine gewisse Ähnlichkeit zu Slomans Herangehensweise auf. Beide Ansätze streben die Konstruktion eines kompletten autonomen Systems ohne separates Emotionsmodul an und betrachten Emotion als einen funktionalen Mechanismus, der infolge von Regulationsprozessen des Gesamtsystems zustande kommt.

Eine Erweiterung des PSI-Modells stellt die MicroPSI-Architektur für einen autonomen situierten Agenten dar, die viele Thesen der Dörner-Theorie in eine Agentenarchitektur umsetzt und weiterentwickelt. In diesem Abschnitt werden zuerst die wesentlichen Eigenschaften des PSI-Modells beschrieben und anschließend die Architekturkomponenten dargestellt, die in MicroPSI teilweise modifiziert oder neu hinzugefügt wurden.

Dörners PSI-Modell

Dem PSI-Modell liegt die Theorie der Persönlichkeits-System-Interaktionen von Dietrich Dörner ([DLD⁺01], [DSD01]) zugrunde, die menschliches Handeln auf eine Reihe von Informationsverarbeitungsprozessen zurückführt und dabei die wesentlichen psychischen Funktionen wie Wahrnehmung, Denken, Lernen, Emotion, Motivation und Gedächtnis untersucht: “Die Theorie erklärt zielgerichtetes Verhalten genau so gut wie viele verschiedene Formen emotional modulierten Verhaltens, wie z.B. das Verhalten bei Angst, Panik oder Stress.“ ([DSD01], S. 1).

Insbesondere möchte Dörner untersuchen, wie Menschen Ziele und Absichten in Abhängigkeit von den aktuellen Bedürfnissen und Fähigkeiten und von Umgebungs- und Situationsmerkmalen entwickeln, nach welchen Heuristiken sie sich für eines der zahlreichen Ziele entscheiden und wie sie die anschließenden zielgerichteten Verhaltensweisen - wie Planen, Explorieren oder Handeln - auswählen und der jeweiligen Situation anpassen. Dörner betrachtet schwerpunktmäßig die Rolle der Emotion bei Perzeption, Aktionswahl, Planen und Gedächtniszugriff ([DLD⁺01], [DSD01]).

Das PSI-Modell baut damit auf einer theoretischen Grundlage auf, die speziell auf eine spätere Umsetzung in einem Agentenmodell ausgerichtet ist. Deshalb zielt sie nicht primär darauf ab, mentale Prozesse wie Emotionen, Lernen usw. zu simulieren, sondern sie will diese vielmehr formal und robust genug definieren, um eine Implementierung dieser Konzepte zu ermöglichen:

“Konzepte z.B. zur Wahrnehmung, zu Emotionen und zu Gedächtnisprozessen sind soweit formalisiert, dass sie als Computerprogramm der menschlichen Psyche implementiert sind. Das Verhalten dieser Computerimplementation wurde bereits hinsichtlich einiger Aspekte mit dem Verhalten von Menschen (Schaub, 1993, 1997; Hille, 1997; Bartl und Dörner, 1998) und kognitiven Architekturen (Westermann und Heise, 1996; Detje, 1999) verglichen“ ([DLD⁺01], S. 3)

Die PSI-Agenten sind virtuelle Dampfmaschinen, die sich in einer simulierten Inselwelt bewegen und mit dieser über verschiedene Aktuatoren und Sensoren interagieren (siehe Abb. 7.14).

7. Architekturen emotionaler Agenten

Affiliation wird durch die *Legitimitätssignale* (*L-Signale*) erhöht. Legitimitätssignale indizieren, dass ein Individuum sich im Einklang mit den Gruppennormen befindet.

Energie und Affiliation verschwinden mit der Zeit und müssen durch die Interaktion der PSI-Agenten mit der Umgebung “aufgetankt” werden. Die kognitiven Bedürfnisse nach Kompetenz und Bestimmtheit motivieren den Agenten zur Erzeugung neuer explorativer Strategien, begrenzen diese jedoch auf die Bereiche, in denen sich die Interaktion mit der Umgebung bereits als erfolgreich erwiesen hat.

Motivation: Wenn ein Ist-Wert eines bestimmten Bedürfnisses von seinem Sollwert abweicht, entsteht ein *Bedarf* (*urge*), dessen Intensität zu dem Ausmaß und der zeitlichen Ausdehnung der Abweichung proportional ist. Ein Bedarf aktiviert einen *Motivator*, dessen Aktivierungsgrad der Intensität des Bedarfs entspricht. Der Motivator kann die Handlungssteuerung übernehmen, um den Bedarf, der ihn verursacht hat, zu beseitigen, d.h. eine Zielsituation erreichen, in der die Bedürfnisbefriedigung möglich ist. Ziele “erkennt” der Motivator dadurch, dass er bestimmte sensorische Schemata im Gedächtnis von PSI (s.u.) referenziert.

Damit stellen die Bedürfnisse einen Antriebsmechanismus für das Handeln und die Interaktion von PSI mit der Umwelt dar. Die verschiedenen Grundbedürfnisse von PSI sind *Unversehrtheit, Energie, Wasser, Bestimmtheit, Kompetenz und Affiliation*). Wird eines dieser Bedürfnisse nicht befriedigt, entsteht ein Bedarf.

Bei materiellen Bedürfnissen (Energie, Wasser und Unversehrtheit) ist das der Fall, wenn die entsprechenden Sensoren einen mangelnden Energie- bzw. Wasserpegel melden, oder eine Verletzung der äußeren Hülle von PSI feststellen.

Bei informationellen Bedürfnissen wird ein Bedarf aktiv, wenn die entsprechenden Signale überwiegen: Z.B. entsteht der Bedarf nach Bestimmtheit, wenn die Anzahl der Unbestimmtheitssignale die Bestimmtheitssignale dominiert. Der Bedarf nach Kompetenz wird ausgelöst, wenn die Zahl der Ineffizienz-Signale die der Effizienzsignale übersteigt. Schließlich wird das Bedürfnis nach sozialer Anerkennung aktiviert, wenn die Anzahl der Legitimitätssignale kleiner als der vorgegebene konstante Sollwert ist.

Die Aktiviertheit eines Bedarfs wird bei Dörner als Akkumulation der Signale über die Abweichung des Ist- vom Sollzustand beschrieben. Jeder Bedarf aktiviert einen spezifischen Motivator, welcher umso aktiver ist, je größer die Sollwertabweichung ist und je länger sie andauert. Die Aktivität des Motivators wird als Logarithmus der Summe der Abweichungssignale berechnet. Die quantitative Charakterisierung der Bedarfsentstehung und Motivatorgenerierung kann mit einfachen / gefärbten Petrinetzen nicht ohne Weiteres beschrieben werden und wird an dieser Stelle ausgelassen.

Der Motivator versucht nun, die Handlungssteuerung von PSI zu übernehmen, um Handlungen in Gang zu setzen, die dazu führen, dass das Bedürfnis befriedigt wird. Um zwischen mehreren aktiven Motivatoren entscheiden zu können, verfügt PSI über einen Motivselektor. Dieser wählt jeweils einen der Motivatoren aus und lässt ihn zur Handlungsleitung zu. Dabei wird der Motivselektor demjenigen Motivator den Zugang zur Handlungsleitung gewähren, dem das stärkste Bedürfnis (mit dem augenblicklich größten “Wert”) zugrunde liegt und bei dem das Streben nach der konsumatorischen Endhandlung den augenblicklich

größten Erfolg verspricht (die Erfolgserwartung also am größten ist). Diese beiden Größen werden im einfachsten Fall multipliziert und ergeben die Motivstärke eines Bedürfnisses; es wird das Bedürfnis mit der maximalen Motivstärke ausgewählt. Der ausgewählte Motivator übernimmt die Steuerung der Handlungswahl und -konstruktion und übt Einfluss auf die Wahrnehmung von PSI aus.

Der Mechanismus der Motivauswahl sowie der gesamte Motivationsprozess könnten auf einfache Weise mit einem Referenznetz modelliert werden. Diese Modellierung wird jedoch hier nicht vorgenommen, da wir in diesem Kapitel in erster Linie die allgemeinen Aufbauprinzipien der Architekturen emotionaler Agenten und ihre prinzipielle Abbildbarkeit mit Petri- bzw. Referenznetzen zeigen wollen, ohne bei der Modellierung einzelner Architekturelemente zu sehr ins Detail zu gehen.

Emotion: Außer über den motivationalen “Antrieb“ verfügt die PSI-Modellstruktur über eine emotionale Verhaltenssteuerung [DLD⁺01]. Emotionen besitzen in dem Agentenmodell keine explizite Repräsentation: Vielmehr werden sie durch das Verhalten des Agenten deutlich. Die Konzeption von Emotion als einer Klasse von mentalen Zuständen, die sich im Verhalten eines eingebetteten Agenten manifestiert, ist dem Konzept der “intentional stance“ [Den71] bzw. der Sicht auf Emotionen als wichtige Wirkmechanismen innerhalb der Agentenarchitektur von Aaron Sloman [Slo01] relativ nah.

Dörner [DLD⁺01] vertritt in seiner Theorie folgende Auffassung über Emotion bzw. einen emotionalen Zustand, die auch von Bach geteilt wird:

“If we say, that someone “has an emotion“, we imply that this individual is inclined to certain ways of perception, may prefer certain kinds of action or may plan in a different way, and so on. In the understanding of Dörner’s ‘Psi’ theory, these implications are what makes up these emotions, i.e. if the system has sufficient means of perception, interaction, planning etc. and can modify them to adapt to changing situations, then this system is effectively undergoing (but not necessarily experiencing) emotional states“ ([Bac02], S. 4).

Abb. 7.15 veranschaulicht die emotionale Regulation in PSI.

Emotionen entstehen in der PSI als Wechselspiel der *Modulationsparameter*, nämlich von *Aktiviertheit*, *Selektionsschwelle* und *Auflösungsgrad*, und der Bedürfnispegel, insbesondere der Bedürfnisse “Bestimmtheit“ (Stelle *Bestimmtheit*) und “Kompetenz“ (Stelle *Kompetenz*).

Die *Aktiviertheit* (Stelle *Aktiviertheit*) ist eine Art Handlungsbereitschaft von PSI. Die erhöhte Aktiviertheit entsteht z.B. bei erhöhtem Bedürfnisdruck. Sie steuert die beiden anderen Modulationsparameter (Stellen *Auflösungsgrad* und *Selektionsschwelle*).

Der *Auflösungsgrad* steuert die Genauigkeit der Wahrnehmung (Transition *Wahrnehmung*): “Da Wahrnehmen auf Vergleichen der Reize mit Gedächtnisschemata basiert, ergibt sich bei einem geringen Auflösungsgrad “Überinklusivität“. PSI kann unter solchen Umständen Dinge, die einander nur ähnlich sind, miteinander verwechseln.“ ([DSD01], S. 8). Die mit einem gesenkten Auflösungsgrad verbundene Überinklusivität führt zu groben, kurzen Plänen, die fälschlicherweise leicht für zielführend gehalten werden. Daraus resultieren risikobehaftete Aktionen. Je höher der Auflösungsgrad, desto mehr Vergleiche gibt es

7. Architekturen emotionaler Agenten

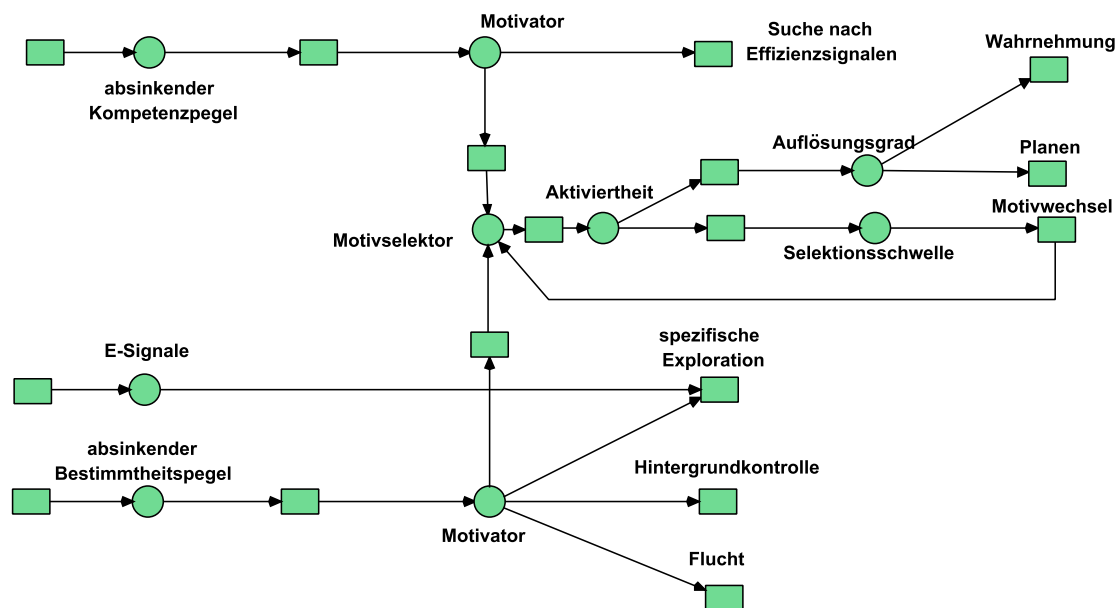


Abbildung 7.15.: Emotionale Regulation eines PSI-Agenten

zwischen den wahrgenommenen Reizen und internen Schemata und desto geringer ist folglich die Verwechslungsgefahr und desto genauer bzw. sicherer sind die erstellten Handlungspläne. Der Einfluss des Auflösungsgrades auf das Planen wird durch die Transition **Planen** modelliert.

Die *Selektionsschwelle* bestimmt die Konzentration von PSI auf seine aktuelle Absicht bzw. sein Motiv (Transition **Motivwechsel**). Die Selektionsschwelle steigt bei erhöhter Aktiviertheit. „Ist die Selektionsschwelle hoch, so kann PSI nur schwer abgelenkt werden; es konzentriert sich auf seine Aufgabe und ist relativ unempfindlich für „Seitenreize“ ([DLD⁺01], S. 29).

Die Modulationsparameter und die sich ändernden Bedürfnispegel für Bestimmtheit und Kompetenz modifizieren sich gegenseitig und beeinflussen auf diese Weise das gesamte kognitive System von PSI: „The modulation is designed to allocate mental resources in way that is suitable to a given situation and reduce the computational complexity of the tasks at hand.“ ([Bac03], S. 16).

Die absinkende Bestimmtheit z.B. erhöht die Empfindlichkeit für Reize aus der Umgebung und führt zur verstärkten Umgebungsüberwachung (**Hintergrundkontrolle**): PSI wendet sich häufiger als gewöhnlich von seiner augenblicklichen Absicht ab, um die Umgebung zu kontrollieren. Die Tendenz zu Fluchtverhalten (Transition **Flucht** - bei gleichzeitig niedrigem Kompetenzpegel) oder aber zu Verhaltensweisen der spezifischen Exploration (Transition **spezifische Exploration** - bei noch relativ hohen Kompetenzwerten) wird auf diese Weise erhöht. Ein niedriger Bestimmtheitspegel führt dazu, dass PSI „zögerlicher in seinem Verhalten wird, nicht so schnell zu Aktionen übergeht, länger plant als es dies unter anderen Umständen tun würde, nicht so mutig ist beim Erkunden. Flucht bedeutet also eine große Menge verschiedenartiger Aktivitäten, die aber alle das Ziel haben, das Ausmaß des

Kontaktes mit der unbestimmten Realität möglichst zu vermindern.“ ([DLD⁺01], S. 28).

Die absinkende Kompetenz erhöht das Bestreben, Effizienzsignale zu finden (Transition *Suche nach Effizienzsignalen*). Diese Verhaltenstendenz kann sehr verschiedene Formen annehmen. Die primitive Form der Suche nach Effizienz ist eine unmittelbare Aktion, die sofortige Effekte zeigt. Eine konstruktivere Form der Kompetenzsteigerung ist z.B. eine durchgeplante Befriedigung eines Bedürfnisses, etwa nach Energie oder Affiliation.

Zusammenfassend sieht man, dass Emotionen in PSI nicht deliberativ, etwa in einem speziellen Emotionsmodul erzeugt werden, sondern als Seiteneffekte der kognitiven Regulationsprozesse entstehen und deshalb nicht nur am simulierten Gesichtsausdruck von PSI, sondern auch an seinen Verhaltensweisen beobachtbar sind, die oft emotionale Färbung aufweisen. “Das Zusammenwirken der verhaltenssteuernden Wirkung des Kompetenzpegels und des Bestimmtheitspegels zusammen mit den Schwankungen der Modulationsparameter “Aktiviertheit“, “Selektionsschwelle“, “Auflösungsgrad“ erzeugt bei PSI unterschiedliche Verhaltensweisen, die als verschiedene Emotionen identifizierbar sind“ ([DSD01], S. 9).

Das Gedächtnis von PSI: Das Gedächtnis von PSI ist ein konzeptuelles Gedächtnis, in dem die erworbenen *sensorischen und motorischen Schemata* gespeichert werden. Sensorische Schemata werden durch Abstraktion der perzipierten Eigenschaften der Objekte und Situationen zu allgemeinen *Kategorien* gewonnen. Sie dienen der Identifizierung und Klassifikation der neuen Objekte, denen PSI erstmalig begegnet. Motorische Schemata, auch *Automatismen* genannt [Bac03], sind die in Form von Aktionsketten gespeicherten Pläne oder *Verhaltensprogramme*, die sich zu früheren Zeiten in bestimmten Situationen als erfolgreich erwiesen haben. Die Gedächtniskomponente der PSI-Architektur weist keine Unterteilung in Langzeit- und Kurzzeitgedächtnis auf. Dörner weist jedoch darauf hin, dass sein Absicht-Konzept (s.u.) in etwa dem Arbeitsgedächtnis von PSI entspricht [DSD01]. Durch dieses Konzept werden die mit dem aktuellen Motivator verknüpften Informationen über anzustrebende Ziele, anwendbare Operatoren oder Aktionsschemata sowie vergangene vergebliche Ansätze und Pläne zusammengefasst.

Lernen: Grundbedürfnisse von PSI sind fest vorgegeben, die Vorgehensweisen zu ihrer Befriedigung müssen jedoch erst erlernt werden. Um das konzeptuelle Wissen über Objekte und Situationen zu erwerben, verfügt PSI über eine “perzeptive Lernfähigkeit“ ([DSD01], S. 7), die das Entwickeln von sensorischen Schemata ermöglicht. Das Lernen erfolgt sowohl über die Perzeption von externen Situations- und Objekteigenschaften mit Hilfe von externen Sensoren, als auch über die Wahrnehmung von Signalen der internen Sensoren, die Bedürfnisbefriedigung signalisieren. Auf diese Weise lernt ein PSI-Agent im Laufe der Zeit, welche Situationen geeignet sind, bestimmte Bedürfnisse zu befriedigen, durch welche Handlungsketten sie zu erreichen sind, welche Strategien besonders erfolgversprechend sind usw.

Handlungsregulation: Die Handlungssteuerung wird von einer *Absicht* geleitet, die eine Kombination ist aus dem jeweils ausgewählten Motiv und den Schemata, die mit dem aktiven Motivator im Gedächtnis verknüpft sind. Der Motivator referenziert z.B. das

7. Architekturen emotionaler Agenten

sensorische Schema der Zielsituation sowie die motorischen Schemata der anwendbaren Handlungsstrategien und der vergangenen, effizienten (oder auch erfolglosen) Aktionsketten und Pläne. Zunächst versucht PSI, in seinem Gedächtnis einen Weg von der augenblicklichen Situation zum angestrebten Ziel zu finden, einen sogenannten Automatismus. Wird kein Automatismus gefunden, versucht es, durch das Planen eine effektive Aktionskette zu bilden, die zur Zielsituation führt. Das Planen wird in PSI als einfache Bottom-up-Suche (Vorwärtsverkettung von Aktionsschemata) implementiert. Längeres erfolgloses Planen führt zur stetigen Kompetenzminderung und irgendwann zum Abbruch. Wenn das Planen kein Ergebnis liefert, generiert PSI ein Verhalten nach dem Versuch-Irrtum-Prinzip. Es probiert zufällige Handlungen aus, wobei es solche bevorzugt, die es in dieser Situation noch nicht verwendet hat.

In Abb. 7.16 wird die Gesamtarchitektur eines PSI-Agenten mit einem Petrinetz-Modell veranschaulicht. Dabei werden insbesondere die strukturellen Zusammenhänge formalisiert, während von den quantitativen Implementierungsdetails abstrahiert wird.

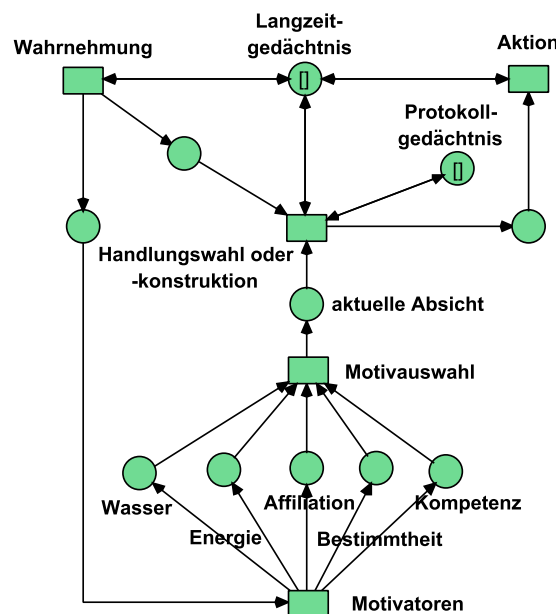


Abbildung 7.16.: Architektur eines PSI-Agenten

Die Motivauswahl wird vereinfacht als willkürliche Auswahl zwischen den fünf Motivatoren dargestellt. Durch die Konzentration auf ein Motiv (Transition **Motivauswahl**) wird eine handlungsleitende Absicht (Stelle **aktuelle Absicht**) festgelegt und verfolgt. Diese steuert zusammen mit den Informationen aus der Umgebung (Transition **Wahrnehmung**) die aktuellen Handlungen eines Agenten (Transition **Handlungswahl oder -konstruktion**). Bei der Handlungswahl werden die Informationen bezüglich verschiedener Handlungsalternativen aus dem Gedächtnis von PSI abgerufen (Stellen **Langzeitgedächtnis** und **Protokollgedächtnis**). Wenn keine zielführenden Automatismen zu finden sind, werden neue Handlungen

konstruiert. Diese werden wiederum protokolliert und können in den zukünftigen Handlungswahlsituationen benutzt werden. Die generierten Handlungen werden anschließend ausgeführt (Stelle Aktion).

Bach: MicroPSI-Architektur

Der Einfluss der PSI-Theorie innerhalb der KI- und VKI-Gemeinschaft blieb begrenzt, teilweise wegen der fehlenden Formalisierung und der manchmal unpräzisen und unvollständigen Formulierungen ([Bac03], S. 15; [vSM04], S. 13). Um die PSI-Theorie für ein breiteres Anwendungsfeld verwendbar und zugänglich zu machen, hat Bach zur Formalisierung der Theorie seine MicroPSI-Architektur für situierte autonome Agenten entwickelt ([Bac02], [Bac03], [Bac08]). Die MicroPSI-Architektur übernimmt wesentliche Aufbauelemente und Mechanismen der PSI-Architektur, implementiert jedoch eine Reihe von Erweiterungen und Verfeinerungen der ursprünglichen Architektur. Im Modell von Bach werden die *NodeNets* zur Repräsentation aller Arten von Schemata verwendet - ein Formalismus, welcher die Grundprinzipien der neuronalen und der semantischen Netze vereint. Ferner werden das Gedächtnissystem in Langzeit- und Kurzzeitkomponenten aufgeteilt, zusätzliche Kontrollstrukturen wie Metamanagement und Alarms eingeführt sowie ein modularer Architekturaufbau realisiert.

MicroPSI besteht aus mehreren Komponenten, die über eine Menge von somatischen Parametern (entsprechend den Grundbedürfnissen von PSI) und externen Sensoren mit der Umgebung kommunizieren. Aus der Modulation dieser Parameter werden automatisch Grundbedürfnisse (*urges*) des Agenten abgeleitet. Das Handlungsrepertoire des Agenten wird durch eine Reihe von internen und externen Verhaltensmodulen bestimmt, wobei die Ersteren die "mentalen" Aktivitäten repräsentieren, während die Letzteren die Interaktionssequenzen mit der Umgebung bewerkstelligen. Abb. 7.17 zeigt die Architektur eines MicroPSI-Agenten.

Repräsentation mit NodeNets Dörner hat in [Dör94] Verwendung einer Art neuronaler Netze vorgeschlagen, um Kontrollstrukturen, deklaratives Gedächtnis und Protokollgedächtnis zu repräsentieren. Generell diskutiert er die Anwendung einer Datenstruktur, mit der sowohl symbolisches Wissen als auch Verknüpfungen zwischen Wissenskategorien repräsentiert werden könnten.

In MicroPSI-Agenten werden die genannten Strukturen mit NodeNets repräsentiert. Sie fungieren als eine universelle Datenstruktur für Perzeption, Speicherung und Planen: Objekte, Situationen, Kategorien, Aktionen, Episoden und Pläne sind in MicroPSI als hierarchische Netzwerke von Knoten repräsentiert. Knoten können in die gewichteten Konjunktionen und Disjunktionen der untergeordneten Knotennetze expandiert werden und schließlich in ihre Referenzen auf Sensoren und Aktuatoren zerlegt werden. Ebenso wie bei PSI-Agenten werden dadurch die Objekte und Situationen sowie die Aktionspläne zu Kategorien zusammengefasst und in Form von sensorischen und motorischen Schemata im Gedächtnis gespeichert. "Thus, the semantics of all acquired representations result from interaction with the environment or from somatic responses of the agent to external or internal situations." ([Bac03], S. 16).

7. Architekturen emotionaler Agenten

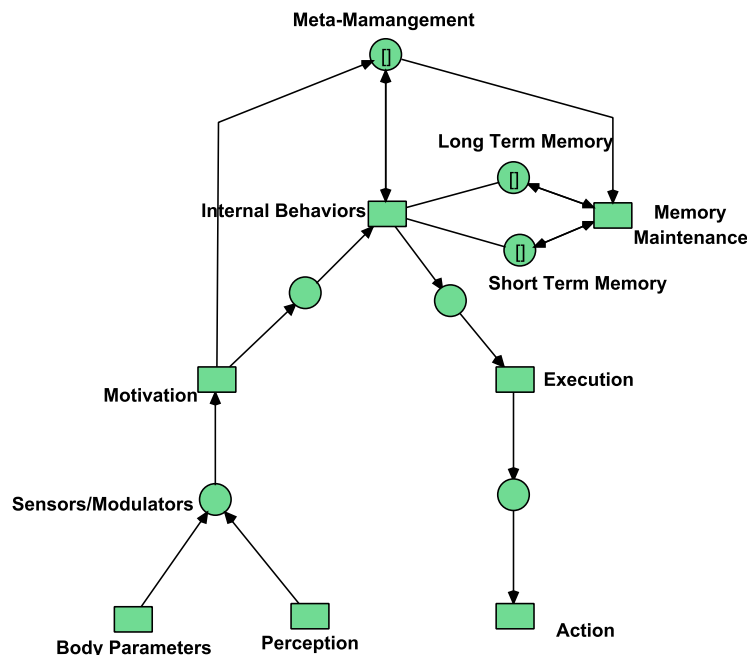


Abbildung 7.17.: Architektur eines MicroPSI-Agenten (aus [Bac03], S. 16).

NodeNets sind gerichtete Netzwerke, die sowohl symbolisches Wissen als auch konnektionistische Konfigurationen repräsentieren können. Sie bestehen aus Knoten, die Eingabe (*slots*), Ausgabe (*gates*), Ausbreitungs- und Knotenfunktion besitzen. Wenn ein Knoten aktiviert wird, führt er seine Knotenfunktion aus. Die Aktivierung kann an der Ausgabe (gate) eines Knotens gespeichert werden, welche einen Schwellenwert und einen Verstärkungswert besitzt. Über die Kanten, die die Ausgänge eines Knotens mit den Eingängen von anderen verbinden, kann sich die Aktivierung über das Netzwerk ausbreiten.

Grundsätzlich kann diese Netzstruktur auch mit Petrinetzen beschrieben werden (vgl. Abb. 7.18). Ein Knoten kann durch eine Transition repräsentiert werden, wobei der Vorbereich der Transition (Stelle **Ein**) dem Eingang und der Nachbereich (Stelle **Aus**) dem Ausgang des NodeNet-Knotens entspricht. Die Ausbreitungs- bzw. Knotenfunktionen können mittels Erweiterung eines einfachen P/T-Netzes zu einem Referenznetz modelliert werden: Beim Schalten der Transition **Knoten aktivieren**, welche die Aktivierung eines NodeNet-Knoten repräsentiert, wird die Knotenfunktion ausgeführt, indem die entsprechende Netzinstanz (Transitionen **Knotenfunktion**) über einen synchronen Kanal aufgerufen wird. Die explizite Modellierung der Ausbreitungsfunktion entfällt, da sie bei der Repräsentation mit dem Petrinetz-Formalismus implizit durch die Netzstruktur abgebildet wird. Die Schwellenwerte an den Ausgängen der NodeNet-Knoten können durch die entsprechende Gewichtung von Ausgangskanten der Referenznetz-Transitionen wiedergegeben werden.

Knoten können eine unterschiedliche Anzahl von Eingängen und Ausgängen enthalten, welche verschiedene Datentypen übertragen bzw. verschiedene Typen von Beziehungen

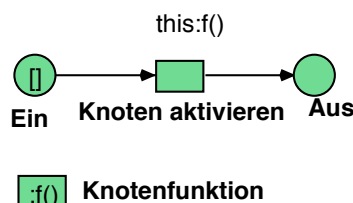


Abbildung 7.18.: Modellierung eines NodeNet-Elementes (Knoten) mit einem Referenznetz.

zwischen den Knoten repräsentieren. Diese unterschiedlichen Datentypen könnten durch ein gefärbtes Netz modelliert werden, indem Vor- und Nachbereiche der Transition **Knoten aktivieren** durch mehrere Stellen mit verschiedenen Farbzusweisungen repräsentiert werden.

Der formale Beweis der Transformierbarkeit der NetNodes zu P/T-Netzen bzw. CPN oder Referenznetzen würde den Rahmen dieses Kapitels sprengen, deshalb kann an dieser Stelle keine abschließende Beurteilung dahingehend erfolgen, ob alle Eigenschaften von NodeNets mit den Petrinetz-Formalismen repräsentiert werden können. Aufgrund einer gewissen Ähnlichkeit in der Struktur der Grundelemente und in der Art der Aktivierungsverbreitung liegt jedoch die Vermutung nahe, dass sich die Grundfunktionalität der NodeNets mit Referenznetzen auf eine relativ einfache Weise abbilden lässt.

Es werden mehrere verschiedene Typen von Knoten unterschieden:

- **Konzept-Knoten** sind typische Bausteine von MicroPSI-NodeNets. Sie repräsentieren semantische Beziehungen, insbesondere kausale Vorwärts- und Rückwärts-Verknüpfungen, *Teil-von*- und *enthält*-Relationen. Üblicherweise besitzt ein Konzept-Knoten einen Eingang und neun Ausgänge, von denen jeder eine bestimmte Beziehung zu dem darüber verknüpften Knoten ausdrückt: So leitet ein *gen*-Ausgang die allgemeine Aktivierung weiter, *por* / *ret*-Ausgänge repräsentieren eine kausale Beziehung im Sinne von “wenn, dann“, *sur* / *sub*-Ausgänge stellen die *Teil-von*- und *enthält*-Relationen dar, *cat* / *exp*-Ausgänge - die *Kategorie*- und *Exemplar*-Beziehungen und *gsym* / *ref*-Ausgänge dienen der Benennung von Konzepten.
- **Register-Knoten** bilden einen weiteren grundlegenden Knotentyp. Sie bestehen aus genau einem Eingang und einem Ausgang, beide vom Typ *gen*, und dienen der Übergabe von Parametern an Module.
- **Sensor-Knoten** geben Werte von externen Datenquellen weiter.
- **Aktoren-Knoten** geben Werte an externe Datenziele weiter.

Zusätzlich zu diesen vier Basisknotentypen wurde eine Reihe von abgewandelten Register-Knoten eingeführt, die die Kontrollstrukturen über NodeNets implementieren (*activators*, *general activation nodes*, *general deactivation nodes*, *associator nodes*, *disassociator nodes*).

Konzept-Knoten ermöglichen das Bilden von komplexen Strukturen und Abstraktionen: *Knoten-Skripte* bestehen aus Ketten von *por* / *ret*-Verknüpfungen; mittels *sur* /

7. Architekturen emotionaler Agenten

sub-Verknüpfungen werden Makros und Hierarchien definiert. Durch eine entsprechende Verknüpfung der Teile können Alternativen oder Konjunktionen realisiert werden. Die unterste Ebene der Hierarchien bilden immer Sensor- oder Aktoren-Knoten.

Aufgrund dieser Strukturen ist es möglich, hierarchische Pläne mit Tausenden von Basisaktionen auf der untersten Hierarchieebene und mit wenigen abstrakten Elementen auf der obersten Ebene auszuführen. Dadurch wird die rechnerische Komplexität der Konstruktion von Plänen beträchtlich reduziert.

Des Weiteren werden sogenannte *native modules* definiert. Das sind Netzeinheiten, die eine beliebige Anzahl von Ein- und Ausgängen haben und als Knotenfunktion einen Programmcode in einer beliebigen Programmiersprache beherbergen können. In der aktuellen Implementierung können *native modules* Java Code enthalten und machen dadurch beliebige Operationen über die Knotennetze möglich, was einen Vergleich zu Objektnetzen nahelegt. Mit dem Modul *Script Execution* wird die Ausführung von Scripten (hierarchischen Plänen) realisiert. Weitere Module implementieren die Wahrnehmung (Modul *Perception*) und die emotionale Steuerung (Modul *Emotion Regulation*) eines MicroPSI-Agenten.

Hypothesenbasierte Perzeption: Externe Perzeption wird von Hypothesen über die Umgebung abgeleitet, die anschließend mit den unmittelbaren externen Perzeptionsergebnissen verglichen werden: “Thus, activation spreads upward, “suggesting“ theories of what is there to be perceived in the environment, then down, “testing“ those theories, then up again, confirming or disconfirming them. The remaining confirmed concepts are considered the immediate percepts of the agent and are made available in a distinct node-space for further processing“ ([Bac05] S. 13). Ein solcher Perzeptionsmechanismus wurde in Dörners Theorie beschrieben (vgl. [DLD⁺01], S. 15) und dort als *hypothesenbasierte Perzeption* bezeichnet.

Durch den Mechanismus der *Assimilation* können die wahrgenommenen Objekte den bereits existierenden Schemata zugeordnet und somit bestimmten Objektklassen zugewiesen werden.

Wenn die Hypothesen des Agenten über die Zugehörigkeit des Objektes zu einer bestimmten Objektklasse nicht bestätigt werden und keine Erklärungstheorie (passendes Schema) für die perzipierten Phänomene im Gedächtnis gefunden werden kann (Assimilation also nicht möglich ist), wird ein neues Objektschema angelegt (*Accommodation*). Externe Perzepte werden auf die Sensoren (*space of sensors*) abgebildet, aus denen eine Repräsentation der Umgebung eines Agenten erzeugt wird (*local perceptual space*). Die wahrgenommenen Umgebungsveränderungen werden im Kurzzeitgedächtnis des Agenten gespeichert.

Emotionale Steuerung Das Modul *Emotion Regulation* ist für die Steuerung der Emotionen eines MicroPSI-Agenten zuständig. Dieses Modul kalkuliert die emotionalen Parameter aus Bedürfnissen, Relevanzsignalen und Werten der vorangegangenen Zustände. Der emotionale Zustand des MicroPSI-Agenten wird, entsprechend der PSI-Theorie, durch die Modulationsparameter Aktiviertheit, Selektionsschwelle, Auflösungsgrad, Bestimmtheit und Kompetenz festgelegt (vgl. Kapitel 7.2.3).

Diese Parameter finden sich an den Ausgängen (*gates*) des Moduls wieder. Alle Subsysteme des Agenten, die durch Emotionen gesteuert werden, sind mit diesen Ausgängen verlinkt,

so dass sie die aktuellen Werte der emotionalen Parameter mittels Ausbreitungsfunktion jederzeit ablesen können.

Diese emotionalen Parameter beeinflussen das Verhalten des MicroPSI-Agenten in vielerlei Hinsicht. So spielt bei der Perzeption insbesondere der Modulationsparameter „Auflösungsgrad“ eine wichtige Rolle: Wenn dieser niedrig ist, kann eine zu überprüfende Hypothese als wahr angenommen werden, ohne dass jedes ihrer Elemente geprüft wird. Ein niedriger Auflösungsgrad führt zu einer schnelleren, aber weniger exakten Perzeption, während bei hoher Auflösung eine langsame, aber präzisere Wahrnehmung erfolgt.

Motivation: Ein MicroPSI-Agent besitzt dieselben vorgegebenen Grundbedürfnisse wie ein PSI-Agent (Unversehrtheit, Energie, Affiliation, Kompetenz und Bestimmtheit). Diese Bedürfnisse führen zu einem Bedarf-Zustand, wenn sie nicht befriedigt werden, und dienen so als Quellen von *Motivationen*. Ereignisse, die einen Bedarf auslösen bzw. verstärken, werden als negative Signale interpretiert, während die Befriedigung eines Bedarfs positive Signale erzeugt. Die Ausführung der internen Aktivitäten und die Auswertung der Unbestimmtheit der extern wahrnehmbaren Ereignisse erzeugen einen Feedback für die Modulationsparameter und die kognitiven Bedürfnisse eines Agenten.

Aktionen und Planen: MicroPSI-Agenten repräsentieren Aktionen als Knotentripel, in denen der erste Knoten die Situationselemente referenziert, die die Vorbedingungen einer Aktion formen, der zweite einen Aktuator, der zu Veränderung in der Umwelt führt, und der dritte die Nachbedingungen der Aktion darstellt. Die Aktuatoren können auf weitere Aktionsketten (*macros* oder *scripts*) verweisen, so dass lange Aktionspläne ermöglicht werden. Da alle internen Aktionen (wie Perzeption, Zielidentifikation, Planen, Meta management usw.) auch als Knotenkette (NodeNets) formuliert und von dem Agenten ausgewertet und reorganisiert werden können, verfügt dadurch der Agent über die Fähigkeit, eigene Strategien zu prüfen und zu revidieren.

Das Planen erfolgt bei MicroPSI-Agenten genauso wie bei ihren PSI-Prototypen: Nachdem eine Zielsituation aus dem motivationalen Prozess abgeleitet wurde, sucht der Agent in seinem Speicher nach einer Aktionskette, die von der aktuellen Situation zur Zielsituation führt. Wenn kein solches Handlungsschema gefunden werden kann, fängt der Agent an, eine Aktionskette zu konstruieren, die zum Ziel führt. Dazu werden verschiedene Suchalgorithmen eingesetzt (Bottom-up, Top-down, A* usw.), wobei Suchtiefe und -breite durch die Modulationsparameter (s.o.) kontrolliert werden.

Metamanagement: Obwohl es keine zentrale Kontrollstruktur in der MicroPSI-Architektur gibt, bekommen verschiedene Prozesse, die das interne Verhalten und den Speicherzugriff steuern, je nach Situation eine verschiedene Ressourcenkapazität zugewiesen. Dies geschieht, indem Prozesse mit unterschiedlicher Frequenz aufgerufen werden oder Algorithmen benutzt werden, die unterschiedliche Zeit- bzw. Speicherplatzkomplexität besitzen. Damit können verschiedene Abstufungen der Reaktivität innerhalb eines Agenten realisiert werden: Kognitive Prozesse, deren erfolgreiche Ausführung mehr Aufmerksamkeit benötigt, werden entsprechend unterstützt und zwar auf Kosten von Prozessen, die auch ohne Unterstützung

7. Architekturen emotionaler Agenten

problemlos laufen. Die Auswertung der Performance von solchen Prozessen ist Aufgabe des Metamanagements.

Alarms: Zusätzlich zum Metamanagement implementieren die MicroPSI-Agenten einen Kontrollmechanismus, um schnell auf plötzliche Veränderungen in der Umwelt reagieren zu können: die Alarms. Dörner hat ein sogenanntes *absicherndes Verhalten* (*securing behavior*) vorgeschlagen, das von einem Agenten regelmäßig ausgeführt wird, um die Umwelt zu scannen.

In MicroPSI-Architektur erfolgt ein *Orientierungsverhalten* (*orientation behavior*), falls schnelle Veränderungen in der niederen Perzeption oder Bedürfniserkennung auftreten.

Langzeit- und Arbeitsgedächtnis: Der wichtige Unterschied der MicroPSI zu Dörners Modell ist die Aufteilung des Gedächtnisses in Langzeit- und Arbeitsspeicher. Diese Modifikation ermöglicht schnelle Datenrepräsentation und -manipulation und die Generierung und Auswertung von neuen Hypothesen, ohne dass die bereits existierenden, langfristig gespeicherten Konzepte aus dem Langzeitspeicher vorzeitig gelöscht bzw. modifiziert werden müssen. In dem Arbeitsspeicher werden die aktuelle Situation, laufende Ereignisse, kontextuelle Hintergründe der Objekte, die in der aktuellen Situation eine Rolle spielen, sowie die derzeit aktiven Ziele und Pläne gespeichert. Der Langzeitspeicher enthält Informationen über individuelle Objekte, Objektkategorien, die Agentenbiografie in Form von Gedächtnisprotokollen (*protocol memory*), eine Bibliothek der Pläne und Plänetile sowie die Abbildung der Agentenumwelt.

In Abb. 7.19 wird der Zusammenhang der beschriebenen Elemente der MicroPSI-Architektur als Petrinetz repräsentiert.

Der MicroPSI-Agent besitzt zwei Arten von Sensoren. Zum einen kann er über externe Sensoren (Stelle *PerceptSensor*) Informationen über die Veränderung der Umgebung wahrnehmen (*Transition Perception*). Die wahrgenommenen Umgebungsveränderungen werden dann im Kurzzeitgedächtnis des Agenten gespeichert (Stelle *Short Term Memory*). Zum anderen werden über interne Sensoren (Stelle *UrgeSensor*) somatische Signale über den internen Zustand des Agenten, insbesondere den Stand seiner Grundbedürfnis-Pegel (Kompetenz, Bestimmtheit, Energie, Wasser, Unversehrtheit) abgeleitet (*Transition Body Parameters*). Daraus wird der aktuelle Bedarf (*urge*) ermittelt, der, wie bei einem PSI-Modell, einen entsprechenden Motivator (*Transition Motivation*) aktiviert.

Die Aktivitäten eines MicroPSI-Agenten sind durch ein Repertoire von internen und externen Verhaltensweisen definiert (Stellen *Internal Behaviors* und *External Behaviors*). Mit internen Verhaltensweisen sind "mentale" Aktionen gemeint, die höhere kognitive Prozesse eines Agenten repräsentieren. Zu den *Internal Behaviors* zählen im MicroPSI-Modell insbesondere die Zielgenerierung und -verifizierung, die Wahl der aktuellen Absicht, das Planen und die Handlungswahl. Diese Prozesse werden durch Motivatoren ausgelöst (*Transition Motivation*) und durch eine Menge von Modulatoren (Aktiviertheit, Auflösungsgrad und Selektionsschwelle; vgl. die Modulationsparameter im PSI-Modell Abb. 7.15) gelenkt und gesteuert. Die Ausführung der durch die genannten internen Verarbeitungsprozesse generierten (ggf. "mentalen") Aktionen erfolgt über die *Transition Plan / Behavior / Script*

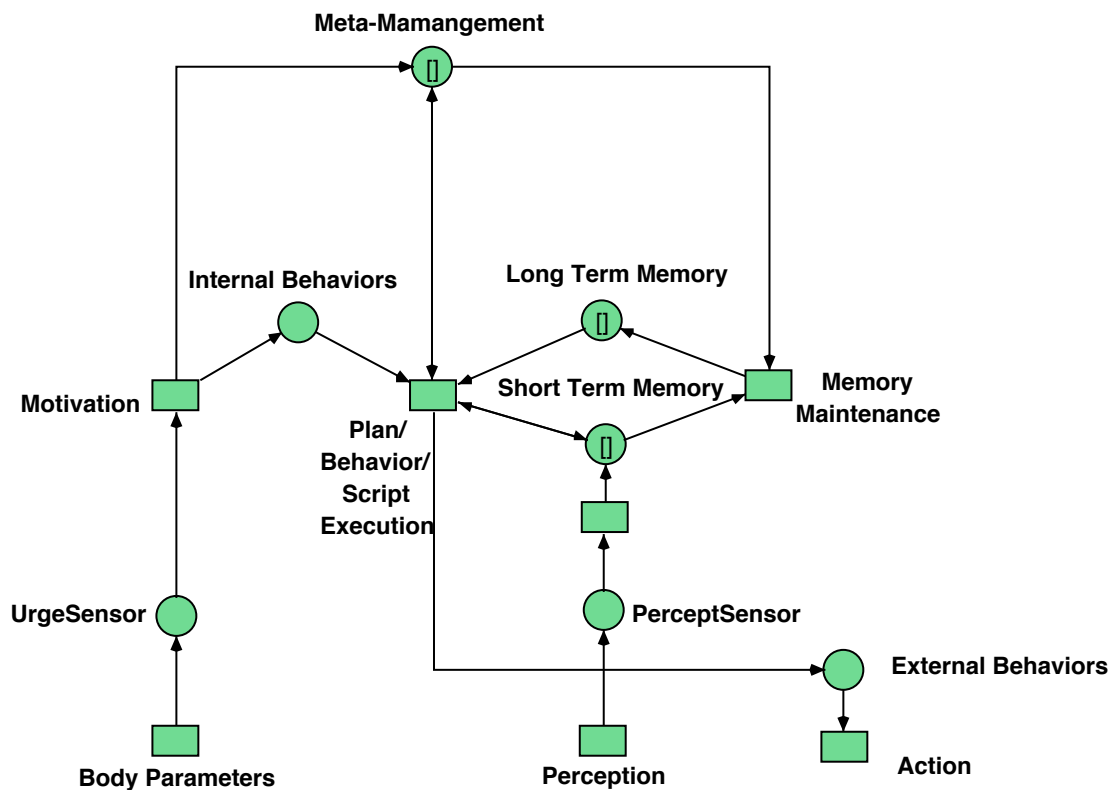


Abbildung 7.19.: Architekturübersicht eines MicroPSI-Agenten als Petrinetz.

Execution. Bei der Ausführung sind ggf. Kurzzeit- und Langzeitgedächtnis eines Agenten (Stellen **Short Term Memory** und **Long Term Memory**) sowie Meta-Managementprozesse (Stelle **Meta-Mamangement**) beteiligt.

Externe Verhaltensweisen sind Aktionen bzw. Aktionssequenzen (*macros* oder *scripts*), die von den Aktuatoren des Agenten direkt ausgeführt werden können (Transition **Action**).

Die Repräsentationen, die aus den externen Perceptionsergebnissen abgeleitet werden, werden zusammen mit dem zuvor akquirierten Wissen im Kurzzeitgedächtnis des Agenten gespeichert (Stelle **Short Term Memory**), wie auch die Informationen über die aktuelle Situation, die Ereignisse sowie aktive Ziele und Pläne. Im Langzeitgedächtnis (Stelle **Long Term Memory**) werden Interaktionssequenzen mit der Umgebung (Gedächtnisprotokolle), Objektkategorien, Pläne und Plänetteile gespeichert, die durch Ausführung der internen Verarbeitungsprozesse (*internal behaviors*) gewonnen werden (Transition **Plan / Behavior / Script Execution**).

Der Informationsaustausch zwischen Kurzzeitgedächtnis und Langzeitgedächtnis findet über eine Reihe von autonomen Prozessen (Transition **Memory Maintenance**) statt, welche auch den Zerfall von Informationen (memory decay), die Generierung neuer Konzepte usw. steuern. Die Verteilung der Verarbeitungsressourcen zwischen den internen Verarbeitungs-

7. Architekturen emotionaler Agenten

prozessen (Transition Execution) und Gedächtnissteuerungsprozessen (Transition Memory Maintenance) erfolgt über die Metamanagementprozesse (Stelle Meta-Management).

Bach sieht die sehr beschränkten Fähigkeiten zum sozialen Verhalten als einen bedeutenden Mangel des MicroPSI-Modells: “Shortcomings of the agent include a very limited potential for social behavior, mainly because means for the assessment of social situations (such as representations about mental states of other agents) are missing, and MicroPsi does not possess urges that would result in dominant, accepting, socially possessive or altruistic behaviors“. ([Bac03], S. 18).

7.2.4. Sloman: CogAff als konzeptuelles Framework für Architekturen emotionaler intelligenter Systeme

Um die Arbeiten von Aaron Sloman richtig einordnen zu können, muss man sie im Kontext seines (erkenntnis-)theoretischen Ansatzes sehen, der nicht die Emotionsmodellierung an sich, sondern die Erforschung und Modellierung natürlicher bzw. die Konstruktion künstlicher intelligenter Systeme in den Mittelpunkt stellt, wobei Modelle der Emotion einen wichtigen Teil der Gesamtarchitektur bilden.

Zunächst werden in diesem Abschnitt Slomans designbasierter Ansatz zur Entwicklung von KI-Systemen kurz skizziert, seine Konzeption von Emotion in einem intelligenten System erläutert und seine Emotionstaxonomie vorgestellt. Anschließend wird Slomans Drei-Schichten-Architektur beschrieben und mit Petrinetzen illustriert.

Der designbasierte Ansatz

Sloman bezeichnet seinen Ansatz zur Entwicklung von KI-Systemen als designbasiert. Dabei wird versucht, das vom System produzierte Verhalten auf das Systemdesign zurückzuführen und dadurch zu begründen ([WSB96], [Slo03]). Durch die Analyse des Systemverhaltens versucht man auf die internen Prozesse und Mechanismen zu schließen, die hinter den vorgeführten Phänomenen (z.B. Emotionalität, intelligentes Verhalten) stecken, um so das Systemdesign zu rekonstruieren: “[Design-based approach] construes AI as a methodology for exploring an abstract space of possible requirement for functioning agents (*niche space*) and a space of possible designs for such agents (*design space*) and the mapping between them“ ([WSB96], S. 105).

Sloman sieht natürliche intelligente Systeme als informationsverarbeitende Architekturen, die sowohl das interne als auch das externe Verhalten des Systems bestimmen. Der designbasierte Ansatz soll eine methodologische Grundlage schaffen, um die ausgewählten Aspekte dieser natürlichen Architekturen auf die künstlichen Systemarchitekturen zu transferieren. Der Schwerpunkt liegt dabei jedoch nicht auf dem Nachbau der Funktionalität der natürlichen Systeme, sondern auf der möglichst präzisen Nachbildung der Gesamtarchitektur, einschließlich der jeweiligen Teilsysteme mit ihren gegenseitigen Wechselwirkungen: “We need to understand global design for *complete systems*, including their functional decomposition into coexisting interacting subsystems“ ([WSB96], S. 106). Die Fähigkeiten eines Agenten werden maßgeblich durch die Agentenarchitektur bestimmt: “We claim that

architecture dominates mechanism ([Slo93]), i.e. global design normally determines global capabilities to a greater extent than implementation details.“ ([WSB96], S. 105).

Slomans Anspruch geht über den Entwurf der Architektur eines einzelnen intelligenten Systems - etwa der Simulation eines natürlichen tierischen oder menschlichen Verhaltens - weit hinaus. Sein Interesse liegt in der Entwicklung einer generellen Architektur, in der sowohl natürliche als auch künstliche und sowohl existierende als auch prinzipiell mögliche intelligente Systeme repräsentiert und analysiert werden könnten: “The design-based approach is a methodology for investigating mechanisms capable of generating mental phenomena, whether introspectively or externally observed, and whether they occur in humans, other animals or robots“ ([WSB96], S. 101). Sloman hält es deshalb für notwendig, bei der Untersuchung und Modellierung der natürlichen intelligenten Systeme unterschiedliche Ausprägungen dieser Systeme bei verschiedenen Spezies (Menschen und Tieren) zu berücksichtigen und individuelle Differenzen (Kinder, Erwachsene, Menschen mit Hirnschäden usw.) zu betrachten. Aus diesen Daten können dann prinzipielle Designkriterien für die Architekturen künstlicher intelligenter Systeme abgeleitet werden.

Eine Architektur für ein intelligentes System besteht nach Sloman im Allgemeinen aus den folgenden grundlegenden Komponenten: mehreren funktional unterschiedlichen Verarbeitungsschichten, Motivations- und Evaluierungsmechanismen, Filtern sowie einem globalen Alarmsystem. Diese Architekturkomponenten können bei natürlichen und künstlichen intelligenten Systemen auf unterschiedliche Weise zusammengesetzt sein, so dass sehr vielfältige Architekturmodelle vorliegen können. Im Kontext dieser unterschiedlichen Architekturen versucht Sloman die Emotionen jeweils entsprechend zu verorten und anhand ihrer Funktion innerhalb der jeweiligen Architektur zu interpretieren und zu klassifizieren (siehe Kapitel 7.2.4).

Die von Sloman postulierten Mechanismen knüpfen vielerorts an die evolutionistischen Emotionstheorien an. So entsprechen z.B. die phylogenetisch angeordneten Verarbeitungsschichten, denen entwicklungsgeschichtlich angeordnete Emotionen entspringen, weitestgehend der Sichtweise von Lazarus (vgl. Kapitel 6.1.2) und die Motivatoren bzw. Dispositionen sind den “concerns“ bzw. Handlungstendenzen in der Theorie von Frijda überaus ähnlich (ebd.). Die architekturbasierte funktionale Betrachtungsweise von Sloman nimmt jedoch eine Sonderstellung in der emotionstheoretischen Forschung ein.

Sein Ansatz ist mit Sicherheit einer der interessantesten bei der Entwicklung emotionaler Agenten, denn seine Betonung der Architektur sowie der theoretischen Tiefe und Breite der durchgeführten Arbeiten eröffnet dem Studium der Emotionen neue Perspektiven. Die Verbindung des design-orientierten Ansatzes, der Evolutionstheorie und der konzeptuellen Architekturklassifizierung ermöglicht eine tief greifende Betrachtung des natürlichen Emotionsphänomens, die bei den bisher vorgestellten Ansätzen zur Konstruktion autonomer Agenten nicht gegeben ist. Von einer globalen Sichtweise bei der Betrachtung der Emotionalität ausgehend, übt Sloman Kritik an undifferenzierten, konzeptuell und theoretisch schlecht belegten Herangehensweisen an die Emotionsmodellierung in der Informatik (siehe Kapitel 7.4.2).

7. Architekturen emotionaler Agenten

Emotionen und Motivatoren

Emotionen sind für Sloman keine eigenständigen Prozesse, sondern sie entstehen als emergentes Phänomen aus dem Zusammenspiel der unterschiedlichen Architekturkomponenten eines intelligenten Systems. Architekturen sind nach Sloman durch eine Menge von möglichen repräsentierbaren Zuständen, Prozessen und kausalen Interaktionen bestimmt. Verschiedene Architekturen weisen unterschiedliche solcher Mengen auf und besitzen deshalb verschiedene mentale Ontologien. Demnach können nicht die gleichen terminologischen Konzepte auf alle Architekturen angewandt werden.

Aus diesem Grund steht Sloman Versuchen, Emotionen explizit zu definieren, um aufgrund dieser Definition eine Emotionstheorie bzw. ein Emotionsmodell zu entwickeln, kritisch gegenüber. Das Phänomen der Emotion ist seiner Meinung nach zu komplex, um mit den einschlägigen Definitionen oder womöglich auch dem Alltagsverständnis gründlich erfasst zu werden. Viele Emotionstheorien beziehen sich in ihrem Verständnis der Emotionen implizit auf sehr unterschiedliche Ausprägungen. Sloman kritisiert, dass die gesamte Theorie sowie die von ihr gestellten Forschungsfragen nicht präzise genug formuliert werden können, wenn die Konzepte, die theoretische Annahmen und Aussagen formulieren sollen, mehrdeutig sind: “For instance, if our concept of “emotion“ is ambiguous, then so also will many questions about emotions be ambiguous, e.g. how emotions evolved, what their functions are, which animals have them, which brain mechanisms produce them, what types there are, whether they are needed for intelligence, whether a foetus has them, etc.“ ([Slo04], S. 1).

Wenn man hingegen Emotionen im Kontext der Architektur betrachtet, die sie produziert, kann man laut Sloman diese Widersprüchlichkeit umgehen. Konzepte wie “Emotionen“, “Bewusstsein“ usw. müssen nach Sloman in verschiedenen Architekturen unterschiedlich - und zwar architekturabhängig - interpretiert werden. Eine Theorie, die Emotionen im Zusammenhang mit Architekturkonzepten analysiert, ist daher in Bezug auf die Emotionsmodellierung besser geeignet und den rein theoretischen Ansätzen vorzuziehen.

Unter Emotionen versteht Sloman also ein Cluster-Konzept, das sich aus einer Menge von Fähigkeiten und Eigenschaften zusammensetzt, die nicht präzise definiert, sondern lediglich durch die Wechselwirkung von verschiedenen Komponenten der Gesamtarchitektur eines intelligenten (Software-)System beschrieben werden können: “Emotions involve complex processes produced by interactions between motives, beliefs, percepts, etc. E.g. real or imagined fulfilment or violation of a motive, or triggering of a “motive-generator“, can disturb processes produced by other motives. To understand emotions, therefore, we need to understand motives and the types of processes they can produce. This leads to a study of the global architecture of a mind“ ([SC81], S. 1).

Nicht die Emotionen an sich stehen also für Sloman im Mittelpunkt der Betrachtung, sondern einerseits die Bedeutung der Gesamtarchitektur eines Systems bei der Generierung der Emotionen und andererseits der Einfluss der Emotionen auf das Gesamtsystem. Wie bei vielen anderen KI-Ansätzen, die in diesem Abschnitt vorgestellt wurden, wird in dem Ansatz von Sloman der Schwerpunkt auf die funktionale Rolle der Emotionen gelegt.

Emotionstaxonomie Die Emotionstaxonomie Slomans unterteilt Emotionen in drei Kategorien: *primäre, sekundäre und tertiäre Emotionen* (z.B. [Slo03], S. 85f.). Diese drei Emoti-

onskategorien korrespondieren mit den drei von Sloman postulierten Architekturschichten: der reaktiven, der deliberativen und der Metamanagement-Schicht (siehe Kapitel 7.2.4)). Dabei können Emotionen sowohl durch interne Prozesse innerhalb jeweils einer dieser Schichten entstehen als auch durch Interaktionen zwischen den Schichten. Des Weiteren kann die Ausprägung der Emotionen aus der jeweiligen Kategorie je nach Systemarchitektur variieren.

- **Primäre Emotionen** beruhen auf der entwicklungsgeschichtlich ältesten und einfachsten reaktiven Schicht. Sie stellen globale Alarmsignale des Organismus auf interne und externe Bedrohungen dar und haben zum Ziel, das reaktive Verhalten und üblicherweise auch das perzeptive und das motorische Subsystem schnell zu mobilisieren und umzuorientieren. Primäre Emotionen in der Definition von Sloman entsprechen in etwa den Reflexen in der Definition von Lazarus (vgl. Kapitel 5.1) bzw. sie sind in der Definition von Damasio als niedere Anpassungsmechanismen (“Reflexe“ sowie “Triebe und Motivationen“) zu klassifizieren, deren Zusammenwirken eine Hintergrundemotion ergibt (vgl. Kapitel 5.5). Beispiele für primäre Emotionen sind Schreck, Furcht, sexuelle Erregung, Hunger usw. Sie können in einfacheren Organismen Verhaltensweisen wie Flucht, Angriff usw. produzieren. In komplexeren Systemen führen sie eher zur Handlungsbereitschaft anstatt zu einer direkten Handlung.
- **Sekundäre Emotionen** entstehen auf der deliberativen Ebene durch die Abschätzung der möglichen Folgen einer Situation oder eines Ereignisses, durch sogenannte “What-if“-Abschätzungen. Sie können während des Planens von und Reflektierens über Ereignisse aus der Vergangenheit oder beim Denken auftreten. Zu den sekundären Emotionen zählt Sloman verschiedene Arten der Nervosität, Erleichterung, Angst, Freude usw. Die Effekte sekundärer Emotionen lassen sich in vielen Verhaltensweisen und internen Veränderungen verfolgen, von der erhöhten Aufmerksamkeit und Vorsicht bei Angst bis zur Unterbrechung und Umorientierung der Metamanagement-Prozesse. Sekundäre Emotionen können sich auf die gleiche Weise manifestieren wie primäre Emotionen.
- **Tertiäre Emotionen**, auf die Sloman sich in seinen älteren Arbeiten als *Perturbances* bezogen hat ([Slo01], S. 16), werden mit einer Metamanagement-Schicht in Beziehung gesetzt. Eine Perturbanz tritt immer dann ein, wenn ein zurückgewiesener, vertagter oder einfach unerwünschter Motivator wiederholt auftaucht und damit das Management anderer, wichtigerer Ziele verhindert oder erschwert.

Perturbanzen sind nach Sloman keine eigenständigen Emotionen, sondern typische Begleiterscheinungen von Zuständen, die gemeinhin als emotional bezeichnet werden. Sie sind mit einem partiellen Kontrollverlust verbunden und können sich als Gefühl der Überwältigung in Freude, Trauer oder Schuldgefühl manifestieren. Damasio nennt solche Emotionen soziale Emotionen und zählt dazu, mit Sloman übereinstimmend, Scham, Schuld, Erniedrigung, Aspekte von Trauer, Stolz und Ärger.

7. Architekturen emotionaler Agenten

Später hat Sloman die Sonderstellung der Perturbanzen innerhalb seiner allgemeinen Emotionsklassifizierung wieder aufgehoben und sie auf eine Teilmenge der tertiären Emotionen reduziert ([Slo03], S. 86).

Motivationale Mechanismen Unter *Motivatoren* versteht Sloman eine spezielle Klasse von Informationsstrukturen, die interne und externe Aktionen veranlassen und beeinflussen. Diese Klasse umfasst Beliefs, Desires, Intentions und Wünsche (wishes). Motivatoren können unterschiedlich komplexe Prozesse auslösen, darunter Evaluierung und Priorisierung, Aktionswahl, Planung und Ausführung von Plänen usw. Ziel dieser Prozesse ist die Erreichung eines angestrebten bzw. die Vermeidung eines unerwünschten (Umwelt-)Zustandes. Die Menge dieser Zustände wird als *Motive* bezeichnet.

Zwischen Motiven und Emotionen bestehen grundlegende Wechselwirkungen: Das Erreichen eines Motivs bzw. das endgültige Ausscheiden eines Motivs aus der Menge der erreichbaren Zustände kann eine Emotion auslösen. Durch Emotionen wiederum können neue Motive entstehen. Darüber hinaus wird die Auswahl von Motiven durch Emotionen beeinflusst. Nach Wright besteht darin die Hauptfunktion der Emotion [Wri94].

Motivatoren sind nach Sloman ein zentraler Bestandteil eines jeden intelligenten Systems. Veränderungen im internen Körperzustand oder in der perzipierten Umweltsituation produzieren neue Ziele, die neues reaktives Verhalten auslösen, wenn zum Ziel führende Verhaltensprimitiven im Verhaltensrepertoire des Systems vorhanden sind. Anderenfalls werden deliberative Prozesse aktiviert, welche einen oder mehrere Pläne ausarbeiten, die zum Ziel führen. Dies impliziert jedoch die Notwendigkeit eines Mechanismus zur Produktion von Zielen, den Sloman Motivgenerator (*motive generator*) nennt ([Slo03], S. 67).

Es gibt nach Sloman viele verschiedene Arten von Zielen oder Motiven: Sie können kurzfristig, langfristig oder sogar permanent sein; sie können von internen physiologischen Veränderungen, externen Perzeptionsergebnissen, deliberativen oder Metamanagement-Prozessen ausgelöst werden. Deshalb wird eine Menge von unterschiedlichen Motivgeneratoren benötigt, die in verschiedene Komponenten des Gesamtsystems integriert werden. Sie können asynchron und parallel zu anderen Prozessen aktiviert werden, weil während der Ausführung von ganz anderen Aufgaben oft alternative Motivationen auftreten können (z.B. kann man beim Lesen eines Buches plötzlich Hunger bekommen oder neugierig auf die Geräusche von draußen werden).

Als Folge der asynchronen Generation können Motive in Konflikt miteinander stehen und es werden dann Motivkomparatoren (*motive comparators*) benötigt, um diese Konflikte zu identifizieren und zu lösen. Diese Motivkomparatoren entscheiden, welche Motive in verschiedenen Kontexten für wichtiger zu erachten sind, wobei spezifische Motivkomparatoren auf der reaktiven Ebene arbeiten, andere innerhalb der Problemlösungs- und Planungsmechanismen (z.B. indem sie Subziele vergleichen) und wieder andere als Teile des Metamanagements, wo sie bestimmen, welche Oberziele bei Konflikten gewählt werden.

Einige Organismen haben starre, eingeborene Motivgeneratoren und -komparatoren, während andere neue Strategien der Zielgenerierung und des Zielvergleichs erlernen können.

Evaluationsmechanismen Um die verschiedenen Motive sinnvoll miteinander vergleichen zu können, bedarf es eines Systems von Evaluationsmechanismen (*evaluators*). Die Evaluation erfolgt anhand der neurobiologischen Konzepte “Lust“ und “Schmerz“: Sloman definiert den Lustzustand funktional als einen Zustand, in dem die aktuelle Situation oder Aktivität positiv ausgewertet wird und eine Disposition produziert wird, diese Situation oder Aktivität aufrechtzuerhalten und den Faktoren entgegenzuwirken, die diese zu unterbinden oder abzuschwächen drohen ([Slo03], S. 70). Entsprechend ist Schmerz mit negativer Auswertung verbunden und produziert die Disposition, den schmerzlichen Zustand zu beenden oder abzumildern. Die Auswertungseffekte bestehen lediglich in der Tendenz, zu agieren (vgl. Frijdas Handlungstendenzen); in vielen Kontexten werden diese Handlungsdispositionen durch andere Faktoren dominiert, die für die Ziele des Systems für wichtiger erachtet werden, als z.B. nur die momentane Unlust zu vermeiden. Die Auswertung kann auch unabhängig vom Vergleich mit den Zielen und Motiven erfolgen, z.B. wenn ein Zustand als ästhetisch angenehm empfunden wird.

Drei-Schichten-Architektur

Sloman definiert eine hybride Architektur mit drei Verarbeitungsschichten (*Layers*), von denen jede zu einem bestimmten Grad als separates intelligentes System funktionieren kann:

- eine reaktive Schicht, die automatische und fest verdrahtete Prozesse einschließt,
- eine deliberative Schicht zum Planen, Evaluieren, Ressourcen zuteilen usw. sowie
- eine Metamanagement-Schicht, die Überwachungs- und Evaluationsmechanismen beinhaltet.

Die unterschiedlichen Schichten der Architektur produzieren unterschiedliche Emotionen. Ein nur aus einer reaktiven Schicht bestehendes intelligente(s) System(-architektur) wird im Folgenden *reaktives System* genannt, eine aus einer reaktiven und einer deliberativen Schicht bestehende Architektur ein *deliberatives System*.

Reaktive Schicht Ein rein reaktives System antwortet auf die Umweltreize mit fest einprogrammierten Reaktionen (siehe Abb. 7.20). Diese Schicht ist verantwortlich für die Generierung der Emotionen, die durch spezifische Reize ausgelöst werden, wie Ekel, sexuelle Erregung oder Erschrecken vor großen, sich schnell nähernden Objekten (vgl. Reflexe und physiologische Triebe, Kapitel 5.1). Grundsätzlich kann die Architektur des reaktiven Schichtes mit dem allgemeinen Modell des reaktiven Agenten nach [Röl04] abgebildet werden (vgl. Kapitel 3.1). Ein entsprechendes Petrinetz-Modell findet sich in Abb. 7.21.

Obwohl prinzipiell jedes beliebige Problem reaktiv lösbar ist, erweist es sich in der Praxis oft als unmöglich, nur mit starren Reiz-Reaktion-Verbindungen auszukommen. Das reaktive System ist deshalb optimal für eine begrenzte Anzahl spezieller Aufgaben geeignet, da die reaktive Verarbeitung besonders gute Zeitkomplexitätseigenschaften hat und die Reaktionsgeschwindigkeit wesentlich höher als ist bei zwei anderen Verarbeitungsschichten.

7. Architekturen emotionaler Agenten

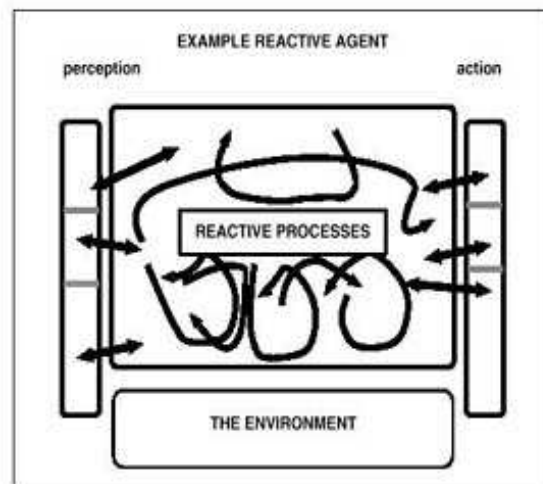


Abbildung 7.20.: Reaktive Architektur (aus [Slo01], S. 14).

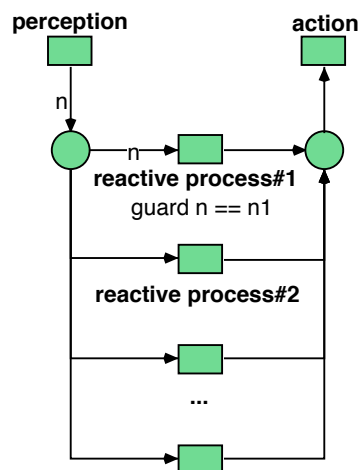


Abbildung 7.21.: Petrinetz-Modell einer reaktiven Architektur nach Sloman (vgl. auch Abb. 3.1).

Dafür ist das Fehlen jeglicher Flexibilität in einem rein reaktiven System offensichtlich. Da weder Planen noch Schlussfolgerungsmechanismen auf der reaktiven Verarbeitungsebene vorgesehen sind, kann es ausschließlich nur die festgelegten Aufgaben und Strukturen verarbeiten.

Deliberative Schicht Eine zweite, phylogenetisch jüngere Schicht, die deliberative Schicht (siehe Abb. 7.22), implementiert komplexere Mechanismen wie Planen, Entscheiden, Schlussfolgern usw. und ist vor allem durch die Fähigkeit zum sogenannten “what-if“-Reasoning charakterisiert. Darunter versteht Sloman das Schließen über die Ketten der möglichen

zukünftigen Handlungen, um Pläne zu bilden, sowie die damit eng verwandte Fähigkeit zum Schließen über die Menge der möglichen Handlungen in der Vergangenheit, um Erklärungen für die vergangenen Erfolge und Misserfolge zu bilden. Das Schließen erfolgt über die Verarbeitung der kontrafaktischen inneren Zustände, die mögliche Umweltereignisse und Systemreaktionen repräsentieren. Durch die Evaluation dieser inneren Zustände werden die möglichen Handlungsergebnisse antizipiert.

Ein deliberatives System kann sein Handlungsrepertoire auf diese Weise beliebig neu kombinieren, Pläne entwickeln und vor der Ausführung evaluieren. Für die Implementierung der deliberativen Prozesse werden weitere Mechanismen benötigt, u.a. eine *Kurzzeitgedächtniskomponente*, mit deren Hilfe die Pläne aus den einzelnen Aktionen zusammengestellt werden. Dazu gehört die Speicherung der einzelnen Aktionen eines Plans und deren erwartete Konsequenzen, der Vergleich der einzelnen Planschritte und Planteile untereinander und schließlich die Erstellung des gesamten Planes. Die erstellten und die noch nicht komplettierten Pläne werden in der *Langzeitgedächtniskomponente* gespeichert. Hier werden auch die wahrscheinlichen Konsequenzen von Plänen abgelegt und später evaluiert (vgl. Abb. 7.22).

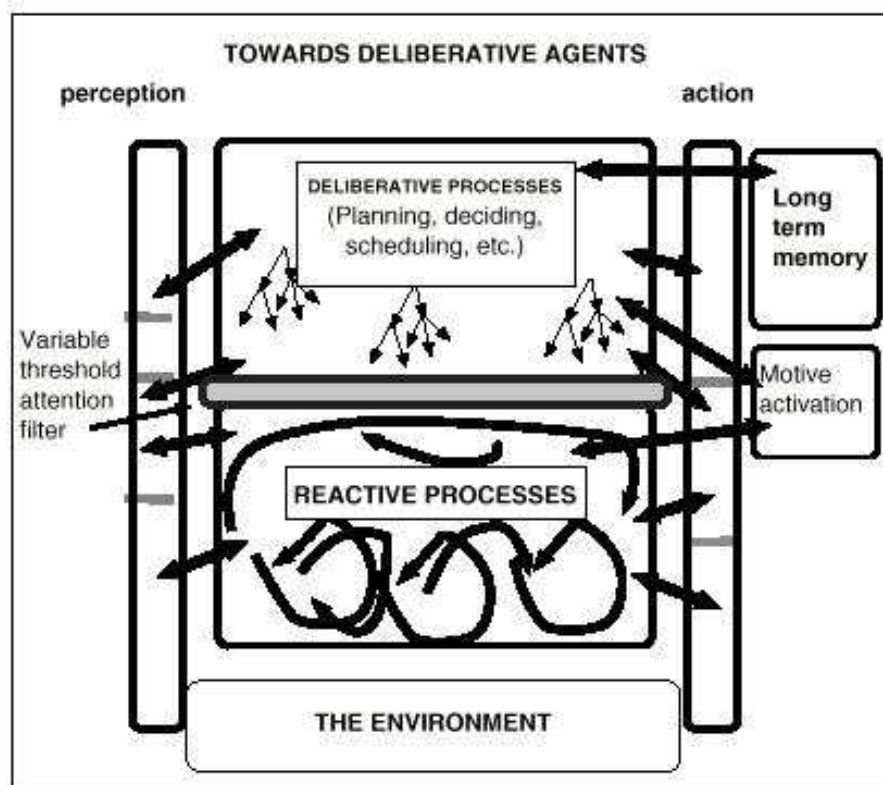


Abbildung 7.22.: Hybride Architektur mit deliberativer Schicht, Langzeitgedächtniskomponente und motivationalem System (H-CogAff) (aus [Slo01], S. 15).

7. Architekturen emotionaler Agenten

Deliberative Mechanismen sind im Allgemeinen langsamer als reaktive, da viele Prozesse in der deliberativen Schicht wegen der Notwendigkeit des Zugriffs auf den überschreibbaren Kurzzeitspeicher und wegen der Anwendung derselben assoziativen Langzeitspeicherkomponente seriell und diskret arbeiten. Die serielle Arbeitsweise bietet aber auch eine Reihe von Vorteilen: Dem System ist jederzeit klar, welcher Plan zu einem Erfolg geführt hat, die gleichzeitige Ausführung sich widersprechender Pläne wird verhindert und die Kommunikation mit dem Speicher verläuft weitgehend fehlerfrei.

Eine hybride Agentenarchitektur mit einer reaktiven und einer deliberativen Architekturschicht erhält man durch die Zusammenführung des Modells eines reaktiven Agenten nach [Röl04] (siehe Abb. 3.1) mit dem allgemeinen Modell einer hybriden Architektur mit horizontaler Schichtung (siehe Abb. 3.3).

Ein entsprechendes Petrinetz-Modell findet sich in Abb. 7.23. Die Perzeptionsergebnisse werden über die Transition **perception** zeitgleich an die reaktive und die deliberative Architekturschicht übertragen. Die Schichten arbeiten unabhängig voneinander wie eigenständige Agenten. Die Unabhängigkeit der Verarbeitung auf verschiedenen Ebenen wird durch die Petrinetzsemantik der nebenläufig zueinander ausführbaren Transitionen **deliberative processes** bzw. **reactive processes** angedeutet. Bei der Verarbeitung auf der deliberativen Schicht können die erstellten Pläne in der Langzeitgedächtniskomponente, welche durch die Stelle **long term memory** modelliert wurde, abgelegt und wieder abgerufen werden. Aus den Ausgaben der Schichten wird ein tatsächlich auszuführender Plan ausgewählt (Transition **plan selection**) und zur Ausführung gebracht (Transition **action**).

Die deliberative Schicht ist verantwortlich für die Generierung der sekundären Emotionen. Dazu gehören emotionale Phänomene wie Frustration durch Versagen, Erleichterung durch Gefahrenvermeidung, Angst vor Fehlschlägen oder angenehme Überraschung durch Erfolg. Sekundäre Emotionen entstehen durch kognitive Auswertung der Ereignisse bzw. deren Konsequenzen im Hinblick auf Ziele und Motive des Systems und entsprechen damit dem traditionellen einschätzungstheoretischen Emotionskonzept. Für ihre Implementierung sind deliberative Prozesse unbedingt erforderlich, deshalb sind sie der deliberativen Architekturschicht zuzuschreiben.

Metamanagement Die phylogenetisch jüngste Schicht des Gesamtsystems ist das sogenannte Metamanagement. Hierbei handelt es sich um einen Mechanismus, welcher die internen Vorgänge in allen Schichten des Systems verfolgt und evaluiert (siehe Abb. 7.24). Ein solcher Mechanismus ist notwendig, um die in der deliberativen Schicht entwickelten Pläne und Strategien in Hinblick auf Schleifen, schlechte Performanz usw. zu überprüfen und ggf. zu verwerfen, wiederkehrende Muster im deliberativen Subsystem zu erkennen, zu kategorisieren und auszuwerten bzw. langfristige Strategien zu entwickeln und effektiv mit anderen Schichten zu kommunizieren.

Abb. 7.25 Petrinetz-Modell einer hybriden Architektur mit der reaktiven, der deliberativen und der Metamanagement-Schicht. Die Verarbeitung auf der Metamanagement-Ebene kann sowohl durch Perzeption neuer Umweltinformationen (Transition **perception**) als auch durch die Verarbeitungsprozesse der deliberativen bzw. der reaktiven Ebene initiiert werden (Transitionen **deliberative processes** bzw. **reactive process**). Die Ergebnisse der Verarbeitung

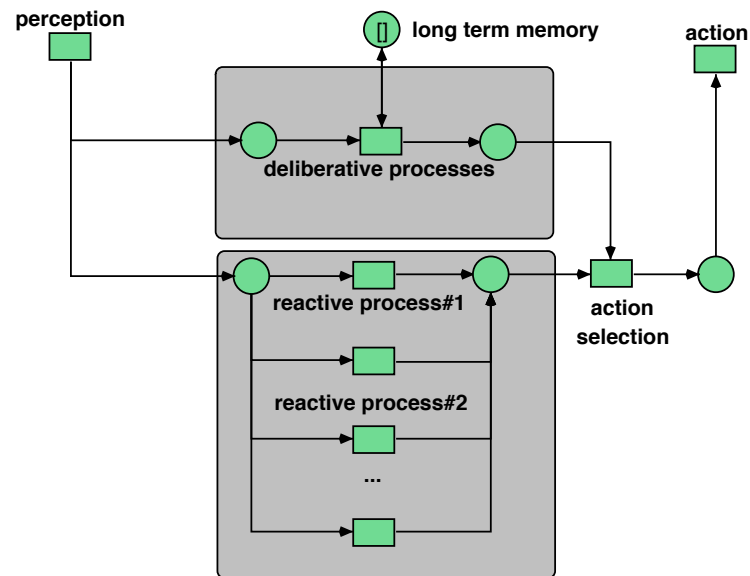


Abbildung 7.23.: Petrinetz-Modell einer hybriden Architektur mit reaktiver und deliberativer Schicht nach Sloman.

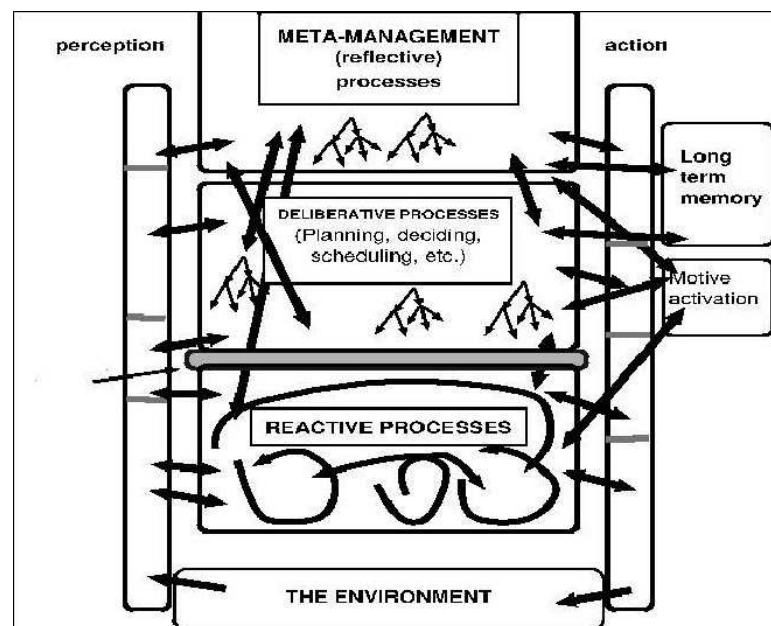


Abbildung 7.24.: Drei-Schichten-Architektur H-CogAff mit Metamanagement (aus [Slo01], S. 16).

auf der Metamanagement-Ebene stehen den Prozessen anderer Ebenen über die Stelle *evaluation results* zur Verfügung.

Die Metamanagement-Schicht hat jedoch keinen umfassenden Zugriff auf alle internen

7. Architekturen emotionaler Agenten

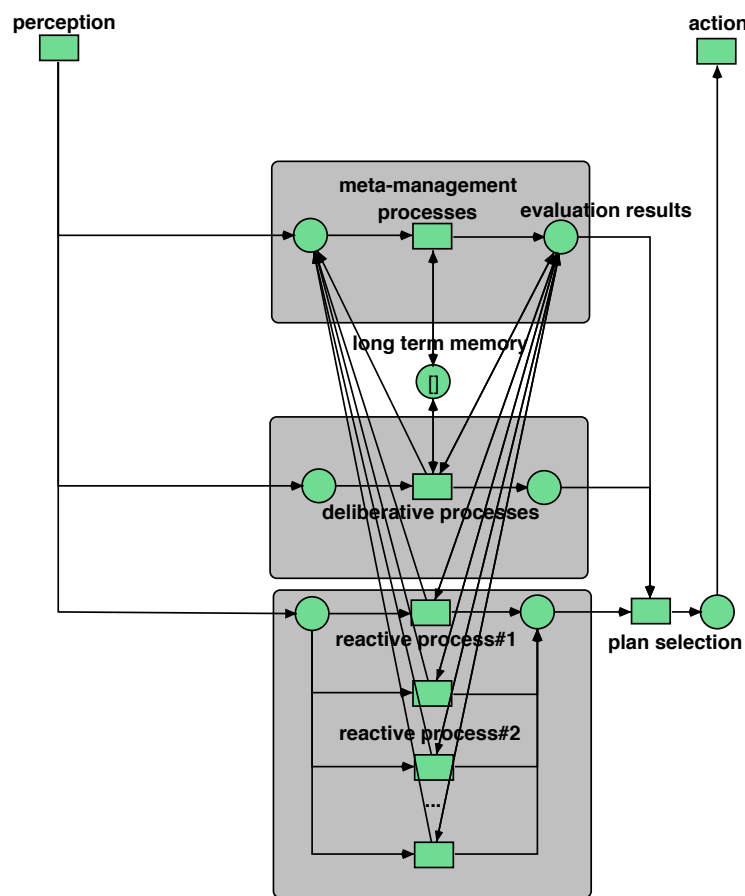


Abbildung 7.25.: Petrinetz-Modell einer hybriden Architektur mit reaktiver, deliberativer und Meta-Management-Schicht nach Sloman.

Zustände und Prozesse. Die Kontrolle über das deliberative Subsystem ist niemals vollständig und wird manchmal sogar abgeschaltet, daher können die Selbst-Evaluationen auf falschen Voraussetzungen beruhen. Durch die Unvollkommenheit der Kontrolle der Metaebene entstehen tertiäre Emotionen, die als Episoden partiellen Kontrollverlusts definiert werden und das Management anderer, wichtigerer Ziele verhindern können. Zu den tertiären Emotionen gehören Scham, Erniedrigung, Aspekte von Trauer, Ärger, Stolz usw. Nach der Definition von Damasio werden diese Emotionen zu den sozialen Emotionen gezählt (vgl. Kapitel 5.5).

Aufmerksamkeitsfilter Die verschiedenen Motivatoren konkurrieren miteinander um die begrenzten Ressourcen des deliberativen Subsystems. Die neu generierten Motive, Ziele, Alarmsignale oder wichtige perzipierte Informationen aus der reaktiven Schicht gelangen parallel zueinander in die deliberative Schicht, um dann in einer festzulegenden Reihenfolge sequenziell verarbeitet zu werden. Diese Signale können jedoch destruktiv wirken, wenn sie

z.B. eine gerade ablaufende Bearbeitung eines wichtigen Ziels, Plans oder einer Aktivität unterbrechen, die ungeteilte und ununterbrochene Aufmerksamkeit verlangt, um erfolgreich abgeschlossen zu werden.

Um die Ablenkung durch neue Signale und häufige Aufmerksamkeitswechsel zu reduzieren und dem deliberativen Subsystem effektives Funktionieren zu ermöglichen, wurde ein Mechanismus erschaffen, der verhindert, dass neue Motivatoren jederzeit die Aufmerksamkeit auf sich ziehen können. Zu diesem Zweck wurde ein Aufmerksamkeitsfilter mit dynamischen Schwellenwerten entworfen, der entscheidet, welche neuen Ziele, Alarmsignale oder andere potenziellen Interrupts die Verarbeitung in der deliberativen Schicht unterbrechen dürfen.

Der Filter legt einen Schwellenwert fest, den ein Motivator übertreffen muss, um Aufmerksamkeitsressourcen für sich beanspruchen zu können. Die Filterschwellenwerte sind variabel und können z.B. bei der Verarbeitung einer wichtigen oder anspruchsvollen Aufgabe sehr hoch sein, während sie niedrig und leicht zu überbieten sind, wenn die aktuelle Aufgabe nicht besonders dringend oder wichtig ist. Prinzipiell darf ein Filter ein Interrupt nie völlig ausschließen, da eine Unterbrechung durch eine lebensnotwendige Information immer möglich sein muss. Ein Filter kann sich u.a. durch Lernen ändern, wenn z.B. eine motorische Fähigkeit automatisiert wird und nicht mehr die volle Aufmerksamkeit benötigt.

Der Aufmerksamkeitsfilter wurde durch die Stelle **attention filter** in Abb. 7.26 modelliert, welche die Eingaben des Perzeptionssystems (Transition **perception**) und des motivationalen Systems (Transition **motive activation**) empfängt, filtert und an die deliberative Verarbeitungsschicht (Transition **deliberative processes**) weitergibt.

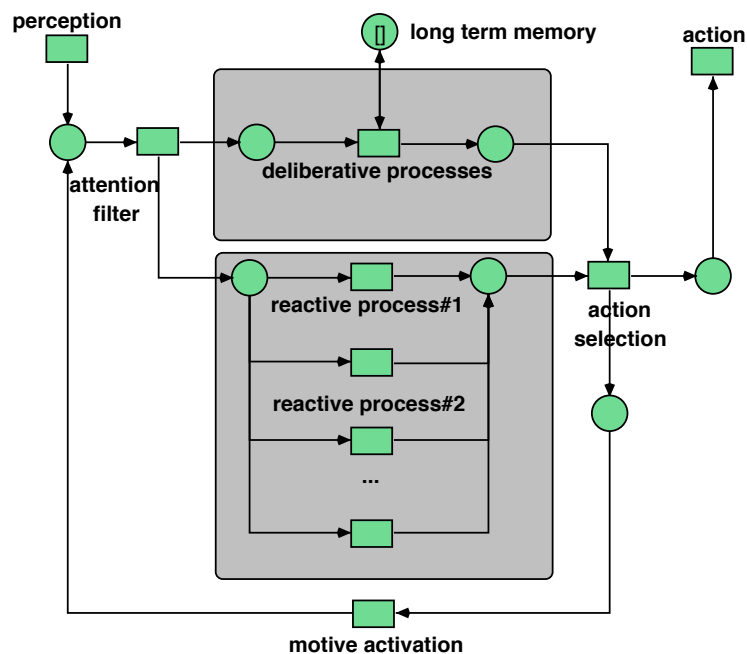


Abbildung 7.26.: Petrinetz-Modell einer hybriden Architektur mit reaktiver und deliberativer Schicht und Aufmerksamkeitsfilter.

7. Architekturen emotionaler Agenten

Die *Insistenz* eines Motivators - die entscheidende Variable für das Überwinden des Filters - wird definiert in Relation zu der Fähigkeit (*power*) des Motivators, durch den Interrupt-Filter hindurchzudringen, und stellt einen schnell zu berechnenden heuristischen Wert der Wichtigkeit und Dringlichkeit des Motivators dar. Im Gegensatz dazu wird die *Intensität* eines Motivators definiert als seine Fähigkeit, aktiv zu bleiben, nachdem er schon für die Bearbeitung ausgewählt wurde.

“Interruption might occur if the new goal is judged more important than, and inconsistent with, the purpose of the current activity, or if the new one is judged to be very urgent (although not necessarily very important) whereas the (more important) current activity is not time-critical and can be temporarily suspended: for instance stopping for a meal because one has plenty of time before the important meeting. Alternatively a highly insistent motive that gets through the filters can be considered and then rejected as relatively unimportant, without interrupting any important current action. So insistence, the propensity to divert attention, is not the same as a propensity to interrupt current actions, except those that require full attention“ ([Slo92], S. 15).

Alarmsysteme Zum Überleben in einer dynamischen Umwelt muss das intelligente System in der Lage sein, schnell auf Umweltveränderungen zu reagieren. Insbesondere aufgrund der Notwendigkeit von schnellen Verhaltensreaktionen sind deliberative Systeme wegen ihrer vielen langsamen seriellen Prozesse den Anforderungen der dynamischen Umwelt oft nicht gewachsen. Auch reaktive Mechanismen können unter Umständen langsam arbeiten, z.B. wenn zwischen sensorischem Input und motorischem Output eine lange Kette von sequenziellen internen Reaktionen liegt. In einigen Situationen ist aber jegliche Reaktionsverzögerung fatal oder zumindest nicht optimal für das Überleben des Organismus, z.B. wenn dadurch eine Möglichkeit zur Triebbefriedigung versäumt oder auf Bedrohungen nicht zeitig reagiert wird. Deshalb hält Sloman die Implementierung von *Alarmmechanismen* (auch in rein reaktiven Systemarchitekturen) für nötig, die schnell die Notwendigkeit zu sofortigem Verhaltenswechsel erkennen und durch Kontrollsignale an andere Subsysteme weiterleiten.

Das Alarmsystem bekommt Input von allen wichtigen Architekturkomponenten und kann ihnen seine Kontrollsignale senden. Es kann verschiedene globale Aktionen auslösen (wie Flucht, Angriff, sich verkriechen, einer Gefahr ausweichen, Wechsel der Bewegungsgeschwindigkeit oder -richtung usw.). Es kann auch Reaktionen in Form von Aufmerksamkeitsveränderungen oder Veränderungen der kognitiven und physiologischen Zustände generieren (vgl. Handlungsbereitschaften nach [Fri86] bzw. [Fri07]; siehe auch Kapitel 6.1.2).

Die unterschiedlichen Schichten des Systems können verschiedene Alarmer auslösen und werden von diesen auf unterschiedliche Art und Weise beeinflusst (vgl. Abb. 7.27). So kann beispielsweise beim Vorausplanen auf der deliberativen Ebene eine Gefahr oder eine Chance identifiziert werden, auf die umgehend mit einem Alarmsignal zum Wechsel der Strategie reagiert werden muss. Auf der reaktiven Ebene löst üblicherweise ein bestimmter sensorischer Input einen Alarm aus, dem mit einer bestimmten internen oder externen Reaktion entgegnet wird. Diese verschiedenen Alarmarten werden durch eine virtuelle Stelle **alarms** modelliert.

Diese kann sowohl durch die Verarbeitungsprozesse der drei Ebenen (Transitionen **meta-management processes**, **deliberative processes**, **reactive processes**) als auch durch Informationen aus der Umwelt, welche den Schwellenwert des Aufmerksamkeitsfilter passieren konnten (Transition **attention filter**), aktiviert werden und dann die Verarbeitungsprozesse auf allen drei Verarbeitungsebenen beeinflussen.

Ferner wurden in Abb. 7.27 die drei verschiedenen Emotionsklassen modelliert (Stellen **primary emotions**, **secondary emotions** und **tertiary emotions**), welche jeweils verschiedenen Architekturschichten zugeordnet sind und zusammen mit den Ergebnissen der kognitiven Verarbeitung und Alarmen auf der jeweiligen Schicht letztendlich die Aktionswahl (Transition **plan selection**) beeinflussen. Neben der Aktionswahl bestimmen diese drei Parameter auch die Wahl der Motivatoren (Transition **motive activation**), entsprechend Slomans Vorstellung, dass die Auswahl von Motiven durch Emotionen beeinflusst wird (vgl. Kapitel 7.2.4). Abb. 7.27 bietet einen Gesamtüberblick über eine H-CogAff-Architektur mit reaktiver, deliberativer und Metamanagement-Schicht.

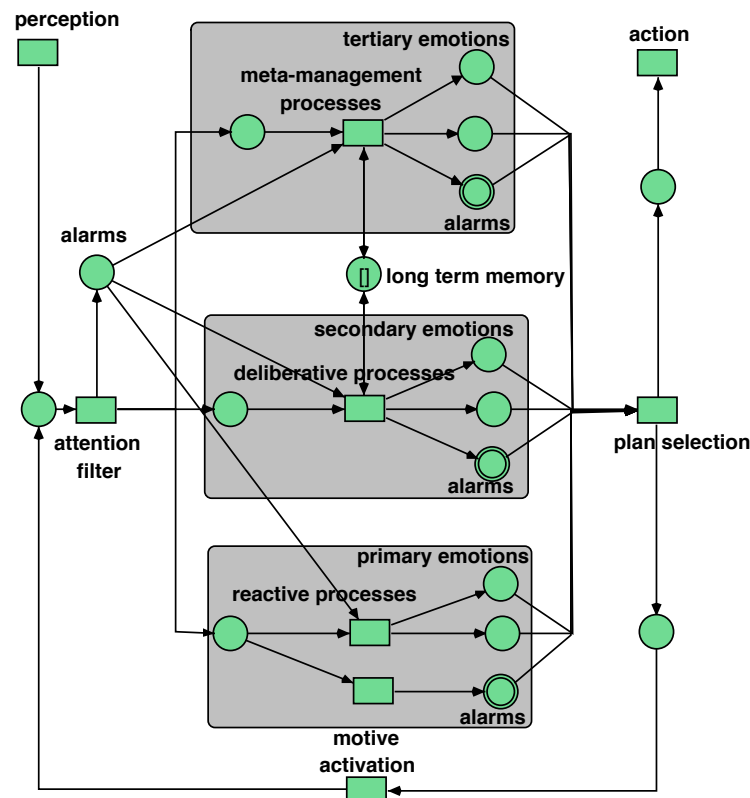


Abbildung 7.27.: Petrinetz-Modell einer hybriden Architektur mit reaktiver, deliberativer und Meta-Management- Schicht und einem Alarmsystem.

7.3. VKI

In diesem Abschnitt werden Modelle vorgestellt, die Emotion in einen Kontext zur VKI-Forschung setzen (vgl. zu grundlegenden Motivationen bei der Emotionsimplementierung im VKI-Kontext Kapitel 2.2). Insbesondere werden wir die Implementierung der Emotion in den zwei bekanntesten emotionalen Multiagentenarchitekturen - von Elliot und von Bates und Reilly - untersuchen. Außerdem wird der Ansatz von Aubé und Senteni vorgestellt, in dem die Emotionen zur Ressourcensteuerung und Agentenkommunikation in einem MAS eingesetzt werden. Nach einer Übersichtsdarstellung der Emotionsmodelle diskutieren wir die vorgestellten Architekturen und analysieren insbesondere ihre emotionstheoretische Fundierung. Die entsprechenden psychologischen und soziologischen emotionstheoretischen Grundlagen wurden in Kapitel 6 eingeführt.

7.3.1. Elliott: Affective Reasoner

Das Modell von Elliott stellt bis dato eine der umfassendsten Implementierungen einer Emotionstheorie in einem Multiagentensystem dar. Das Grundgerüst des Systems wurde in [Ell92] beschrieben und später um die Emotionsintensität [EG93] und einige weitere Komponenten ergänzt [EO92].

Das Modell von Elliott [Ell92] ist eine breit angelegte Plattform zur Erforschung verschiedener Aspekte der Emotionalität. Das Forschungsinteresse gilt dabei besonders der Rolle der Emotionen in sozialen Interaktionen. Das Modell gründet auf der Einschätzungstheorie von Ortony et al. ([OCC88], siehe Kapitel 6.1.2). Der Modellierungsanspruch konzentriert sich auf die Ab- bzw. Nachbildung von sozialemotionalen Interaktionen zwischen

- (1) Agenten und ihrer Umgebung,
- (2) Agenten und anderen Agenten sowie
- (3) Agenten und Benutzern.

Die funktionale Rolle der Agentenemotionen innerhalb des MAS, etwa zur Koordination oder Strukturierung der Agentengesellschaft, wird im Affective-Reasoner-Modell (AR) nicht thematisiert. Die primäre Zielsetzung des Modells ist vielmehr die Generierung, Darstellung (Ausdruck) und Kommunikation von Emotionen zwischen Agenten.

Als Domäne wählt Elliott die Welt der Taxifahrer (TaxiWorld) mit ihren interpersonellen Beziehungen. Das emotionale Leben der Agenten in dieser simulierten Welt basiert auf 24 Emotionsklassen und ungefähr 1400 verschiedenen Ausdrucksmöglichkeiten für Emotionen. Wenn eine Emotion generiert wird, haben Agenten ein Spektrum von ungefähr 440 Reaktionsmustern (*emotion manifestation categories*) zur Verfügung, um diese Emotion auszudrücken (etwa 20 Kanäle für jede Emotion).

Die Reaktionsmuster sind entlang eines Kontinuums von spontanen (*not intentional*) zu geplanten (*highly intentional*) Reaktionen gegliedert. Zum Beispiel kann ein Agent seinen Ärger somatisch ausdrücken, indem er "errötet" (nicht intentional) oder kognitiv, indem er einen Racheplan erstellt. Dazwischen findet man Reaktionsmuster wie "verbale Ausdrücke" (Agent teilt seine Emotion verbal mit), "Emotionsunterdrückung" (Agent tut so, als wäre nichts passiert) usw. Die Reaktionskanäle werden in Abhängigkeit von den Temperamenten

der Agenten (s.u.) aktiviert. Zusätzlich können die Agenten über Emotionen “nachsinnen“ und Emotionen der anderen Agenten mit eigenen Emotionen erwidern.

Die Fähigkeit seines Systems, Emotionen korrekt darzustellen, wurde in einem Experiment mit 141 Versuchspersonen empirisch überprüft [Ell96]. Es hat sich gezeigt, dass die Versuchspersonen die dargestellte Emotion aus der mimischen, sprachlichen und musikalischen Emotionsrepräsentation des Affective Reasoner signifikant besser ableiten konnten als aus der Darstellung eines speziell trainierten Schauspielers (70 zu 53 Prozent).

AR-Architektur Für seine Zielsetzung, in einer simulierten Umgebung virtuelle Agenten an emotionalen Episoden teilnehmen zu lassen, identifiziert Elliott folgende notwendige Systemkomponenten:

“For this to be useful we must have

1. a simulated world which is rich enough to test the many subtle variations a treatment of emotion reasoning requires,
2. agents capable of (a) a wide range of affective states, (b) an interesting array of interpretations of situations leading to those states and (c) a reasonable set of reactions to those states,
3. a way to capture a theory of emotions, and
4. a way for agents to interact and to reason about the affective states of one another.

The Affective Reasoner supports these requirements“ ([Ell92], S. 2).

Im Folgenden werden wir auf das funktionale Grundgerüst der AR etwas genauer eingehen. Jeder Agent hat Zugriff auf folgende Datenbanken ([EO92], S. 810):

1. Eine BDI-Datenbank, wobei die BDIs die *concern structure* eines Agenten festlegen und zugleich seine *interpretive personality component* (s.u.) definieren.
2. Eine Datenbank mit 24 Emotionstypen, die im Wesentlichen den Emotionstypen nach [OCC88] entsprechen und von Elliott um die beiden Typen *love* und *hate* erweitert wurden. Jedem dieser Emotionstypen sind spezielle *emotion eliciting conditions* (EEC) zugeordnet.
3. Eine COO-Datenbank (Concerns-of-Others), eine Datenbank der BDIs anderer Agenten. Da es sich hier um vom Agenten gelernte Daten handelt, sind sie meistens unvollkommen und können auch falsche Annahmen enthalten.
4. Eine Datenbank mit Reaktionsmustern, die, je nach Emotionstyp, in bis zu 20 unterschiedliche Gruppen (*temperament traits*) unterteilt sind.

7. Architekturen emotionaler Agenten

Persönlichkeit der AR-Agenten Die Agenten besitzen einzigartige Persönlichkeiten, die es ihnen ermöglichen, unterschiedliche Interpretationen der gleichen Situation auszubilden und unterschiedliche Reaktionsmuster aus dem gleichen emotionalen Zustand heraus zu entwickeln.

Um diese Ambivalenzen zu implementieren, untergliedert Elliott die Emotionsverarbeitung eines Agenten in zwei Teilprozesse ([EO92], S. 809f.):

- Die *strong-theory component* beschreibt den Weg von der Situation, die Emotion verursacht, zur “Entstehung“ der Emotion; sie basiert in ihrem Kern auf der OCC-Theorie (vgl. Kapitel 6.1.2).
- Die *weak-theory component* beschreibt den Teilprozess, der von der Emotion zur emotionalen (Verhaltens-)Reaktion führt. Diese Komponente besitzt keine emotions-theoretische Fundierung, da die OCC-Theorie lediglich Aussagen über den Emotions-generierungsprozess liefert und das Problem des funktionalen Einflusses der Emotion, etwa auf das Verhalten und die Handlungswahl, nicht betrachtet.

Die Emotionsverarbeitung nach Elliott wurde in Abb. 7.28 schematisch dargestellt. Hier sieht man, dass die von Elliott vorgeschlagene Aufteilung des Emotionsprozesses in eine Emotions- und eine Verhaltensgenerierungskomponente weitestgehend der funktionalen Emotionsdefinition (vgl. Abb. 5.2) und auch dem generellen Emotionsmodell (vgl. Abb. 5.13) entspricht, welches in Kapitel 5 als Referenzmodell für die Suche nach einer geeigneten Emotionsdefinition bzw. einem allgemeinen KI-Emotionsmodell vorgeschlagen wurden.

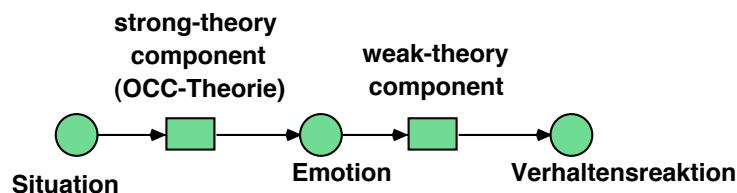


Abbildung 7.28.: Emotionsverarbeitung eines AR-Agenten nach [Ell92].

Um die Aufteilung des Emotionsprozesses in zwei Teilprozesse zu realisieren, wurde die rudimentäre Repräsentation der Agentenpersönlichkeit eingeführt, die in zwei Komponenten zerlegt werden kann:

- Die *interpretive personality component* stellt die individuelle Disposition eines Agenten dar, Situationen in seiner Welt auf bestimmte Art und Weise zu interpretieren. Der Agent stellt zuerst fest, ob ein Ereignis, eine Aktion oder ein Objekt einen Bezug zu seinen BDIs hat. Diese Überprüfung ist in der originalen OCC-Theorie nicht enthalten und wurde von Elliott offensichtlich aus den anderen Einschätzungstheorien übernommen (vgl. die Einschätzungsdimensionen “goal significance check“ bei Scherer in Kapitel 6.1.2, Bewertung der Zielrelevanz bei Lazarus in Kapitel 6.21 bzw. Relevanzeinschätzung bei Frijda in Kapitel 6.25). Anschließend wird die regelbasierte

Verarbeitung genutzt, um die Interpretation einer Situation durch den Agenten an die emotionalen Zustände zu binden.

- Die *manifestative personality component* bestimmt die individuelle Art des Agenten, seine Emotionen auszudrücken. Diese Komponente koppelt seine emotionalen Zustände mit seinen (emotionalen) Reaktionen; ihr liegt keine emotionstheoretische Fundierung zugrunde. Die Komponente enthält eine Menge der Temperamenteigenschaften (*temperament traits*), die dynamisch aktiviert werden, um das Profil der individuellen Verhaltenstendenzen des Agenten zu bestimmen.

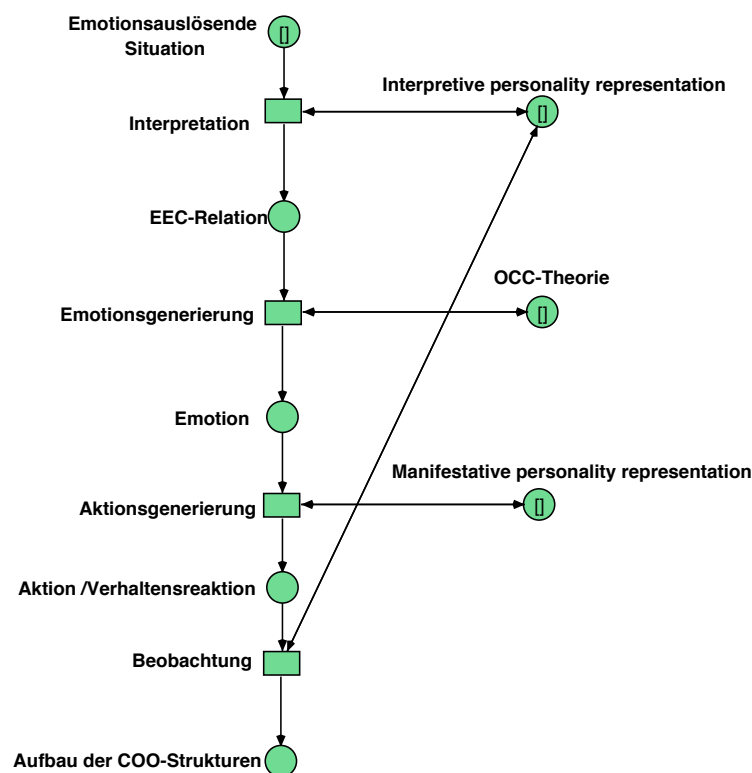


Abbildung 7.29.: Petrinetz-Modell des Emotionsprozesses eines AR-Agenten (nach [Ell92], S. 28).

Emotionsgenese Abb. 7.29 stellt die einzelnen Verarbeitungsschritte des Emotionsprozesses eines AR-Agenten zusammen mit den korrespondierenden Repräsentationsstrukturen als Petrinetz dar (vgl. [Ell92], S. 28).

Die Verarbeitung beginnt mit einer spezifischen Situation, die von dem Agenten so interpretiert wird, dass eine Emotionsgenerierung initiiert wird (Stelle *Emotionsauslösende Situation*). Die Situationsinterpretation (Transition *Interpretation*) beinhaltet die Überprüfung von deren Relevanz für die eigenen BDIs sowie den Vergleich bestimmter Situationseigenschaften

7. Architekturen emotionaler Agenten

mit den in den BDI- und COO-Datenbanken gespeicherten Mustern, den EECs (*emotion eliciting condition*, vgl. Kapitel 6.5).

Bei Korrespondenz wird eine EEC-Relation gebildet (Stelle EEC-Relation), die zwei oder mehr Werte in jeder Emotionsklasse einnehmen kann. EEC-Relationen setzen sich damit aus den Elementen der aktuellen Situation und deren Interpretation durch den Agenten zusammen, wobei die Art dieser Interpretation durch die “interpretive personality component“ (Stelle **Interpretive personality representation**) des Agenten bestimmt wird. Die Parameter der EEC-Relation entsprechen weitestgehend den Einschätzungsdimensionen der OCC-Theorie. Einige Parameter sind z.B. Erwünschtheit eines Ereignisses - *desire-self (event)*, Lobwürdigkeit einer Handlung - *evaluation (act)*, Attraktivität eines Objekts - *appealingness (object)* usw. (für eine komplette Auflistung siehe [Ell92], S. 37).

Wurden eine oder mehrere EEC-Relationen gebildet, werden diese benutzt, um entsprechend den Vorgaben der OCC-Theorie Emotionen zu generieren (Transition **Emotionsgenerierung**). In dieser Phase tritt eine Reihe von Problemen auf, die bereits in Kapitel 6.1.2 erläutert wurden, weil sie in der Theorie von Ortony et al. nicht ausreichend berücksichtigt werden.

So diskutiert Elliott [Ell92] das Problem der zusammengesetzten (compound) Emotionen (z.B. *anger*, *gratification*, *remorse* usw., vgl. Kapitel 6.7: Der Affective Reasoner konstruiert die EEC-Relationen für die beiden zugrunde liegenden Emotionen und fasst sie anschließend in einer neuen EEC-Relation zusammen. Die konstituierenden Emotionen werden also durch die zusammengesetzte Emotion ersetzt. Elliott betrachtet das nicht als optimale Lösung:

“Does anger really subsume distress? Do compound emotions always subsume their constituent emotions? That is, in feeling anger does a person also feel distress and reproach? This is a difficult question. Unfortunately, since we are implementing a platform that generates discrete instances of emotions, we cannot finesse this issue. Either they do or they do not. There can be no middle ground until the eliciting condition theory is extended, and the EEC relations are extended.“ ([Ell92], S. 42).

Solche Vorgehensweise funktioniert zwar noch bei qualitativ ähnlichen Emotionen (Elliott nennt als Beispiele *distress* und *anger*), bereitet aber spätestens bei mehreren gleichzeitig auftretenden Emotionen Probleme, insbesondere dann, wenn sie einander widersprechen. Elliott räumt ein, dass das OCC-Modell an dieser Stelle Lücken aufweist: “Except for the superficial treatment of conflicting expressions of emotions, the development and implementation of a theory of the expression of multiple emotions is beyond the scope of this work.“ ([Ell92], S. 44f.). In der AR-Architektur wird die “Lösung“ dieses Problems auf den Reaktionsgenerierungs-Modul verlagert (s.u.).

Generierung der Reaktionen Sobald ein emotionaler Zustand für einen Agenten generiert worden ist (Stelle **Emotion**), wird eine daraus resultierende Reaktion initiiert (Transition **Aktionsgenerierung**). Hierbei spielt die “manifestative personality component“ (Stelle **Manifestative personality representation**) eine entscheidende Rolle. Sie enthält ein *emotion manifestation lexicon*, das über drei Dimensionen verfügt: 24 Emotionstypen, etwa 20

Reaktionsmuster (*emotion manifestation categories*) sowie eine Intensitätshierarchie der möglichen Reaktionen (die in der ersten Version des AR nicht implementiert wurde). Die Reaktionsmuster des AR sind hierarchisch organisiert; jede Hierarchiestufe ist außerdem entlang eines Kontinuums von spontanen (nicht intentionalen) zu geplanten (hochintentionalen) Reaktionen gegliedert.

Für jeden Agenten können einzelne Kanäle vor Beginn der Simulation aktiviert bzw. deaktiviert werden. Dieses spezifische Muster aktiver und nicht aktiver Kategorien konstituiert die individuellen Dispositionen (*manifestative personality*) eines Agenten, auf bestimmte Art zu reagieren. Die aktivierten Kategorien können als individuelles Temperament (*temperamental traits*) eines Agenten aufgefasst werden.

Um Konflikte zwischen sich widersprechenden Emotionen und damit auch zwischen sich widersprechenden Handlungen zu vermeiden, enthält das Reaktionsmodul sogenannte *action exclusion sets*. Sie werden gebildet, indem die möglichen Reaktionen in Äquivalenzklassen eingeteilt werden. Ein Element einer dieser Klassen kann niemals gemeinsam mit einem Element einer anderen Klasse in der letztendlich bestimmten Reaktionsmenge auftauchen.

Die AR-Architekturkomponente, die für die Emotionsmanifestation verantwortlich ist (*manifestative personality component*), nimmt keinen expliziten theoretischen Bezug auf die sozialwissenschaftlichen Emotionstheorien und basiert eher auf intuitiven Konstrukten.

Concerns of Others (COO) Die AR-Agenten sind in der Lage, die Reaktionen der anderen Agenten auf bestimmte Situationen zu erklären und vorherzusagen. Ebenso können sie die Emotionen anderer Agenten “nachempfinden“. Dazu werden interne Modelle der Dispositionsstrukturen der anderen (*interpretive personality*) sowie ihre Reaktionsmuster (*manifestative personality*) in einer eigens dafür angelegten COO-Datenbank gespeichert. Die Ersteren, um die Emotionen der anderen situationsabhängig ableiten zu können, die Letzteren, um die zugrunde liegenden Emotionen aus der Reaktionsweise der anderen zu erschließen.

Ein Agent beobachtet die emotionalen Reaktionen eines anderen Agenten (Transition Beobachtung) und leitet daraus ein eigenes Modell seiner BDIs ab, das er in der COO-Datenbank abspeichert (Stelle Aufbau der COO-Strukturen). Wenn der Agent anstelle der eigenen BDIs nun die abgeleiteten BDIs des beobachteten Agenten einsetzt, kann er unter Anwendung desselben Mechanismus, mit dem er eigene Emotionen generiert, Schlüsse über die Emotionen bzw. emotionalen Reaktionen anderer Agenten ziehen.

Im Idealfall müsste demnach die COO-Datenbank die exakte Kopie der BDI-Datenbank des beobachteten Agenten sein. Weil es sich bei den COOs jedoch um eine von dem Agenten durch Lernen inkrementell aufgebaute Datenbank handelt, enthält sie auch unkomplettes, unsicheres und eventuell falsches Wissen.

Der Lernprozess, der dem Agenten erlaubt, über die Emotionen des anderen Agenten Schlussfolgerungen zu ziehen (Elliott / Ortony 1992, S. 813), wird mit Hilfe des Programms Protos (Bareiss 1989) implementiert. Dieser Lernprozess läuft folgendermaßen ab: Zunächst wird das beobachtete Reaktionsmuster mit einer Datenbank von emotionalen Reaktionen verglichen, um die zugrunde liegende Emotion zu identifizieren. Sodann wird das beobachtete Ereignis durch die COO-Datenbank für den beobachteten Agenten gefiltert, um festzustellen,

7. Architekturen emotionaler Agenten

ob diese Reaktion bereits darin enthalten ist. Ist dies der Fall, so kann davon ausgegangen werden, dass die Datenbank eine korrekte Repräsentation der emotionsauslösenden Situation enthält. Auf dieser Grundlage kann der beobachtende Agent dann eine Erklärung für das Verhalten des beobachteten Agenten entwickeln.

Sollte die Repräsentation in der COO-Datenbank mit dem beobachteten Verhalten nicht übereinstimmen, wird sie aus der Datenbank entfernt und die Datenbank wird erneut durchsucht. Sollte keine korrekte Repräsentation gefunden werden, kann der Agent auf Default-Werte zurückgreifen, die in die COO-Datenbank integriert werden. Es werden zwei Arten von Default-Werten definiert: Entweder können die Reaktionen eines typischen Agenten auf die beobachtete Situation oder die eigenen Reaktionen in einer entsprechenden Situation als Default-Reaktionen erfolgen. Zudem besteht die Möglichkeit, die COOs eines Dritten, d.h. eines beliebigen anderen Agenten einzusetzen, falls die COOs in der eigenen COO-Datenbank bereits vorhanden sind.

Hat der Agent nun eine COO-Datenbank für einen anderen Agenten aufgebaut, kann er sie für zwei Zwecke einsetzen. Zum einen kann ein Agent in Abhängigkeit von der Art der Beziehung zwischen den beiden Agenten aufgrund der beobachteten Erfolge oder Misserfolge eines anderen Agenten Emotionen generieren. Im Affective Reasoner sind drei verschiedene Beziehungstypen zwischen Agenten möglich: *friendship*, *animosity* und *empathetic unit*.

Bei dem freundschaftlichen Beziehungstyp wird der beobachtende Agent mit dem Beobachteten mitfühlen, bei der Rivalität wird er konträre Emotionen “verspüren“ (z.B. Schadenfreude, Neid), im Falle einer *empathetic unit* wird der beobachtende Agent die COOs des beobachteten Agenten anstelle seiner eigenen BDIs einsetzen, um eine vollkommene Empathie mit dem anderen zu erzeugen. Zum anderen ist es aufgrund der aufgebauten COO-Datenbank möglich, die emotionalen Reaktionen der anderen Agenten zu verstehen und vorherzusagen.

Ein Problem bei der Reaktionsinterpretation stellt jedoch die Tatsache dar, dass es, wenn überhaupt, eine nur sehr vage Abbildung der Reaktionen auf die zugrunde liegenden Emotionen gibt: Man lächelt, wenn man zufrieden ist, aber auch, um negative Emotionen zu vertuschen. Erschwerend hinzu kommen die individuell unterschiedlichen Arten, Emotionen auszudrücken, die wiederum durch die verschiedenen Launen der Individuen beeinflusst werden.

7.3.2. Das Oz-Projekt von Bates und Reilly

Das Modell von Bates und Reilly hat primär auf dem Gebiet der MCI breiten Anklang gefunden und wurde in weiteren Forschungsprojekten zur Gestaltung der Benutzerinteraktion wiederaufgegriffen (vgl. z.B. [MG03], [LB02]).

Bates et al. verfolgen mit ihrem *Oz Project* das Ziel, in einer simulierten interaktiven Welt agierende Charaktere so zu gestalten, dass ihr Verhalten für einen menschlichen Betrachter nachvollziehbar und überzeugend erscheint. In [Rei96] stellt Reilly die These auf, dass interaktive animierte Charaktere, für die er den Begriff *believable agents* einführt, weder intelligent noch realistisch sein müssen (im Sinne der realitätsgetreuen Nachbildung eines natürlichen Aussehens oder der Verhaltensweise von existierenden Lebewesen), um auf einen menschlichen Betrachter glaubhaft zu wirken. Vielmehr sollen sie eine “starke“

Persönlichkeit besitzen, die das Aussehen und alle Verhaltensweisen des Agenten bestimmt, von der Bewegungsart bis zum sprachlichen Ausdruck.

Bates hält die Darstellung der Emotion für einen wichtigen Mechanismus, um die Glaubhaftigkeit seiner interaktiven Charaktere zu steigern: “Emotion is one of the primary means to achieve this believability, this illusion of life, because it helps us know that characters really care about what happens in the world, that they truly have desires“ ([Bat94], S. 8).

Bates und Reilly realisieren die angestrebte Glaubhaftigkeit ihrer Charaktere durch eine besonders breit angelegte Agentenplattform, die über multiple “menschliche“ Fähigkeiten verfügt. Sie nennen diese Herangehensweise “*broad but shallow approach*“. Die Agenten verfügen über Fähigkeiten wie Perzeption, Reaktivität, zielbasiertes Verhalten, Emotionen, soziales Verhalten, Erkennung und Generierung natürlicher Sprache u.v.m. Doch werden diese Fähigkeiten nicht tiefer ausgebaut, als die Aufgaben der Agenten in der Domäne es verlangen. Die Autoren behaupten, dass die geringe Kompetenz der Agenten in den einzelnen Fähigkeitsbereichen keine Auswirkungen auf ihre Glaubhaftigkeit hat, weil dieser Mangel an Kompetenz durch die Breite ihres Fähigkeitsspektrums ausgeglichen wird [Rei96].

In diesem Zusammenhang führen die Autoren den Begriff *broad shallow agent* ein. Bei solchen Agenten werden viele I/O-Module implementiert, u.a. um die Emotionen auf viele verschiedene Weisen auszudrücken. Diese Fähigkeiten beschränken sich aber nur auf das in der Aufgabendomäne notwendige Niveau:

“The concept is simple: the agent will be emotionally richer if there are more things to have emotions about and more ways to express them. This reliance on breadth has also meant that I have been able to create simple emotion models that rely on perception and motivation instead of deep modeling of other agents and complex cognitive processing.“ ([Rei96], S. 3).

Die modellierte Welt, Oz-Welt genannt, besteht aus einer simulierten Umgebung, einer Menge von Agenten, die diese Umgebung bevölkern, einer Benutzerschnittstelle, durch die ein oder mehrere Benutzer am Geschehen in der Welt teilhaben können, und einem Planer, der sich mit der Benutzermodellierung befasst. Die Agenten in der Oz-Welt werden konfrontiert mit der Unmöglichkeit, ihre Welt vollständig zu modellieren, mit ungenauen und fehlerhaften Wahrnehmungen sowie mit der Notwendigkeit, in Echtzeit zu reagieren.

Später wurden noch Module zur Sprachanalyse und Spracherzeugung implementiert [Rei96]. Auf Basis des Modells von Bates und Reilly haben Mateas und Stern [MAS02] ein System zur Kodierung des Verhaltens und der Körpersprache für künstliche Charaktere (genannt ABL - A behaviour Language) entwickelt, das sie im interaktiven Drama eingesetzt haben.

Im Gegensatz zu Elliott strebt Reilly in seiner Architektur eine minimale Modellierung der BDIs und Emotionen anderer Agenten an, er behauptet jedoch, mit seinem Modell trotzdem komplexe soziale und emotionale Beziehungen wie Ablehnung oder Freundschaft nachbilden zu können: “I propose an approach to modeling other agents that calls for limiting the amount of modeling of other agents to that which is sufficient to create the desired behavior. Using this technique, I have been able to build robust social behaviors that use surprisingly little representation.“ ([Rei96], S. 3).

7. Architekturen emotionaler Agenten

Reilly weist ausdrücklich darauf hin, dass sein Emotionsmodell nicht den Anspruch erhebt, im psychologischen Sinne korrekt zu sein, sondern dass es lediglich einen Ausgangspunkt zur Erzeugung glaubwürdiger emotionaler Agenten darstelle.

Die Agentenarchitektur namens **Tok** besteht aus drei wesentlichen Komponenten:

- einem Modul mit dem Namen **Hap** für zielbasiertes Verhalten und Reaktivität
- einem Modul für Emotionen und soziale Beziehungen, das **Em**, und
- einem Perzeptionsmodul für die Sinneswahrnehmungen.

Abb. 7.30 zeigt die abstrakte Architektur eines Tok-Agenten.

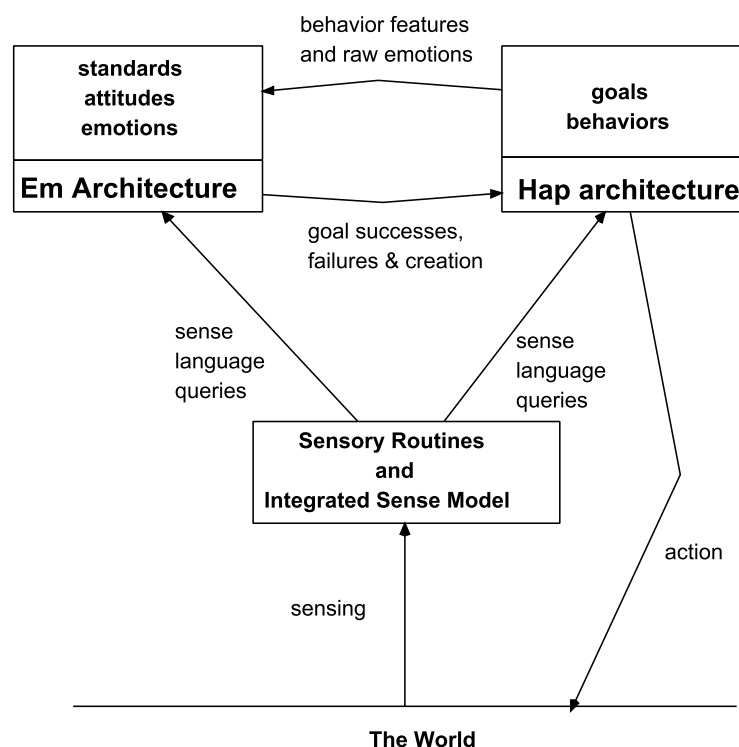


Abbildung 7.30.: Tok-Architektur (aus [RB92], S. 2)

Im Folgenden konzentrieren wir uns auf dessen emotionale Architekturkomponente, die **Em**. Die Architektur der **Em**-Komponente wird in Abb. 7.31 gezeigt (aus [Rei96], S. 45).

In Abb. 7.32 wird der Emotionsprozess nach Bates und Reilly als Petrinetz dargestellt.

Em bekommt den Input vom **Hap** (Transition *Input*) und entscheidet, ob emotional reagiert werden muss. Die *Emotion Generators box* enthält eine Menge der Produktionsregeln für Emotionsgenerierung, die den **Hap**-Input in eine Emotionsstruktur (Stelle **Emotion Structure**) transformiert (Transition **Emotionsgenerierung**).

Eine Emotionsstruktur setzt sich zusammen aus einem Emotionstyp (z.B. Angst), einer Intensität (z.B. 7 aus 10), möglicherweise einer Richtung der Emotion (z.B. Angst vor einem

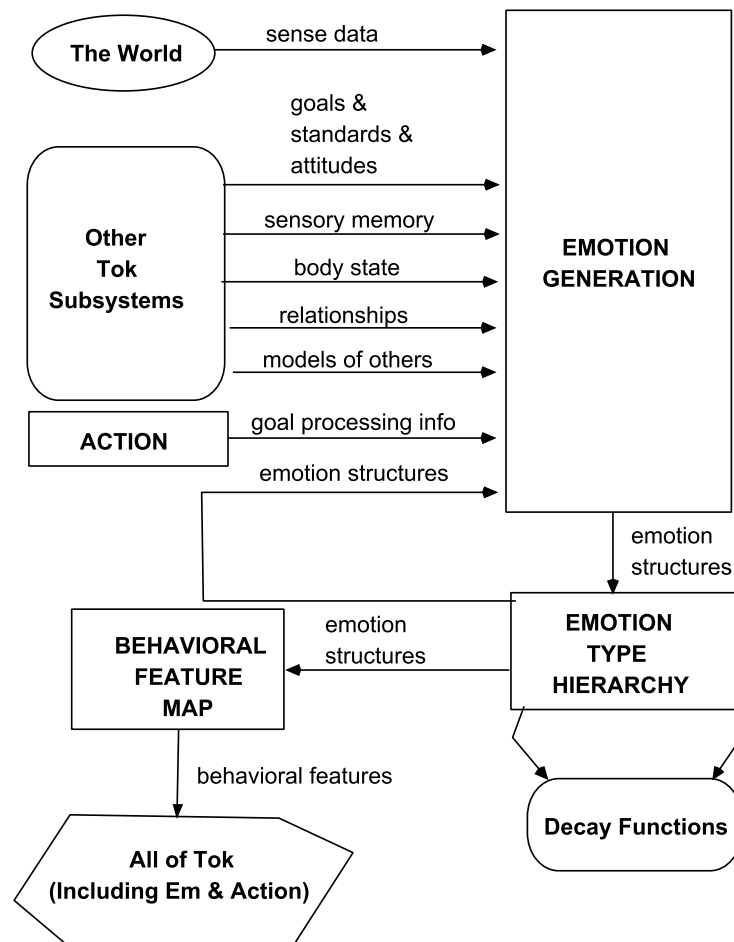


Abbildung 7.31.: Em-Architektur (aus [Rei96], S. 45).

Hund) und einen Grund der Emotion (z.B. der Hund hat den Agenten schon einmal gebissen). Die generierten Emotionsstrukturen werden in Form einer Emotionstypenhierarchie gespeichert (Transition Emotionsspeicherung), wobei die Emotionen mit den allgemeineren Auswirkungen auf den oberen Hierarchiestufen angesiedelt werden, während die Emotionen mit spezielleren Auswirkungen auf den unteren Stufen der Hierarchie zu finden sind. “For example, the hierarchy might have a distress type that represents general forms of distress expression, like frowning, crying, and moving slowly. Below that type might be subtypes, such as grief, homesickness, and lovesickness, which inherit the general effects of their common parent, but that also have more specific means of expression as well, such as thinking about home when homesick” ([Rei96], S. 32).

Die Em-Architektur bietet die Möglichkeit, die Ereignis-Emotion-Regeln für jeden Agenten neu zu definieren. Die Regeln definieren nicht nur, welche Emotion in Abhängigkeit vom Stimulus generiert werden muss, sondern legen auch die Emotionsintensität für die

7. Architekturen emotionaler Agenten

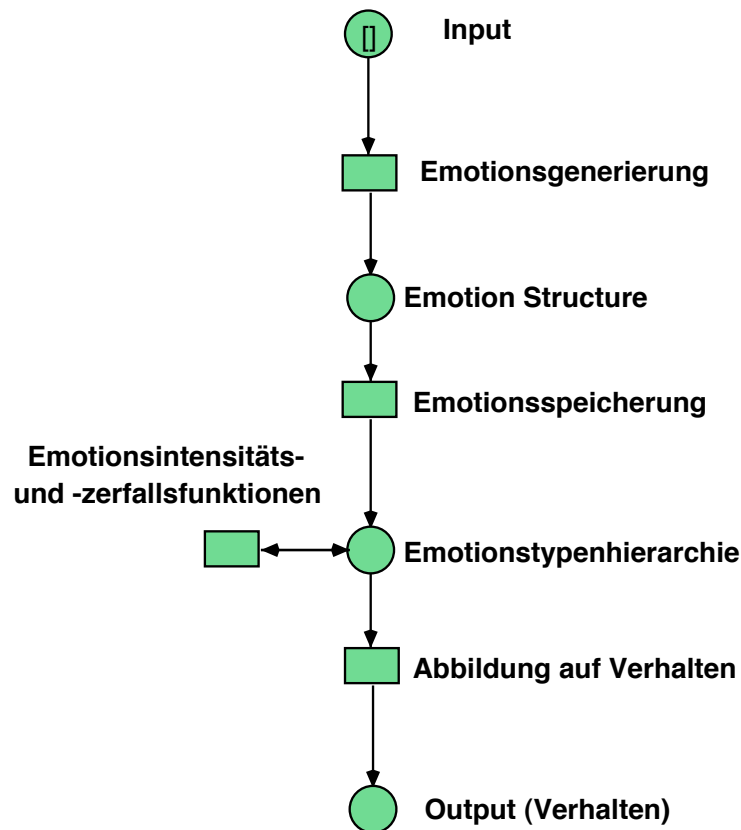


Abbildung 7.32.: Emotionsprozess nach Bates und Reilly als Petrinetz-Modell.

individuellen Ziele des Agenten in Abhängigkeit von der Bedeutung des Ereignisses fest (Transition *Emotionsintensitäts- und -zerfallsfunktionen*). “An example emotion generation rule is the following: when an agent has a goal failure and the goal has importance X, generate an emotion structure of type distress and with intensity X” ([Rei96], S. 32). Für jede Emotionsstruktur kann man ferner eine eigene *Emotion Decay Function* definieren. Diese Funktion bestimmt, wie schnell die jeweilige Emotionsstruktur mit der Zeit zerfällt: So kann z.B. Ärger wegen einer Beleidigung langsamer zerfallen als wegen der Verspätung eines Freundes.

Die Emotionsstrukturen werden mit einer *Behavioral Feature Map* auf entsprechende Verhaltensweisen abgebildet (Transition *Abbildung auf Verhalten*). Diese Abbildung ist im Modul **Hap** implementiert. Die aus den Emotionsstrukturen abgeleiteten Verhaltensweisen werden als **Em-Output** an die **Hap**-Komponente weitergegeben (Stelle *Output (Verhalten)*). Sie beeinflussen verschiedene Aspekte der Informationsverarbeitung des Agenten.

Em implementiert ein Emotionssystem, das auf dem Modell von Ortony et al. [OCC88] basiert: Die vordefinierten Emotionen werden durch Einschätzung der externen Ereignisse, Aktionen und Objekte bezüglich der vorgegebenen Einschätzungskriterien generiert.

Ähnlich wie Elliott unterteilen Bates und Reilly das OCC-Modell in zwei Komponenten: Sie unterscheiden zwischen Prozessen der Erkennung der Emotionen von anderen und emotionalen (Verhaltens-)Antworten. Eine Art der Emotionserkennung wurde nach der OCC-Modell umgesetzt, d.h. die Emotionen von anderen werden mit Hilfe der Modellierung der Ziele von anderen Agenten und eines Schlussfolgerungsmechanismus abgeleitet. Diese Art von Ableitung entspricht weitestgehend der Einschätzung der eigenen Emotionen eines Agenten und wird in der Emotionsgenerierungskomponente des Tok-Agenten abgebildet. Eine andere Möglichkeit, die Emotionen eines anderen Agenten zu erkennen, besteht darin, die emotionalen Verhaltensweisen bei diesem Agenten zu beobachten: “Em keeps the recognition of sadness apart from the emotional response, which allows for multiple ways of coming to know about the emotions of others. One way is to do reasoning and modeling, but another way, for example, is to see that an agent is crying. The Em model is more complete than the OCC model in cases such as agent A seeing that agent B is sad but not knowing why. In the OCC case, when agent A does not know why agent B is unhappy, the criteria for pity is not met. Because the default Em emotions generators require only that agent A believe that agent B is unhappy, which can be perceived in this case, Em generates pity“ ([Rei96], S. 54f.). Auf diese Weise sollen Emotionsstrukturen (*Emotion Structures*) entstehen, die kompletter und einfacher anzuwenden sind als ihre rein kognitiven einschätzungstheoretischen Vorbilder.

Reilly hat die Glaubwürdigkeit eines mit Em ausgestatteten virtuellen Charakters überprüft: Versuchspersonen wurden mit zwei virtuellen Welten konfrontiert, in denen zwei virtuelle Charaktere agierten, wobei einer der Charaktere mit Em ausgestattet wurde, der andere aber nicht darüber verfügte. Die Versuchspersonen stufen den Em-Charakter als emotionaler ein als den Nicht-Em-Charakter. Auch seine Glaubwürdigkeit wurde höher bewertet. Zugleich gaben die Versuchspersonen an, dass die Persönlichkeit des Em-Charakters deutlicher ausgeprägt war und dass sie bei ihm weniger häufig das Gefühl hatten, es mit einem fiktiven Charakter zu tun zu haben. Die Signifikanz der Ergebnisse schwankt allerdings deutlich, so dass Reilly einräumen muss, Em sei lediglich “moderately successful.“ ([Rei96], S. 129).

7.3.3. Der Ansatz von Aubé und Senteni: Emotion als Metapher für Kontrollstrukturen in MAS

Aubé und Senteni suchen nach leistungsfähigen Kontrollstrukturen für effektives Ressourcenmanagement in einem verteilten System. In [AS96] stellen sie zum ersten Mal ein Architekturmodell vor, in dem eine emotionsbasierte Kontrollstruktur (“emotion-like control structures“) für die Verbesserung der Ressourcenzuteilung im Kontext eines Multiagentensystems eingeführt wird. Die Autoren erklären, durch diesen Ansatz die Zuverlässigkeit, Autonomie und Adaptivität von MAS steigern zu können.

Aubé und Senteni entwickeln die Ideen von Maes [Mae91] und Norman und Long [NL95] weiter und erklären, dass Agenten autonom werden können, wenn sie die Kontrolle über ihre Motivationen selbst übernehmen und die Möglichkeit erhalten, sich die eigenen Ziele selbstständig zu setzen. Die Motivationen des Agenten - nach Aubé und Senteni sind dies *Bedürfnisse* und *Emotionen* - haben indirekt mit dem Ressourcenmanagement zu tun, da

7. Architekturen emotionaler Agenten

sie seine unmittelbaren Ziele bestimmen. Emotionen stehen somit mit den internen Zielen eines Agenten in einer wechselseitigen Beziehung und spielen deshalb auch beim Planen eine Rolle (vgl. auch [Aub05], [Aub09] für eine generelle Architekturübersicht und Auswertung).

Ressourcen Ressourcen werden von Aubé und Senteni definiert als alles was notwendig ist, um Leistung zu erbringen: “It could be physical energy like heat or electricity or food, but also time, money, affiliation, power, knowledge [...]” ([AS96], S. 2f.). Die Autoren unterscheiden zwei Arten von Ressourcen:

Ressourcen erster Ordnung (*first-order-Ressources*) sind solche, die dem Agenten unmittelbar zur Verfügung stehen und direkt konsumiert werden können (z.B. verfügbares Wasser und Nahrung für Tiere bzw. Zeit und Speicherplatz für künstliche Agenten); Bedürfnisse wie Hunger, Durst, Müdigkeit usw. sind Motivationsprozesse (*motivational processes*), die das Management und die Regulation des Konsums der Ressourcen erster Ordnung bewerkstelligen.

Ressourcen zweiter Ordnung (*second-order-Ressources*) sind für einen Agenten nur indirekt erreichbar, und zwar in dem Fall, dass er die anderen Agenten verpflichten kann, ihm diese Ressourcen zu erteilen. Zum Beispiel können Neugeborene nur Zugriff auf Nahrung erhalten, falls sie ihre Eltern verpflichten, ihnen diese Ressource zukommen zu lassen. Diese sozialen Verpflichtungen (*“commitments“*) bilden den Kontrollmechanismus über die Ressourcen zweiter Ordnung.

Commitments *Commitments* erfüllen eine der Grundaufgaben der gemeinsamen Ressourcenverwaltung in MAS: “By regulating exchange, they enable using each other as a way to access additional resources, that would remain otherwise inaccessible, or would be too costly to obtain” ([AS96], S. 3). Ein *Commitment* bedeutet im Kontext einer MAS, dass einem Agenten A bestimmte Ressourcen von einem Agenten B zugesprochen werden. Damit erhält A zusätzliche Ressourcen, wobei B in der Nutzung seiner eigenen Ressourcen eingeschränkt wird. Zum Beispiel kann B sich verpflichten, nur einen Teil seiner Ressourcen zu verbrauchen und den anderen Teil für A aufzuheben, oder wenn B die A zugesprochenen Ressourcen noch nicht besitzt, kann er sich verpflichten, sich das Ziel zu setzen, diese Ressourcen zu beschaffen, um sie dann an A abzutreten. Aubé und Senteni schlussfolgern aus diesen Zusammenhängen, dass soziale Verpflichtungen ihrerseits auch Ressourcen sind, nämlich Ressourcen zweiter Ordnung.

Emotion Die zentrale Annahme von Aubé und Senteni lautet, dass Emotionen (wie Ärger, Furcht, Freude, Trauer, Schuld usw.) als Motivationsprozesse für das Management von Ressourcen zweiter Ordnung - insbesondere der “Commitments” - fungieren. Emotionen werden als “commitment operators” gesehen, die dazu bestimmt sind, eine Reihe von Operationen über die Menge von “Commitments” auszuführen, wie z.B. “establish or create new commitments (joy, gratitude), protect, sustain, or reinvest old ones (joy, hope, gratitude, pride), prevent the breaking of commitments by self or others (pride, guilt, gratitude, anger), or call on ‘committed’ others in cases of necessity, danger, and helplessness (sadness, fear)” ([Aub98], S. 15).

So wie Bedürfnisse für das Management der Ressourcen erster Ordnung sorgen, übernehmen Emotionen eine ebensolche Rolle bei der Verwaltung der Ressourcen zweiter Ordnung: “Very much like needs, emotions are powerful control structures that act so as to protect agents - and societies of agents - from running out of resources“ ([AS96], S. 4).

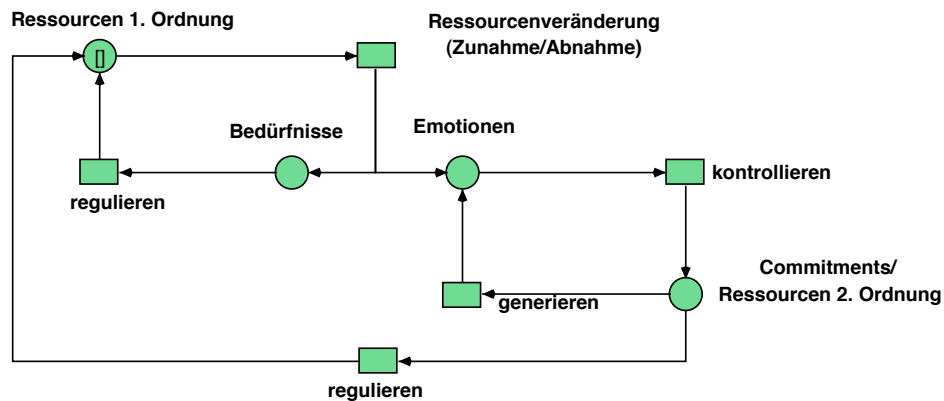


Abbildung 7.33.: Petrinetzmodell der Ressourcenmanagement durch Emotionen nach Aubé und Senteni.

Abb. 7.33 veranschaulicht die von den Autoren postulierte Kontrollfunktion der Emotion für das Management von Ressourcen zweiter Ordnung. Durch interne oder externe Ereignisse können Ressourcen erster Ordnung ab- oder zunehmen (Transition **Ressourcenveränderung (Zunahme / Abnahme)**). Bei Veränderung der Ressourcenmenge werden Bedürfnisse generiert, die nach Aubé und Senteni für das Management und die Regulation des Konsums der Ressourcen erster Ordnung zuständig sind. Die regulatorische Funktion der Bedürfnisse wird durch die Transition **regulieren** modelliert, welche durch die Generierung von Bedürfnissen (Markierung der Stelle **Bedürfnisse**) aktiviert wird, und der (initial markierten) Stelle **Ressourcen 1. Ordnung** wiederum eine Marke hinzufügt.

Neben der Generierung von Bedürfnissen führt die Ressourcenveränderung gleichzeitig zur Generierung von Emotionen (Stelle **Emotionen** wird markiert). Emotionen können sowohl vom antizipierten als auch vom tatsächlichen Ressourcengewinn bzw. -verlust ausgelöst werden und regulieren die Ziele des Agenten und seine Ressourcenverteilung nach einer Strategie zur Vermeidung der Ressourcenverluste (und damit verbundener negativer Emotionen, z.B. Furcht) und zur Erhöhung der Ressourcengewinne (und positiver Emotionen, z.B. Hoffnung).

Aubé und Senteni definieren zwei Emotionsklassen in Bezug auf Ressourcen ([AS96], S. 4):

- *Positive Emotionen* werden durch Ressourcengewinn ausgelöst und sollen diesen verstärken und zeitlich ausdehnen, indem sie als Ressourcenakkumulatoren fungieren (z.B. Freude und Hoffnung).
- *Negative Emotionen* werden durch Ressourcenverlust hervorgerufen und sollen diesen aufhalten bzw. abbremsen (z.B. Trauer und Furcht). Die schützende Funktion der

7. Architekturen emotionaler Agenten

negativen Emotionen könnte etwa dadurch implementiert werden, dass ein Agent, der Ressourcenverlust erleidet, von den anderen isoliert wird und sein Aktivitätsniveau herabsetzt, wie es oft mit verängstigten oder trauernden Menschen geschieht.

Somit besteht die zentrale Funktion der Emotionen in der Steuerung und Verwaltung von Ressourcen zweiter Ordnung, insbesondere der “Commitments” (Transition **kontrollieren**). Emotionen bestimmen u.a., ob Agenten neue Verpflichtungen eingehen bzw. bestehende erfüllen oder brechen. Die benannten Modifikationen der Commitments-Menge werden durch die Markierung der Stelle **Commitments / Ressourcen 2. Ordnung** modelliert. Die Veränderung der Commitments kann wiederum zur Generierung neuer Emotionen führen (Transition **generieren**). Alternativ können sie zur Wiederherstellung der initialen Menge der Ressourcen erster Ordnung beitragen (Transition **regulieren** schaltet, Stelle **Ressourcen 1. Ordnung** wird markiert), z.B. wenn ein neues Commitment geschlossen wird, das einen anderen Agenten verpflichtet, seine Ressourcen zur Verfügung zu stellen.

Emotionen haben eine höhere Verarbeitungspriorität als andere Mechanismen, d.h. sie bekommen sofort bei ihrem Auftreten Zugang zum Prozessor, wobei die laufenden Prozesse unterbrochen werden. Emotionen können sowohl durch einen aktuellen als auch durch einen antizipierten Gewinn oder Verlust ausgelöst werden und entweder zur intensiveren Zielverfolgung (beim antizipierten Gewinn) oder zur Änderung der aktuellen Planungsstrategie (falls ein Verlust erwartet wird) beitragen. Im letzten Fall würden die für die Erstellung eines neuen Plans benötigten Ressourcen dem aktuellen Planungsprozess entzogen, so dass dieser zeitweise (oder u.U. komplett) unterbrochen wird. So haben Emotionen im Modell u.a. die Rolle eines Interrupt-Mechanismus: Sie können die aktuelle Verarbeitung lenken und so die Ziele und Pläne eines Agenten modifizieren.

Mit *agency* bezeichnet Aubé [AS96] die kausale Zuweisung von Ereignisursachen, wobei ein Agent A einen Agenten B für bestimmte Ereignisse in einem MAS verantwortlich macht. Ärger wird in diesem Zusammenhang erklärt als Prozess, der die Situationskontrolle übernimmt, wenn ein signifikanter Ressourcenverlust auf die Aktionen eines anderen Agenten zurückgeführt wird. Wenn ein Agent sich selbst die Verlustursache zuweist, entsteht die Emotion Schuld. Entsprechend entstehen beim Ressourcengewinn durch andere oder sich selbst Dankbarkeit oder Stolz. Wenn diese Emotionen antizipiert werden, versucht ein Agent die Aktionen von anderen, aber auch eigene Handlungen im Sinne der Maximierung des Ressourcengewinns zu beeinflussen. Wenn keine “Verantwortlichen” gefunden werden oder sie nicht beeinflusst werden können, entstehen die “ungerichteten” Emotionen Freude, Hoffnung, Traurigkeit und Furcht.

Die so entstehende Emotionstaxonomie (siehe Abb. 7.34) wird von den Autoren mit den Einschätzungstheorien der Psychologie in Verbindung gebracht: “[T]he taxonomy is pretty much congruent with the intersecting aspects of the current appraisal theories in the psychology of emotion“ ([AS96], S. 6). Aubé und Senteni verweisen dabei auf die Emotionstheorien von Ortony, Clore und Collins [OCC88], Scherer [Sch88], Roseman [Ros96] sowie von Smith und Ellsworth [SE85], jedoch wurden die Konzepte dieser Theorien, insbesondere die Einschätzungsdimensionen, offensichtlich nicht direkt übernommen. Die Emotionen Freude, Hoffnung, Traurigkeit und Furcht generieren neue *Commitments* zwischen Agenten (z.B. wenn diese Ressourcen teilen oder an andere abtreten, sozialen Zusammenhalt stärken,

nach Hilfe und Unterstützung suchen usw.). Die Emotionen Stolz, Schuld, Dankbarkeit und Ärger dienen der Aufrechterhaltung und Stärkung der zwischen Agenten bestehenden *Commitments* (z.B. durch Lob oder Drohungen).

		POSITIVE		NEGATIVE	
		actual	anticipated	actual	anticipated
A G E N T C Y	none	JOY	HOPE	SADNESS	FEAR
	others	GRATITUDE		ANGER	
	self	PRIDE		GUILT	

Abbildung 7.34.: Einschätzung der Basisemotionen (nach [AS96], S. 6).

Aubé und Senteni [AS96] betonen insbesondere die soziale Natur ihrer Emotionsdefinition: Da Emotionen als *commitments operators* fungieren, müssen sie eine bestimmte kommunikative Funktion innerhalb einer Agentengemeinschaft besitzen. “Our basic criterion for counting a motivational process as an emotion, is that it should provide for some communicative means in the service of interagent binding (such as alarm or distress calls, warning growls etc.)“ ([AS96], S. 7).

Im Abb. 7.35 wird der bisher verbal skizzierte Kontrollfluss beim Ressourcenmanagement veranschaulicht.

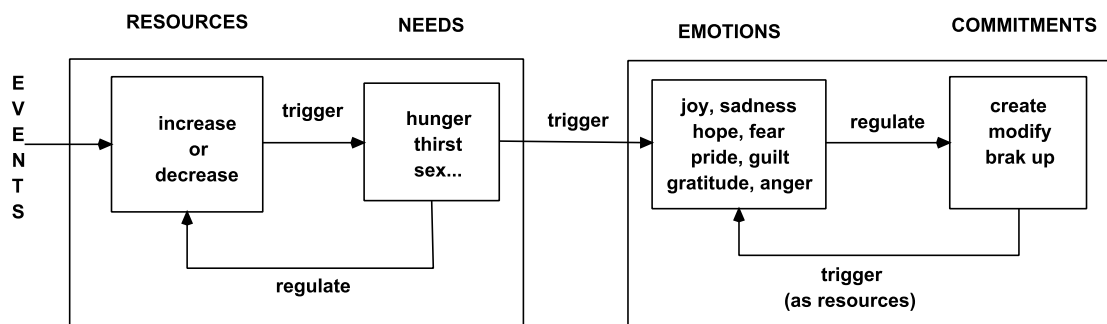


Abbildung 7.35.: Kontrollfluss bei der Ressourcenmanagement (nach [AS96], S. 7).

Auf dem Diagramm sieht man, dass bei dem Versuch, die signifikante Ressourcenveränderung zu regulieren, die Bedürfnisse (needs) ihrerseits Emotionen auslösen können. Dies entspricht weitestgehend der Annahme von Damasio über die Hintergrundemotionen, welche

7. Architekturen emotionaler Agenten

die Basis für die primären Emotionen bilden (vgl. Kapitel 5.5). Ohne auf Damasio's Emotionsmodell bzw. sein "Nesting Principle" direkt zu verweisen, kommen Aubé und Senteni in [AS96] ihrerseits zu den gleichen Ergebnissen in Bezug auf den Kontrollfluss bei der Emotionsaktivierung. Später wurden Damasio's Ergebnisse in das Modell der emotionsbasierten Steuerung von Commitments integriert (vgl. [Aub05], [Aub09]).

Emotion und Kommunikation in MAS Aubé und Senteni [AS96] argumentieren, dass Agenten in der Lage sein sollten, ihre Emotionen einander zu kommunizieren. Dazu schlagen sie die Benutzung der expressiven Sprechakte ("expressive speech act") aus der Taxonomie von Winograd und Flores (1986) vor. Diese Sprechakte wurde in die meisten Formalismen zur Agentenkommunikation nicht aufgenommen. Aubé und Senteni hingegen argumentieren, dass Agenten emotionale (Sprech-)Akte ("emotional act") benötigen: "Hence, communication between agents should not only reflect the verbal aspect of typical human communication but most importantly emotional non-verbal aspects as well" ([AS96], S. 9). Emotionale (Sprech-)Akte werden ausgelöst, wenn das Risiko besteht, gegen aktuelle Commitments zu verstoßen, bzw. wenn sich eine unvorhergesehene Möglichkeit ergibt, neue Commitments herzustellen. Ein Risiko bedeutet in diesem Zusammenhang, dass der Zufluss bzw. Mangel an vorhandenen oder benötigten Ressourcen zweiter Ordnung einen vorher definierten Schwellenwert übersteigt.

Das Modell von Aubé und Senteni wurde in einem Multiagentenszenario erprobt. Es konnte gezeigt werden, dass die in diesem Modell definierten emotionalen Mechanismen gut geeignet sind, den Ressourcenzugriff in einem MAS zu kontrollieren und insbesondere Interrupt-Mechanismen zu implementieren, welche die aktuellen Ziele und Pläne eines Agenten steuern und manipulieren können. Die Autoren plädieren für die Einführung emotionaler Mechanismen zur besseren Umsetzung der grundlegenden Eigenschaften von Agentensystemen: "Their introduction improves the potential autonomy of agents by taking explicitly into account the management of the resources that are the core of their activity and that should hence be part of the fundamental definition of more autonomous, reliable and adaptive multi-agent systems" ([AS96], S. 13).

7.4. Auswertung der informatischen Emotionsmodelle

Wir schließen dieses Kapitel mit einer allgemeinen Auswertung der betrachteten Emotionsarchitekturen.

Zunächst werden im nächsten Abschnitt die aktuellen Entwicklungen auf dem Gebiet der informatischen Emotionsmodellierung in einem Überblick über bisher nicht betrachtete Emotionsarchitekturen vorgestellt. Anschließend werden die informatischen Emotionsarchitekturen zusammenfassend analysiert und im Hinblick auf eine Reihe vordefinierter Bewertungskriterien ausgewertet, insbesondere die Gültigkeitsgrenzen, die implizite Emotionsdefinition, die emotionstheoretische Fundierung, die verwendete Terminologie und die empirische Überprüfung.

7.4.1. Überblick über weitere informatische Emotionsarchitekturen

Die Übersicht in Abb. 7.1 zeigt die in dieser Arbeit betrachteten Emotionsarchitekturen und die zugehörigen Emotionstheorien, die für das jeweilige Modell die theoretische Grundlage bilden. In Kapitel 7.1 wurde die Zusammenstellung der bis dato einflussreichsten Emotionsmodelle nach [MGP12] wiedergegeben (vgl. Abb. 7.1). Beide Übersichtsdarstellungen weisen viele Überschneidungen auf. Insbesondere sieht man, dass der Mehrzahl der betrachteten Emotionsarchitekturen eine Einschätzungstheorie zugrundeliegt.

Zurückgegriffen wird vorwiegend auf Theorien, die sich gut für eine Operationalisierung eignen, vor allem auf die vier in Kapitel 6 dargestellten Einschätzungstheorien: die OCC-Theorie von Ortony, Clore und Collins [OCC88] sowie die Theorien von Lazarus [Laz91], Scherer [Sch93] und Frijda [Fri86]). Die Tabelle 7.1 zeigt die emotionstheoretischen Grundlagen, welche den bisher betrachteten Architekturmodellen zugrunde liegen.

Architekturmodell	Emotionstheoretischen Grundlagen
Affective Reasoner-Architektur von Elliott [Ell92]	OCC-Theorie [OCC88]
Em-Architektur von Bates / Reilly [Rei96]	OCC-Theorie [OCC88]
Modell der emotionsbasierten Ressourcenverteilung von Aubé / Senteni [AS96]	OCC-Theorie [OCC88], Scherer [Sch93], Roseman [Ros96], Smith / Ellsworth [SE85]
TABASCO-Architektur von Staller / Petta [SP98]	Lazarus [Laz91], Frijda [Fri86], Scherer [Sch93], Leventhal / Scherer [LS87]
Rahmenwerk zur emotionsbasierten Kontrolle (EBC) von Velásquez [Vel99]	Damasio [Dam94], Ekman [Ekm92], Izard [Iza91], Johnson-Laird / Oatley [OJL87]
PSI / MicroPsi (Dörner [DLD ⁺ 01], Bach [Bac03])	Dörner [DLD ⁺ 01]
CogAff-Rahmenwerk von Sloman [Slo01]	Eigene Emotionstheorie

Tabelle 7.1.: Emotionstheoretischen Grundlagen der betrachteten Architekturmodelle

Meist weist ein Einschätzungsmodell die Repräsentation einer begrenzten Anzahl diskreter emotionaler Zustände auf. Für die Abbildung der Emotion auf die emotionale Reaktion wird des Öfteren eine regelbasierte Zuordnung gewählt (vgl. Affective Reasoner [Ell92], Em [Rei96]). Dieses Vorgehen wurde jedoch in der Literatur mehrfach kritisiert, einerseits wegen der fehlenden theoretischen Fundierung der Abbildung “Emotion – emotionale Reaktion“, andererseits wegen der Inflexibilität der regelbasierten Verarbeitung (vgl. z.B. [BodAHN03], S. 2 bzw. [SP00], S. 5).

“The reification of emotional states results in the implementation of a full explicit mapping from these states to entailed effects, including internal processing and externally observable overt behaviour. The characteristics of systems engineered

7. Architekturen emotionaler Agenten

according to such a shallow approach are well known from the traditional research area of expert systems in artificial intelligence: the merits of rather straightforward design – for moderate ruleset size – and precisely known coverage stand against a number of problems, including brittleness of system behaviour surfacing with every occurrence of any situation not explicitly anticipated at implementation time; laboriousness of system extension; and the issue of overall system consistency: as the agent’s behaviour is governed by a large collection of independent rules – as opposed to a small set of generating principles – it falls into the responsibility of designers and implementors to ensure that with a growing body of incorporated knowledge the system remains free of inconsistencies and continues to perform in a desired and coherent fashion.“ ([SP00], S. 5).

In den letzten Jahren lässt sich bei den informatischen Einschätzungsmodellen eine Wandlung hin zu einer abstrakteren Modellierung des Emotionsgenerierungsprozesses feststellen. Neuere Arbeiten weisen eine flexiblere domänenunabhängige Abbildung der Einschätzungsvariablen auf Emotionen auf, die auf bekannten KI-Techniken basiert (z.B. die Markov-Ketten in dem Modell von El Nasr et al. [ENJT00]). In dem Modell WILL von Moffat und Frijda [MF00] werden Einschätzungsvariablen an explizite Modelle von Plänen gebunden, welche den kausalen Zusammenhang zwischen Aktionen und Aktionsergebnissen erfassen (vgl. Kapitel 6.1.2).

Vergleichsweise wenige Forscher haben nicht einschätzungstheoretische Modelle des emotionalen Einflusses auf höhere kognitive Prozesse untersucht. Z.B. wurde in [BTS02] versucht, im Rahmen einer sogenannten “Decision Field Theory“ das Konzept der Triebe in die klassische Entscheidungstheorie zu integrieren, um den Einfluss der Emotion auf das Entscheiden zu erklären. Gmytrasiewicz und Lisetti haben in [GL00] die Benutzung der Emotion als Heuristik bei der Berechnung der Utility-Funktion eines Agenten diskutiert.

In einigen Emotionsmodellen wurde versucht, die bisherigen neurowissenschaftlichen Forschungsergebnisse zur Emotionsfunktion auf der neurobiologischen Ebene getreu abzubilden. Beispielsweise wurde in [ASSCL97] ein Modell der Angstkonditionierung ([LeD00], vgl. Kapitel 6.2.1) in einer Computersimulation nachgebildet. In weiteren Modellen werden die physiologischen Effekte der Emotionen thematisiert (vgl. [Vel98]).

Insgesamt zeigte die Betrachtung der KI-Modelle, dass die neurowissenschaftlichen Einflüsse, vor allem die aus den Arbeiten von Damasio ([Dam96], [BDD00], [Dam03], [FADT11]) und LeDoux ([LeD84], [LeD96a], [LeD96b], [LeD98], [LeD00]) hervorgehenden Sichtweisen, die informatische Emotionsmodellierung tief durchdrungen haben. Die Mehrheit der Autoren betrachtet Emotionen als einen wesentlichen Bestandteil des Kontrollsystems eines Agenten und definiert ihre Rolle in einem intelligenten System vor allem funktional (Velásquez, Canamero, Staller / Petta). Emotionen werden inzwischen nicht mehr als Anhängsel des kognitiven Systems, sondern als unverzichtbare Voraussetzung für das zuverlässige Funktionieren von Kognition betrachtet.

Mehrere Emotionsforscher haben natürliche emotionale und motivationale Mechanismen als Vorbilder für den Entwurf des Handlungswahlprozesses in autonomen Agenten, insbesondere in Robotern, genutzt. Emotionen werden hierbei als ein evolutionärer Anpas-

7.4. Auswertung der informatischen Emotionsmodelle

sungsmechanismus gesehen: “We see emotions as mechanisms that aim at modifying or maintaining the relation of the agent with its (external or internal) environment (rather than modifying the environment itself) in order to preserve the agent’s goals.“ ([Can03], S. 115).

In dem Modell von Canamero [Can97] sind künstliche Lebewesen mit sehr beschränkten Wahrnehmungs- und Handlungsfähigkeiten implementiert, die in einer hoch dynamischen, unsicheren Umgebung agieren. Sie verfügen über eine synthetische Physiologie (überlebensrelevante Kontrollvariablen und Hormone), welche ihren internen Zustand konstituiert und interne Emotionen (emotionale Zustände) entstehen lässt, die auf homeostatischem Wege kontrolliert werden können.

Canamero konstruiert also die Verbindung zwischen Emotionen, Motivationen und Verhalten durch die Implementierung einer synthetischen physiologischen Ebene, auf der die Interaktion der drei Architekturelemente stattfindet. Dabei wird ein sehr rudimentäres emotionales System modelliert, indem lediglich Basisemotionen berücksichtigt und die “kognitiven“ Aspekte der Emotionalität wie Einschätzung der Situation und der möglichen Folgen emotionaler Reaktion, Emotionskontrolle usw. nicht einbezogen werden. Diese Lösung hat sich in Bezug auf die spezielle Aufgabe des schnellen Reagierens in hoch dynamischen Umgebungen gut bewährt, ist aber für komplexere Aufgabenstellungen, insbesondere wenn soziale Beziehungen mit ins Spiel kommen, weniger geeignet ([Can03], S. 139). Die theoretische Begründung für den Ansatz von Canamero bilden ausgewählte Gesichtspunkte der Emotionstheorie von Frijda (siehe Kapitel 6.25).

In den Modellen von Gratch und Marsella wird neben Emotionsgenerierung insbesondere der Einfluss der Emotion auf höhere kognitive Prozesse und Verhalten, speziell die Anpassung, betrachtet.

Das Einschätzungsmodell *Émile* [Gra00] basiert auf Elementen der *Em*-Architektur [Rei96] und der Affective-Reasoner-Architektur [Ell92] und hat ihre emotionstheoretische Fundierung in den Emotionstheorien von Ortony, Clore und Collins [OCC88] und Lazarus [Laz91]. Das Modell kombiniert zwei fundamentale Elemente der Emotionsmodellierung: Einschätzung und Anpassung (*coping*), wobei Einschätzung als emotionale Signifikanz eines Ereignisses bzw. einer Umgebungssituation für die Agentenziele beschrieben wird, während *Coping* als Anpassung des Agenten an diese emotionale Signifikanz durch Modifikation seiner Handlungen oder Ziele definiert wird [GM03]. Ein signifikanter Unterschied zu der *Em*-Architektur von Bates und Reilly [Rei96] besteht darin, dass *Émile*-Agenten den emotionalen Zustand anderer Agenten beobachten und ihr Verhalten entsprechend anpassen können [Gra00].

Als besonderer Vorteil von *Émile* wird in [SMBB08] die unmittelbare Integration der Anpassungsmechanismen in die Agentenarchitektur hervorgehoben (z.B. [MG02]). Die Ausweitung von *Coping* auf Agentenziele und -überzeugungen wird durch die Annahme begründet, dass auch die menschliche Emotionsbewältigung eine Art Mischung aus problembasierter und emotionsbasierter Verarbeitung ist, wobei die problembasierte Verarbeitung darauf abzielt, die emotionsauslösende Situation oder das emotionsauslösende Ereignis zu verändern bzw. zu beeinflussen, während bei der emotionsbasierten Verarbeitung versucht wird, die individuelle Sicht der Situation zu modifizieren bzw. die Situation anders einzuschätzen, um ihre emotionale Signifikanz zu verändern, etwa durch eine andere

7. Architekturen emotionaler Agenten

Verantwortungszuschreibung. Ein anderer Vorteil von *Émile* ist die Einbeziehung der Emotionsintensität unter Berücksichtigung der Relevanz der korrespondierenden Agentenziele [GM04]. Diese Architektur ist sehr gut geeignet für Anwendungen, in denen Agenten gute Planungsfähigkeiten mit Reaktionsfähigkeiten kombinieren müssen.

Auf Basis von *Émile* haben Gratch und Marsella ([GM04]) ein allgemeineres Emotionsmodell EMA (Emotion and Adaptation) entwickelt, das auf der kognitiv-motivational-emotionalen Theorie von Smith and Lazarus basiert und Elemente anderer wichtigen Emotionstheorien (von Scherer [Sch84], Frijda [Fri86], OCC-Theorie [OCC88] und Lazarus [Laz91]) integriert. Ferner wurde ein affektiver Planer auf Basis von *Émile* erstellt, der in der Agentenarchitektur *FearNot!* [ADP06] eingesetzt und evaluiert wurde.

7.4.2. Auswertung der Emotionsmodelle

Im Folgenden werden die Ergebnisse der Architekturbetrachtung in diesem Kapitel zusammengefasst und auf einer allgemeinen Ebene ausgewertet. Die Betrachtung der informatischen Emotionsmodelle erfolgt insbesondere unter folgenden Gesichtspunkten:

1. *Gültigkeitsgrenzen der Modelle:* Vor dem Hintergrund des Fehlens einer eindeutigen theoretischen Emotionsdefinition sollte beurteilt werden, wie die informatischen Emotionsmodelle ihre Modellierungsgegenstände abgrenzen und wie die verschiedenen (impliziten und expliziten) Emotionskonzepte miteinander korrelieren.
2. *Theoretische Fundierung der Emotionsmodelle:* Der Mangel an Anwendungswissen aufseiten der Informatik in Bezug auf sozialwissenschaftliche Emotionstheorien kann einerseits zu Missverständnissen gegenüber bestimmten Konzepten führen, andererseits aber auch zur Entwicklung eigener intuitiver Konstrukte. Es wurde deshalb die emotionstheoretische Fundierung der betrachteten Emotionsarchitekturen ausgewertet. Insbesondere wurde geprüft, wie die informatische Modellierung mit dem Fehlen einer allgemeinen Emotionstheorie und einer starken Forschungsfragmentierung aufseiten der sozialwissenschaftlichen Emotionsforschung umgeht, welche Emotionstheorien besonders oft implementiert werden und sich als geeignet für die Umsetzung in informatischen Modelle erwiesen haben, inwieweit die den Modellen zugrunde liegenden Originaltheorien unverändert implementiert bzw. bei der Implementierung modifiziert oder ergänzt werden usw.
3. *Vergleich von Modellarchitekturen:* Das Fehlen universeller Repräsentationstechniken für emotionale Architekturen führt zu Schwierigkeiten bei der Gegenüberstellung verschiedener emotionaler Architekturen ([Sch02], [Slo04], [GM09a]). Es wurden die Möglichkeiten diskutiert, wie der Vergleich bzw. die Diskussion der Architekturen erleichtert werden könnte.
4. *Berücksichtigung funktionaler Zusammenhänge und sozialer Aspekte der Emotion:* Es wurde bewertet, inwieweit die Emotion eine funktionale Rolle im System spielt. Ferner wurde die Modellierung der sozialen Faktoren Emotionsgenese und Emotionsexpression betrachtet.

5. Schließlich wurde die *empirische Überprüfung* der Emotionsmodelle bewertet.

Eingrenzung des Emotionsbegriffs Die theoretische Diskussion um den Emotionsbegriff spiegelt sich in der Emotionsmodellierung wider. Da selbst die natürlichen Emotionen keineswegs eindeutig definiert oder zumindest zu anderen mentalen Phänomenen abgegrenzt werden können (vgl. Kapitel 5.6), gibt es viele unterschiedliche Auffassungen, wenn es um die künstlichen Emotionen (“artificial emotions“) geht (vgl. [Bac02], [Pic03], [SS02]).

“Like for its ‘natural’ counterpart, there are many dissimilar notions, which focus on different aspects, typically depending on the function authors attribute to emotion, or on the application within the project“ ([Bac02], S. 4).

Obwohl dieses Kapitel nur einen kleinen Ausschnitt aus dem Gebiet der Emotionsmodellierung präsentiert, wird aus den dargestellten Modellen deutlich, dass die Emotionskonzepte verschiedener Forscher teilweise sehr weit auseinanderliegen. Es gibt keine grundsätzliche Übereinkunft darüber, ob die physiologischen Emotionsaspekte, situationsbasierten Bewertungen, Motivationen, Wünsche oder Aspekte von expressivem Verhalten unter dem Emotionsbegriff miterfasst werden. Es werden in den Emotionsmodellen sehr unterschiedliche Zusammenhänge zwischen Emotionen und Trieben, Wünschen, Überzeugungen, situationalen Begebenheiten, Perzeption, physiologischen Reaktionen, Verhaltenstendenzen und anderen Komponenten des emotionalen Geschehens unterstellt:

“Some researchers, for instance, see emotions as special kinds of motivations (e.g., (Breazeal 1998)), whereas others draw a clear distinction between motivations and emotions (e.g., (Canamero 1997)). Such subtle changes in terminology have the unhappy consequence that the research community as a whole can easily lose track of what its various members are talking about. This is not only undesirable for any discipline, but may eventually culminate in a terminology so conflated that claims involving any of its terms are nearly meaningless.“ ([Sch02], S. 3).

Während für Bates und Reilly [Rei96] und auch für Elliott [EO92] Emotionen durch sichtbares expressives Verhalten definiert werden, fungieren sie in den Modellen von Velásquez ([Vel97], [Vel99]) und Dörner ([DLD⁺01], [DSD01]) als Kontrolloperatoren und steuern die Gesamtarchitektur, speziell die Perzeption, das Planen und die Aktionswahl bei Dörner bzw. die Entscheidungsfindungs- und Deliberationsmechanismen bei Velásquez. Für Sloman [Slo01] wiederum sind Emotionen ein globales Cluster-Konzept, das nur in Abhängigkeit von der Systemarchitektur zu definieren ist.

Das Fehlen einer einheitlichen Emotionsdefinition in der Emotionsmodellierung wurde in der Literatur bereits heftig kritisiert:

“Because of this conceptual mess, anyone can produce a program, label some component of it the “emotion module“ and proudly announce that they have developed a robot or software agent that has emotions. It will be hard to argue against such claims when there is no agreement on what emotions are.“ ([Slo03], S. 38).

7. Architekturen emotionaler Agenten

Emotionstheoretische Fundierung der Modelle Die Entwicklung von Agentenarchitekturen, die sich in ihrem Aufbau an den Mechanismen der menschlichen Kognition orientieren bzw. menschliches Verhalten nachbilden, erfordern als theoretische Basis psychologische Theorien der Emotion, der Motivation, der Wahrnehmung und des Gedächtnisses, die eine funktionale Beschreibung der genannten Systeme liefern und zugleich hinreichend detailliert und formalisierbar sind.

“Unfortunately, there are few theories in existence that meet this demand, partly because current psychological methodology does not encourage the formulation of comprehensive functional theories. (Newell 1990)“ ([Bac03], S. 1)

Wie die Übersicht der Emotionsmodelle im vorherigen Kapitel gezeigt hat, basiert deren absolute Mehrheit auf Einschätzungstheorien, wobei besonders oft auf die Theorien von Ortony, Clore und Collins [OCC88], Scherer [Sch84], Frijda [Fri86] und Lazarus [Laz91] zurückgegriffen wird (vgl. Kapitel 6.1.3).

Da die meisten Einschätzungstheorien lediglich den Prozess der Emotionsgenerierung betrachten und über die Funktionen der generierten Emotionen innerhalb des Gesamtsystems keine Aussagen machen (vgl. Kapitel 6.1.3), müssen bei ihrer Umsetzung in einem Agentensystem weitere Theorien hinzugezogen werden. Bei vielen Umsetzungen werden an dieser Stelle intuitive Konstrukte der Entwickler eingesetzt, um die fehlende Spezifizierung der Auswirkungen der Emotion auf Agentenverhalten zu ersetzen:

“[R]eification of emotions as identifiable system components [...] engenders the problem of how to proceed from these emotions for further system processing, leading to the adoption of ad-hoc constructions of dubious validity.“ ([SP00], S. 6).

Einige Forscher halten deshalb die Einschätzungsmodelle bei der Modellierung der Emotionsauswirkungen auf das Systemverhalten für ungeeignet und schlagen die Integration neurobiologischer Emotionsgrundlagen in die Modellierung vor:

“[A]lthough suitable to test theories of cognitive evaluation of emotional experiences, computational models of emotion appraisal are very limited for the understanding on how the brain might process emotions. This is primarily due to the fact that these models are based on post-hoc cognitive recollections of emotional experiences and they are not concerned with the neural mechanisms behind such processes.“ ([Vel98], S. 3).

Um die Unvollständigkeit der einschätzungstheoretischen Grundlagen auszugleichen, sind die Forscher aufseiten der Informatik meist gezwungen, gleichzeitig mehrere Emotionstheorien in ihren Modellen umzusetzen oder die Aspekte verschiedener Modelle zu kombinieren. So basiert das Modell von Staller und Petta ([SP98], [Pet03], [PT06]) auf der Kombination der Emotionstheorien von Frijda [Fri86], Scherer [Sch93] und Lazarus [Laz91]. Canamero [Can03] kombiniert die Aspekte der Theorien von Frijda [Fri86] und Rolls [Rol05] mit dem neurobiologischen Konzept der homeostatischen Regulation und das von Velásquez ([Vel97],

[Vel98], [Vel99]) konzipierte emotionale System integriert gleich mehrere psychologische und neurowissenschaftliche Emotionstheorien von Ekman [Ekm92], Izard [Iza91], Johnson-Laird und Oatley [OJL87] sowie Panksepp [Pan98], Damasio [Dam96] und LeDoux [LeD96a].

Wegen der uneinheitlichen Terminologie und allgemeiner konzeptioneller Verwirrung hinsichtlich der Emotionsbegriffe verschiedener Emotionstheorien (vgl. Kapitel 5) bereitet die Integration von mehreren Emotionstheorien zu einem Modell einige grundsätzliche Probleme. So beziehen sich die Emotionstheoretiker in ihrer Auffassung der Emotion oft auf sehr verschiedene Phänomene, ohne dass dies aus den einschlägigen Theoriebeschreibungen ausdrücklich hervorgeht. Deshalb können bestimmte Emotionstheorien nur bedingt miteinander kombiniert werden oder sich sogar als unkompatibel erweisen.

Meist wird die Einschätzungstheorie nicht in ihrer gesamten Komplexität implementiert, u.a. weil sie einige konzeptionelle Probleme beinhaltet, die einer umfassenden Implementierung im Wege stehen (vgl. Kapitel 6.1.3). Vielmehr werden nur bestimmte theoretische Elemente herausgegriffen, die dann um eigene Komponenten erweitert werden, oft ohne dass dies explizit deutlich gemacht wird.

Um schnell Ergebnisse zu erzielen, werden von den entworfenen Modellen wiederum nur Teilbereiche in Simulationen oder Robotern implementiert. Dabei wird zu pragmatischen Lösungen gegriffen, die komplexe Prozesse notwendigerweise auf einige wenige Variablen reduzieren. Diese Variablen werden zudem häufig willkürlich definiert, um das Modell überhaupt umsetzen zu können. Dies macht die Evaluierung der erstellten Modelle oft sehr schwierig.

Fehlen einer gemeinsamen Terminologie und von Vergleichskriterien bei der Architekturbeschreibung Obwohl in den letzten Jahren eine Diskussion über die Architekturen für Agenten allgemein und speziell für emotionale Agenten in erhöhtem Maße stattfand, existiert bisher keine generell akzeptierte Terminologie zur neutralen und architekturunabhängigen Beschreibung verschiedener Architektureigenschaften und -komponenten. Auch wenn die gleichen Labels in Bezug auf Emotionsmodelle benutzt werden (‘reaktiv’, ‘deliberativ’, ‘reflektiv’), beziehen sie sich in verschiedenen Architekturen nicht unbedingt auf die gleichen Architektureigenschaften. Oft verstehen die Forscher unter demselben Begriff völlig verschiedene Mechanismen:

“For instance, some may be inclined to describe any component that can evaluate two possibilities for action and then select one as ‘deliberative’, whereas others would require that to be done in the context of a mechanism that can consider, evaluate, and select between alternative multi-step actions, using a formalism with compositional semantics. The latter is what I thought ‘deliberative’ meant, until I met people using the label for selection processes in a reactive mechanism“ ([Slo04], S. 3).

Immer mehr Emotionsforscher sehen die Notwendigkeit eines Übereinkommens in Bezug auf allgemein akzeptierte Termini zur Architekturbeschreibung, um eine gemeinsame Kommunikations- und Diskussionsgrundlage zwischen verschiedenen Forschungsfeldern

7. Architekturen emotionaler Agenten

innerhalb von KI und VKI zu schaffen. Mit einem gemeinsamen terminologischen Rahmenwerk zur Klassifizierung und Gegenüberstellung verschiedener Architekturen könnte man eventuell der starken Fragmentierung der Forschung, der “Erfindung“ bereits existierender Architekturen und Mechanismen sowie dem terminologischen Durcheinander entgegenwirken. In mehreren Arbeiten (vgl. [Sch02], [SS02], [Slo03], [Slo04], [GM09a], [GM09b]) wurde deshalb für eine gemeinsame Beschreibungstechnik zur Repräsentation und Klassifizierung emotionaler Agentenarchitekturen plädiert:

“We need a framework that explains how there can be all the diverse phenomena that different people refer to when they talk about emotions and other affective states and processes“ ([Slo03], S. 35).

“If people at least try to work with a common framework for talking about architectures and the states they support, it may be possible for more collaboration and more critical comparative analysis to accelerate progress“ ([Slo04], S. 6).

Ein zentrales Anliegen der vorliegenden Arbeit geht dahin, einen gemeinsamen konzeptuellen und terminologischen Rahmen zur Beschreibung der informatischen Emotionsarchitekturen zu entwickeln, der die Analyse derselben aus verschiedenen Perspektiven möglich macht. Auf Basis eines referenzbasierten Modellierungsrahmenwerks, das bereits zur Darstellung von nicht emotionalen Agentenarchitekturen erfolgreich eingesetzt wurde (vgl. z.B. [Röl04]), soll eine einheitliche Modellierung und der Vergleich verschiedener Emotionsarchitekturen möglich werden. Nicht nur die Repräsentation von Prozessen innerhalb der Agentenarchitektur, sondern auch die Repräsentation sozialer Prozesse in Agentengesellschaften soll so innerhalb des gleichen Rahmenwerks erfolgen. Aufgrund der hier aufgestellten Hypothese einer wechselseitigen Abhängigkeitsbeziehung zwischen agenteninternen (psychologischen) Emotionsphänomenen und externen (soziologischen) Gesellschaftsphänomenen (siehe Kapitel 6.3.3) erscheint die gemeinsame Repräsentationstechnik zur Formalisierung dieser verschiedenartigen Prozesse besonders erfolgversprechend, zumal sie auch die Verifizierung der entstehenden Modelle ermöglicht. Die Vorstellung eines Referenzmodells zur Beschreibung von emotionalen Agentenarchitekturen erfolgt in Kapitel 8.

Vernachlässigung der sozialen Komponente Allen psychologischen Emotionstheorien ist die mangelhafte Berücksichtigung der sozialen Komponente von Emotionen gemeinsam. Die Entwickler der VKI-Systeme, bei denen die soziale Komponente der Emotionen eine besonders wichtige Rolle spielt, aber auch die KI-Forscher sind deshalb gezwungen, auf die soziologischen Emotionstheorien zurückzugreifen und sie mit einer Einschätzungstheorie zu kombinieren (vgl. Staller und Petta [Pet03], [PT06]) oder sich mit selbst entwickelten Sozialitätsmodellen zu behelfen (vgl. Elliotts Ergänzung des OCC-Modells um die OOC-Komponente [EO92]). Dadurch kommt zusätzlich viel Willkür in die Modellierung.

Bislang konzentrierte sich die Forschung an emotionalen Agenten auf isolierte Entitäten oder aber auf dyadische Interaktionen (Agent - Agent/Agent - Benutzer). Wie der Überblick der Forschungsprojekte aus den letzten Jahren in Kapitel 7.4.1 gezeigt hat, wird bei der Emotionsmodellierung vorwiegend auf psychologische, seltener auf neurowissenschaftliche

7.4. Auswertung der informatischen Emotionsmodelle

Emotionstheorien zurückgegriffen. Die soziologischen Erklärungsansätze zur Emotionalität sollen jedoch ebenfalls in die informatischen Emotionsmodelle Eingang finden. [vSM04] argumentieren: “Sofern der Anspruch der Informatik darin begründet liegt, diejenigen Ursachen und Auswirkungen der Emotionen in technischen wie soziotechnischen Systemen zu berücksichtigen, die konstitutiv für das natürliche Phänomen Emotion sind und potentiell informatischen Zwecken dienen, kann sie nicht darüber hinwegsehen, eben auch die sozialen Funktionen zu berücksichtigen“ ([vSM04], S. 12).

Nur wenige Forscher beschäftigten sich bisher mit der Rolle und dem Potenzial der Emotionen in Multiagentensystemen in Hinblick auf Probleme wie Koordination, Kooperation oder Strukturausbildung mit Hilfe von Emotion ([Ell93], [AS96], [SP00], [GL00], [vSM04], [MG01], [MG09]). In der soziologischen und neurowissenschaftlichen Diskussion besteht jedoch bezüglich dieser Frage weitgehend Konsens: Es wird angenommen, dass Emotion die Gesellschaftsnormen auf der Mikroebene konstituiert (vgl. die Kapitel 6.3.3 und 6.2.2). Insofern bietet sich die Emotionsmodellierung in Agenten eines MAS an, um die Hypothese der Normkonstituierung durch Emotion simulationstechnisch zu überprüfen.

Empirische Überprüfung der Emotionsmodelle Von den informatischen Emotionsmodellen wurden nur wenige empirisch evaluiert, wobei die Mehrzahl der Evaluierungen die HCI-Aspekte der bewerteten Modelle betrifft. Die durchgeführten Modellbewertungen konzentrieren sich meist auf die Überprüfung des vom Modell erzeugten externen Systemverhaltens, z.B. die Bewertung der Glaubwürdigkeit des Verhaltens eines synthetischen Charakters [Rei96]. Die Modellierung des Emotionsprozesses an sich wird durch eine solche Auswertung nicht direkt evaluiert. Differenziertere Arbeiten in dieser Richtung haben z.B. die Auswirkung des vom System produzierten Verhaltens auf das Stressniveau bzw. das Vertrauen / die Zuversicht der Benutzer bewertet [PMMI03] sowie den Einfluss des emotionalen Feedbacks auf das Lernverhalten der Benutzer evaluiert ([LBS99], [Ell97]). Ferner existiert eine beträchtliche Anzahl an Untersuchungen über die Wirkung des nonverbalen Verhaltens der virtuellen Charaktere auf den menschlichen Betrachter (z.B. [KTB03]). Allgemein konnte gezeigt werden, dass die Qualität der Interaktion im Hinblick auf Nutzerfreundlichkeit und Steigerung der Nutzermotivation durch den Einsatz von Emotion verbessert werden kann ([CK02]).

Teil III.

Anwendungen

8. Referenzmodell für Emotionsmodellierung

Die Modellierung und Diskussion der Emotionstheorien und Agentenarchitekturen in den vorangegangenen Kapiteln hat gezeigt, dass für eine bessere Aufbereitung und Klassifizierung der emotionstheoretischen Grundlagen und besonders für die Integration derselben in die Agenten- und Multiagentenforschung ein hinreichend ausdrucksstarkes Referenzmodell für Emotionsmodellierung notwendig ist.

Unter einem Referenzmodell wird hier eine Menge von aufeinander abgestimmten Beschreibungstechniken verstanden, mit der verschiedene Sichten auf Emotion dargestellt werden können. Der Zweck eines solchen Referenzmodells liegt in der Erschließung der Möglichkeit, verschiedene Theorien der Emotion im Kontext der Agentensysteme zu erforschen, zu klassifizieren, auf verschiedenen Ebenen eines Multiagentensystems zu interpretieren und evtl. zu integrieren.

Die Betrachtung der theoretischen Grundlagen der Emotion in Kapitel 6 ergab, dass Emotionstheorien verschiedener Disziplinen die Emotion auf unterschiedlichen Ebenen beschreiben, ohne dass in den Theorien explizit gemacht wird, auf welcher Betrachtungsebene sich die beschriebenen (Emotions-)Prozesse abspielen. Das Referenzmodell soll eine Strukturierung ermöglichen, durch die diese Beschreibungsebenen explizit benannt und referenziert werden können.

Die Diskussion der Architekturen emotionaler Agentensysteme in Kapitel 7 hat eine starke Forschungsfragmentierung und eine mangelnde theoretische Fundierung vieler Emotionsmodelle erkennen lassen. In diesem Zusammenhang soll das zu erstellende Referenzmodell einen gemeinsamen konzeptuellen und terminologischen Rahmen für die Darstellung der emotionstheoretischen Grundlagen und ihre Integration in die informatischen Emotionsmodelle bilden (vgl. Kapitel 7.4.2).

Letztendlich ist es das Ziel, ein Modell zur Verfügung zu stellen, das sowohl als konzeptionelle Darstellungstechnik zur Modellierung und zum Entwurf einer emotionalen Agentenarchitektur als auch als Referenzmodell zur Analyse, Klassifizierung und zum Vergleich verschiedener emotionaler Systemarchitekturen sowie zur Repräsentation emotionaler Prozesse auf verschiedenen Ebenen eingesetzt werden kann.

In diesem Kapitel stellen wir das kohärente Referenzmodell E-MULAN vor, das die Analyse der Emotion aus den genannten Perspektiven möglich macht. Dieses *Referenzmodell für die Emotionsmodellierung* wurde auf Basis der referenznetzbasierten Multiagentenarchitektur MULAN entwickelt, erlaubt jedoch neben der Modellierung der kognitiven auch eine explizite Modellierung der emotionalen Prozesse.

Zunächst werden die grundsätzlichen Annahmen und Ziele skizziert. Anschließend werden in Kapitel 8.2 das Referenzmodell und seine Implementierung vorgestellt. In Kapitel

8. Referenzmodell für Emotionsmodellierung

8.3 wird gezeigt, wie die in Kapitel 6 vorgestellten emotionalen Prozesse in E-MULAN modelliert werden können. In diesem Abschnitt betrachten wir die E-MULAN-Prozesse zur Laufzeit und veranschaulichen die Wechselwirkung der Referenzmodell-Elemente an ausgewählten Anwendungsszenarien. Im Gegensatz dazu wird in Kapitel 9 der PAOSE-Entwicklungsansatz zur Modellierung z.T. derselben Szenarien angewendet, jedoch auf einer höheren Abstraktionsebene. In Kapitel 8.4 wird die Verknüpfung rationaler und emotionaler Modellelemente diskutiert und die Erweiterung der MULAN-Architektur um die synchrone Kommunikation vorgeschlagen. Schließlich werden in Kapitel 8.5 die Eigenschaften des E-MULAN-Referenzmodells zusammengefasst und diskutiert.

8.1. Grundsätzliche Annahmen und Ziele

Mit einem Referenzmodell für die Emotionsmodellierung wird ein gemeinsames terminologisches Rahmenwerk zur Klassifizierung und Gegenüberstellung verschiedener Emotionstheorien und Emotionsmodelle angestrebt.

Im Folgenden werden die wesentlichen Zielsetzungen und Anwendungsmöglichkeiten eines solchen Referenzmodells skizziert und die grundlegenden Designentscheidungen bei der Entwicklung des Modells erläutert.

8.1.1. Definition der Modellebenen

Die in Kapitel 6 betrachteten Emotionstheorien bzw. -architekturen operieren auf verschiedenen Beschreibungsebenen. Diese Ebenen finden sich in den Emotionstheorien der Psychologie, Soziologie und Neurowissenschaften (vgl. die Kapitel 6.1, 6.2 und 6.3). Durch das Referenzmodell soll eine entsprechende Strukturierung und eine explizite Benennung dieser Beschreibungsebenen möglich werden.

Die Zuordnung der bisher betrachteten emotionstheoretischen Modelle zu den Beschreibungsebenen und die Überprüfung ihrer gegenseitigen Anschlussfähigkeit werden erst durch die Identifikation der verschiedenen Perspektiven auf Emotion und eine entsprechende explizite Definition der Ebenen des Referenzmodells möglich.

Petrinetze erlauben das Aufgreifen der Integration verschiedener Perspektiven auf Emotion und ermöglichen dadurch die Integration der Theorien verschiedener Disziplinen (vgl. Kapitel 4.4.2 für die theoretischen Modellierungsgrundlagen bzw. Kapitel 6.3.3 für ein Anwendungsbeispiel). Diese Zusammenführung der theoretischen Grundlagen auf Basis der Modellierungstechnik war einer der Gründe für die durchgehende Modellierung der Emotionstheorien / -modelle mit den Petri- / Referenznetzen. Nun soll auf Basis eines netzbasierten Referenzmodells die Möglichkeit der Integration der erstellten Modelle geschaffen werden. Damit wird gleichzeitig die Theorienintegration über die Ebenen hinweg ermöglicht.

Die besondere Herausforderung bei dem Versuch einer solchen übergreifenden Integration der Emotionstheorien / -modelle liegt in der Zusammenführung und Identifikation gleichartiger Begriffe, Konzepte, Prozesse, um sie deckungsgleich in den entsprechenden Modellen identifizieren und die ggf. vorhandenen Lücken in der Prozessbeschreibung als solche

erkennen zu können. Dafür benötigt man ein Denkmodell, das mit dem vorgeschlagenen Petrinetz-basierten Referenzmodell geschaffen werden soll.

Eine weitere Zielsetzung besteht darin, mit dem kohärenten Referenzmodell für Emotionsmodellierung eine konzeptionelle Basis für die Integration emotionstheoretischer Grundlagen in die informatischen Emotionsmodelle, insbesondere für (verteilte) künstliche Systeme, sogenannte emotionale Agentensysteme zu erzeugen (vgl. die allgemeinen Ziele dieser Arbeit, wie sie in Kapitel 1.3 dargestellt wurden).

Um eine informatische Ausrichtung des Modells zu erreichen, werden Vorschläge aus der Informatik zur Strukturierung verteilter komplexe Systeme hinzugezogen. Einer der Modellierungsvorschläge, die man dazu findet - im Agentenbereich und in Verbindung mit Petrinetzen - ist die MULAN-Architektur (siehe [KMR01b] zur konzeptionellen Architektur MULAN und [vLMV03] zur Einbettung in die Sozionik). Sie ist angelehnt an die informatische FIPA-Modellierung (siehe Kapitel 3).

Weitere Querbeziehungen bilden ein Vorschlag zur Modellierung der Web-Services in [MOO04] sowie eine organisationstheoretisch fundierte Referenzarchitektur zur Strukturierung von Systemen in [WEKBM08] (Organizational Architecture Nets - ORGAN).

In Anlehnung an diese Architekturvorschläge stellen wir die Hypothese auf, dass die Betrachtung der Emotion und ihrer Wirkungsweise auf vier verschiedenen Ebenen eine sinnvolle Modellierungsmöglichkeit darstellt. Die Aufteilung in vier Ebenen wurde ursprünglich für Multiagentensysteme entwickelt [Röl04] und hat sich in vielen Kontexten bewährt, insbesondere in den bereits erwähnten Bereichen der Web-Services [MOO04], der Modellierung von Organisationen (ORGAN -[WEKBM08]) und der Multiagentensysteme [Röl04]. Die Ebenen der hier vorgestellten Vier-Ebenen-Architektur zur Emotionsmodellierung (im Folgenden: E-MULAN) korrespondieren unmittelbar mit den entsprechenden Ebenen des Modellierungsrahmenwerks MULAN [Röl04].

Durch die Definition der vier Modellierungsebenen wird jedoch eine grundlegende konzeptuelle Einschränkung vorgenommen, da die Einordnung und Analyse der theoretischen Ansätze auf der Grundlage eben dieser Ebenen erzwungen wird, ohne dass ein Nachweis erbracht werden kann, dass die gewählte Aufteilung der Ebenen die einzig mögliche bzw. optimale Art der Aufteilung ist.

8.1.2. Interpretation von E-Mulan als Systemarchitektur und als Referenzmodell

Mit der E-MULAN-Architektur zielen wir darauf ab, ein Referenzmodell zur Verfügung zu stellen, das zum einen als konzeptionelle Darstellungstechnik zur Modellierung und zum Entwurf von emotionalen Agentenarchitektur eingesetzt werden kann, zum anderen aber auch als Referenzmodell zur Analyse, Klassifizierung und Vergleich verschiedener emotionaler Systemarchitekturen und zur Repräsentation emotionaler Prozesse auf verschiedenen Ebenen dienen kann.

In Anlehnung daran, wie die MULAN-Architektur einerseits als konkretes MAS-Modell, andererseits als Modellierungsbasis für ausführbare nebenläufige Agenten und schließlich als Referenzmodell für MAS interpretiert werden kann (vgl. Kapitel 4.2), wollen wir diese sowohl zu einer emotionalen Agenten- bzw. Multiagentensystemarchitektur weiterentwickeln,

8. Referenzmodell für Emotionsmodellierung

als auch zur Repräsentation und zum Vergleich emotionaler Zusammenhänge in Agenten bzw. in der Agentengesellschaft einsetzen. Als Beispiel kann hier die Cogaff-Architektur von Aaron Sloman dienen (vgl. Kapitel 7.2.4), die einerseits als Modellierungstechnik zum Entwurf einer konkreten Systemarchitektur, andererseits als konzeptionelles Rahmenwerk zur Klassifizierung und zum Vergleich verschiedener emotionaler Systemarchitekturen eingesetzt wird ([SS02], [Slo04], [GM09a]).

8.1.3. Analytische Trennung in rationale und emotionale Modellkomponenten

Neurophysiologische Untersuchungen (z.B. [LeD96b]) geben erste Hinweise darauf, dass emotionale und kognitive Verarbeitungsprozesse nach der Wahrnehmung eines Stimulus parallel ablaufen und sich gegenseitig ergänzen (vgl. Kapitel 6.2.1). Unter Berücksichtigung dieser Ergebnisse soll ein Referenzmodell zur Emotionsmodellierung die Repräsentation sowohl rationaler / kognitiver als auch emotionaler Prozesse ermöglichen und insbesondere die Modellierung ihres Zusammenspiels unterstützen. Die explizite Trennung in “rationale” und “emotionale” Modellkomponenten erlaubt einerseits eine klare konzeptionelle Unterscheidung zwischen rationalen und emotionalen Konzepten und Modellen, andererseits die Untersuchung ihrer Wechselwirkungen.

In den psychologischen und soziologischen Emotionstheorien wird jedoch eine solche Trennung zwischen emotionalen und rationalen Prozessen meist nicht explizit hervorgehoben. Sie ist fließend und variiert stark zwischen verschiedenen Theorien. So wird in den psychologischen Theorien unterschiedlich definiert, warum ein Ereignis, ein Gesichtsausdruck etc. als emotional bzw. rational interpretiert wird, ob die Einschätzungsprozesse, die an sich rein kognitive Verarbeitungsprozesse sind, als emotionale oder als rationale Prozesse zu sehen sind, ob eine Handlung bzw. Reaktion emotional bzw. rational bedingt ist etc.

Bei dieser unterschiedlichen Aufteilung der mentalen Verarbeitungsprozesse in emotionale und rationale Prozesse handelt es sich in der Regel um eine rein *analytische Trennung*: Diese Trennung spiegelt keine neurophysiologischen Prozesse wider, vielmehr wird versucht, durch die Aussonderung Aussonderung “rationaler” und “emotionaler” Prozess-Komponenten einen möglichen Erklärungsansatz für bestimmte mentale Phänomene aufzuzeigen.¹

Die Identifikation der emotionalen und der rationalen Komponenten des Modells ist somit ein willkürlicher und meist subjektiver Prozess. Dabei ist es von der modellierten Emotionstheorie, also von der Perspektive des Modellierers bzw. dem Modellierungswunsch abhängig, wie Emotion definiert wird und welche Komponenten als emotional betrachtet werden (vgl. Kapitel 5).

Die Herausforderung bei der Erstellung eines Referenzmodells für die Emotionsmodellierung besteht somit darin, ein generisches Modell zur Verfügung zu stellen, das die Abbildung eines möglichst breiten Spektrums von Aspekten und Betrachtungsweisen der Emotion entsprechend der verschiedenen Emotionstheorien und der verschiedenen Definitionen der Emotion bzw. des emotionalen Prozesses modellierungstechnisch ermöglicht.

¹In diesem Sinne ist die Verwendung dieser Eigenschaftsbezeichnungen im Folgenden zu verstehen. Sie werden daher zumeist nicht in relativierende Anführungszeichen gesetzt.

An dieser Stelle bieten die Petrinetze eine hervorragende Darstellungsmöglichkeit, denn Netzmodelle können mit fließenden Übergängen in rationale und emotionale Modellkomponenten aufgeteilt werden. So kann das rationale Modell in Richtung eines emotionalen Modells ergänzt werden, so dass durch die Auffaltung modellierungstechnisch exakt das dargestellt (‘‘herausgefaltet‘‘) wird, was bei der Modellierung als emotionale Modellkomponenten erkannt wurde (vgl. Kapitel 4.4.2) zur Faltung bzw. Entfaltung).

Durch die Faltung und Entfaltung von Netzen kann man verlustfrei Modelle bzw. Modellteile detailliert betrachten und eine Separierung von Modellteilen vornehmen, die man z.B. durch synchrone Kanäle wieder integrieren kann (vgl. Kapitel 4.4.2 zu synchronen Kanälen).

Die analytische Trennung von Kognition und Emotion dient somit in unserem Referenzmodell in erster Linie zu Forschungszwecken: Zum einen ermöglicht sie die explizite Darstellung und Erforschung der Wechselwirkungen dieser zwei Mechanismen, zum anderen ist dadurch die Auswertung des Potenzials der Emotion für die Optimierung von (zurzeit meist rationalen) (Multi-)Agentenarchitekturen möglich.

Wir halten jedoch Kognition und Emotion für zwei komplementäre und aufeinander bezogene Systemsichten, die in natürlichen intelligenten Systemen oft nicht klar voneinander abgegrenzt werden können vgl. dazu auch die neurowissenschaftlichen Sichtweisen in [Pes08], [DB07]). Die in dem vorgeschlagenen Referenzmodell erzwungene Trennung zielt vor allem auf die konzeptuelle Abgrenzung und Analyse der zugrunde liegenden komplementären Prozesse ab.

Deshalb muss im Referenzmodell die Möglichkeit vorgesehen sein, diese zwei Sichten auf der Modellierungsebene wieder zusammenzuführen. Diese Kopplung oder Synchronisation lässt sich bisher nicht über MULAN-Agenten modellieren, da diese ausschließlich asynchron kommunizieren.

Deshalb muss das herkömmliche MULAN-Modell, das bisher nur die asynchrone Kommunikation der Agenten unterstützt, um einen Synchronisationsmechanismus erweitert werden. Diese technische Erweiterung wurde in [FDM11] vorgestellt und wird in den Kapiteln 8.3.1 bzw. 8.4.3 beschrieben.

In Kapitel 8.3 wird die konkrete Umsetzung der Faltungs- und Entfaltungsmechanismen illustriert. In den Anwendungsbeispielen in Kapitel 9 wird sowohl das aufgefaltete Modell, bei dem rationale und emotionale Komponenten als zwei separate Agenten betrachtet werden (vgl. die Kapitel 9.2 und 9.3), als auch das zusammengefaltete Modell, in dem rationale und emotionale Komponenten in einem Agenten zusammengefasst sind (vgl. Abschnitt 9.5), angewendet.

8.1.4. Ein generisches Referenzmodell auf Basis der Mulan-Architektur

Auf der Basis der bestehenden (kognitiven) MULAN-Architektur wird im Folgenden ein eigenständiges konzeptionelles Referenzmodell E-MULAN entworfen, welches die Modellierung emotionaler Mechanismen auf verschiedenen Modellierungsebenen erlaubt und dadurch als Referenzmodell zur Darstellung verschiedener Emotionstheorien und Emotionsarchitekturen dienen kann. Die Entscheidung für die MULAN-Architektur als technische und konzeptionelle Basis ergibt sich zum einen aus der unmittelbaren Zielsetzung der Erweiterung von MULAN

8. Referenzmodell für Emotionsmodellierung

um die Emotionsmodellierung, zum anderen aus den guten Erfahrungen mit MULAN als Modellierungsrahmenwerk im Kontext der Modellierung von Organisationen [WEKBM08] und von Multiagentensystemen [Röl04].

Bislang lag der Schwerpunkt der Arbeiten zu MULAN weitgehend auf rationalen und/oder sozialen Modellen. Damit standen eher kognitiv ausgerichtete Modellierungsansätze im Vordergrund. Weil der bisherigen Architektur die Abbildung eines BDI-Agenten zugrunde liegt, der die Umweltinformationen durch kognitiven Verarbeitungsprozesse zu Agentenaktionen transformiert, werden Emotionen im einschätzungstheoretischen Sinne in diesem Bild implizit berücksichtigt, da sie nach Definition der Einschätzungstheoretiker aufgrund der kognitiven Bewertung von Umweltereignissen entstehen (vgl. die Übersicht zu einschätzungstheoretischen Emotionsdefinitionen in Kapitel 5.4). Somit bietet das bisherige MULAN-Agentenmodell als typische BDI-Agentenarchitektur keine explizite Abbildung der Emotionen, berücksichtigt diese jedoch implizit, etwa beim Entscheiden und bei der Handlungswahl.

Sofern man jedoch die Emotionen explizit betrachten möchte bzw. an den Wechselwirkungen zwischen emotionalen und rationalen Prozessen interessiert ist (vgl. Definition der Modellierungsziele in Kapitel 8.1), ergeben sich durch die Verwendung des konventionellen MULAN-Modells zur Emotionsmodellierung einige Schwierigkeiten. So ist es z.B. nicht möglich, die Wechselwirkungen zwischen den emotionalen und rationalen Protokollen, die in einem Agenten ablaufen, explizit zu betrachten. Ferner ist man gezwungen, die zur Emotionsrepräsentation notwendigen Architektur Erweiterungen in einem Agentenmodell auch dann mitzuführen, wenn keine Emotionsrepräsentation in dem zu implementierenden Anwendung benötigt wird. Die Komplexität des MULAN-Agentenmodells steigt in diesem Fall, ohne dass es sich auf die Anwendungen, die keine Emotionsimplementierung erfordern, gewinnbringend auswirkt. Des Weiteren bringt eine Veränderung der grundlegenden Architektur eines MULAN-Agenten diverse Nachteile in Bezug auf die Wiederverwendbarkeit der bisherigen Modelle bzw. die Agenten diverse Nachteile in Verwendung der Unterstützungswerkzeuge und Plugins bei der Erstellung eigener Anwendungen.

Sowohl für die konzeptionelle Modellierung als auch für die technische Implementierung, z.B. für Simulationen, wäre es also von Vorteil, eine emotionale Architektur zu konzipieren, welche die konventionelle MULAN-Architektur um die Möglichkeit der Emotionsmodellierung erweitert, ohne sie grundsätzlich zu verändern. Eine mögliche Umsetzung wird im folgenden Abschnitt vorgestellt.

8.2. Vorstellung des allgemeinen Referenzmodells

Das E-MULAN-Modell entsteht durch “Entfaltung” (vgl. Kapitel 4.4.2), d.h. durch die Ausgliederung der emotionalen Prozesse, die in MULAN implizit berücksichtigt werden, in eine getrennte Architektur. Durch die umgekehrte Operation - die Faltung - können beide Sichten zu jedem Zeitpunkt wieder in einer Architektur zusammengefasst werden. Wir können die existierende MULAN-Architektur weiterhin unverändert nutzen, um die “kognitiven” bzw. “rationalen” Komponenten des Anwendungsbereichs abzubilden. Die emotionale Shadow-Architektur zu MULAN wird gebildet, indem zu jeder rationalen /

8.2. Vorstellung des allgemeinen Referenzmodells

kognitiven Modellkomponente eine korrespondierende Komponente in Beziehung gesetzt wird, die die Emotionsmodellierung kapselt (vgl. Abb. 8.1).

So erhalten wir neben dem konventionellen MULAN-Agentenmodell ein strukturell gleiches Modell, das die emotionalen Prozesse eines Agenten abbildet. Die Kopplung der beiden unterschiedlichen Modellebenen (Rationalität und Emotion) erlaubt zum einen die theoretische Untersuchung wechselseitiger Zusammenhänge und Abhängigkeiten und zum anderen die technische Zusammenführung und somit die direkte Umsetzung emotionaler Modelle im Kontext von Multiagentensystemen. Durch eine gleichartige Grundstruktur der emotionalen und rationalen Architektur-Komponenten kann die Anbindung an die existierende MULAN-Architektur ohne wesentlichen zusätzlichen Aufwand erfolgen, sowohl für die konzeptionelle Modellierung als auch für die technische Implementierung.

Somit stellt die E-MULAN-Architektur eine spezielle Sicht auf das System dar, in die insbesondere emotionale Prozesse bzw. Verhaltensaspekte ausgegliedert werden, während “rationale“ / “kognitive“ Verhaltenskomponenten quasi weiterhin in der MULAN-Architektur verbleiben. Um diese beiden Modellkomponenten zu unterscheiden, werden wir im Folgenden die rationalen und die kognitiven Anteile der MULAN-Architektur als Teil des E-MULAN-Modells betrachten und als *rationales Modell* bezeichnen, während die Shadow-Architektur ein *emotionales Modell* genannt wird.

In [FM06] bzw. in [FM09] wurden die verschiedenen Möglichkeiten der Kopplung der Modellkomponenten der E-MULAN-Architektur diskutiert. Die Kommunikation zwischen den beiden Architekturkomponenten kann über eine enge oder eine lose Kopplung erfolgen (siehe Kapitel 8.4) oder - bei einer entsprechender technischer Erweiterung der MULAN-Architektur - über einen Synchronisationsmechanismus, der in Kapitel 8.4.3 vorgestellt wird.

Im Folgenden beschreiben wir die vier Ebenen von E-MULAN und erläutern ihre Beziehung zu den Modellierungsebenen der MULAN-Architektur. Die petrinetzbasierte MULAN-Architektur wurde in Kapitel 4.3 beschrieben. Auf diese Beschreibung Bezug nehmend, wollen wir hier aufzeigen, wie sie erweitert werden kann, um die Implementierung der Emotion auf allen vier Architekturebenen explizit zu machen.

Bei der Beschreibung der vier Modellierungsebenen werden wir versuchen, die MULAN-Terminologie weitestgehend beizubehalten. Jedoch ist die Fixierung der spezifischen emotionsbezogenen Modellaspekte vielerorts mit Schwierigkeiten verbunden, weil die vorgegebenen MULAN-Begriffe an manchen Stellen für die Beschreibung der emotionalen Mechanismen intuitiv nicht passen. An diesen Stellen werden wir eigene Begriffe definieren bzw. bestehende MULAN-Begriffe einschränken bzw. spezifizieren. Da es sich um ein generisches Modell handelt, werden wir mit dem soziologischen Begriff des Akteurs implizit sowohl menschliche Akteure als auch künstliche (Software-)Agenten bezeichnen.

Insgesamt werden in Anlehnung an MULAN vier Ebenen für unterschiedliche Abstraktionsgrade bei der Betrachtung der Emotionalität identifiziert. Jede Ebene umfasst in einer charakteristischen Weise Konzepte von Emotion. Zentral ist, dass für jede Ebene bestimmte Eigenschaften identifiziert werden können und jede Ebene in eine weitere eingebettet ist und / oder eine weitere Detaillierungsebene umfasst.

8. Referenzmodell für Emotionsmodellierung

Da die rein informatische Sicht auf Multiagentensysteme bereits bei der Betrachtung der soziologischen Grundlagen in Verbindung mit der Einbettung in die Sozionik nicht ausreichte (vgl. [vLMV03]), wurde in den letzten Jahren in zahlreichen Diskussionen eine soziologisch inspirierte Variante für das kognitive Modell entwickelt. Somit entsprechen sich folgende Begriffe: Multiagentensystem – Gesellschaft, Agentenplattform – Gruppe / Feld, Agentensystem – Akteur / Person / Individuum, Protokoll – internes Verhalten.

Die Abb. 8.1 zeigt die vier Hierarchiestufen eines Multiagentensystems in MULAN: *Agentensystem*, *Agentenplattform*, *Agent* und *Protokoll*, wobei zu jeder MULAN-Ebene eine korrespondierende Komponente des E-MULAN-Rahmenwerks in Beziehung gesetzt wurde.

Der Schwerpunkt der Modellbeschreibung liegt auf der Akteurs- / Agentenebene (vgl. Kapitel 8.2.3), da die meisten Aussagen der sozialwissenschaftlichen Theorien in diesem Bereich erfolgen und hier die für die Gesamtarchitektur des Modells grundlegenden Designentscheidungen getroffen werden. Die Verbindung zwischen den Modellkomponenten wird in den Kapiteln 8.4 und 8.4.3 diskutiert.

8.2. Vorstellung des allgemeinen Referenzmodells

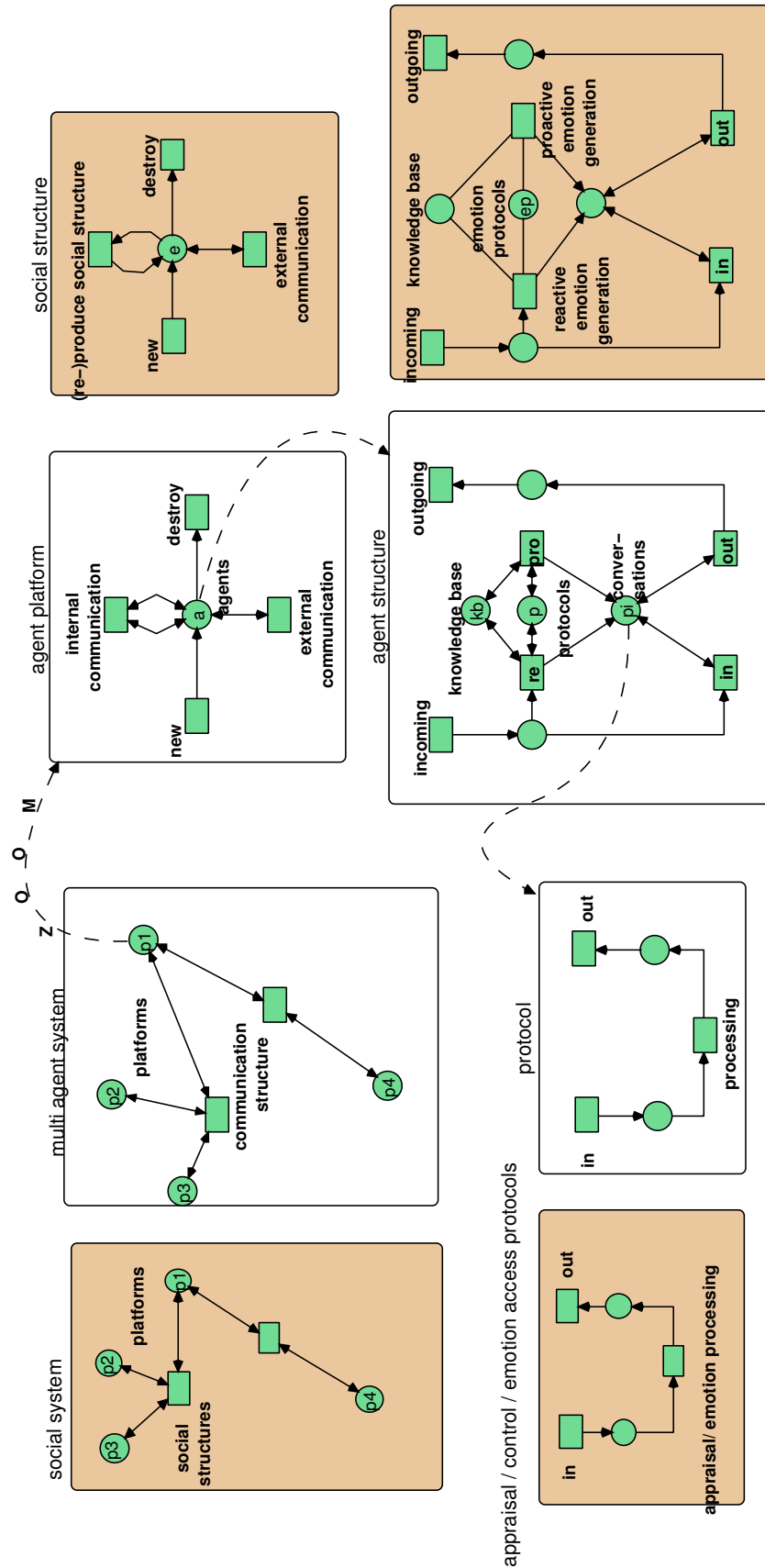


Abbildung 8.1.: E-MULAN - eine *Shadow-Architektur* zu *Mulan*

8. Referenzmodell für Emotionsmodellierung

8.2.1. Gesellschaftsebene

Die Gesellschaftsebene spezifiziert die interne Struktur einer Gesellschaft. Sie ist durch die sie konstituierenden sozialen Aggregate und ihre Wechselbeziehungen untereinander charakterisiert. Auf der Gesellschaftsebene können verschiedene soziale Einheiten modelliert werden, welche die soziale Strukturen und Prozesse der Gesellschaft konstituieren.

Diese sozialen Einheiten können in Bezug auf ihre Interaktion (oder Kommunikation) zueinander in Beziehung gesetzt werden. Entscheidende Eigenschaften, die auf dieser Ebene modelliert werden, sind also die Verbindungen zwischen den Einheiten und die durch sie ermöglichten Prozesse. Die jeweiligen Einheiten werden auf der nächsten Ebene verfeinert und durch ihre Einbettung in ihre jeweilige Infrastruktur eingebunden.

Die auf dieser Ebene modellierten Anteile stellen im Wesentlichen die unveränderlichen Anteile eines Modells dar. Aus der soziologischen Sicht handelt es sich um prinzipielle Relationierungen der verschiedenen sozialen Gruppen (und Akteure). Wie eine Gruppe interagiert, wie sie sozial eingebettet ist und welche sozialen Beziehungen und Zusammenhänge für sie bestehen, wird auf der Gesellschaftsebene festgelegt. In konkreten Kontexten wird dies durch Institutionen, Gesetze, kulturelle, religiöse, ethische und andere fest etablierte Ordnungen und Verhaltensvorschriften festgehalten (siehe [Lan05]).

Auf der Gesellschaftsebene wird das soziale System repräsentiert, indem die sie konstituierenden sozialen Aggregate (durch Stellen) und ihre Wechselbeziehungen (durch Transitionen) spezifiziert werden. Durch die Netzstruktur der Gesellschaftsebene wird die Interaktions- und Kommunikationsinfrastruktur festgelegt, über die sich die Einheiten der nächsttieferen *Plattformebene* austauschen, koordinieren oder bewegen können. Diese Einheiten werden in MULAN "Plattformen" genannt und entsprechen konzeptionell den internen sozialen Strukturen von Aggregaten der Gesellschaftsebene, durch die sie konstituiert werden. Sie repräsentieren also soziale (Kontroll-)Strukturen wie z.B. soziale Normen, Verhaltensregeln und andere (habituierte) Instrumente sozialer Kontrolle, die durch soziale Aggregate definiert und reproduziert werden.

Plattformen bzw. soziale Strukturen können als Marken auf den Stellen der Gesellschaftsebene liegen. Werden z.B. die sozialen Gruppen einer Gesellschaft durch die Stellen der Gesellschaftsebene modelliert, so werden die jeweiligen sozialen Normen jeder Gruppe als Marke auf der entsprechenden Gruppen-Stelle dargestellt. Sie werden auf der Plattformebene verfeinert und durch ihre Einbettung in ihre jeweilige Infrastruktur eingebunden.

In Bezug auf das emotionale Modell werden auf dieser Ebene die gesellschaftlich tradierten Werte und Verhaltensweisen in Bezug auf Emotion festgehalten. Das sind z.B. kodierte Symbole wie Weiß oder Schwarz für das Zeigen von Trauer, die gesetzlichen Regelungen zu Fragen der Religionsfreiheit, der Volljährigkeit oder gesellschaftlich akzeptabler Verhaltensweisen (z.B. das Gesetz über die "Erregung öffentlichen Ärgernisses", 183a StGB).

Ein Spezialfall der auf dieser Ebene modellierten Einheiten sind die gesellschaftlichen Konventionen der Emotionsgenerierung bzw. des Emotionsausdrucks. Unterschiedliche Konventionen im Umgang mit Emotion in verschiedenen Gesellschaftsschichten können kulturell oder ethisch bestimmt sein und passen sich bei dem Übergang von einer Gesellschaft(sschicht) in eine andere der neuen gesellschaftlichen Situation an (vgl. z.B. [Bou70]).

8.2.2. Soziale Struktur

Auf dieser Ebene werden die verfestigten sozialen Strukturen einer Gruppierung von Akteuren spezifiziert. Diese Strukturen legen den Rahmen für die wechselseitige Beziehungen der Akteure fest. Die Ebene der sozialen Struktur wird durch drei Beziehungstypen beschrieben: die Art der Einbettung der Akteure und der sozialen Struktur in die Gesellschaft, die Bereitstellung der Einbindung der jeweiligen Akteure in die soziale Struktur und schließlich die interne Organisation, die innerhalb der sozialen Struktur die Relationen der in ihr enthaltenen Akteure festlegt. Hierbei kann es sich um zwei oder mehrere Personen bis hin zu sozialen oder ethnischen Gruppen handeln.

Die bereitgestellten Dienste der sozialen Struktur ermöglichen die Handlungsmöglichkeiten der Akteure oder schränken sie ein. Beispiele sozialer Strukturen sind soziale Normen, Rollen oder explizite und implizite Verhaltensregeln innerhalb einer Gruppe oder einer Gesellschaftsschicht.

Der Ebene der sozialen Struktur entspricht die Plattformebene von MULAN. Die Plattform stellt einen physikalischen oder logischen Ort dar, den Agenten betreten und verlassen können bzw. in dem sie installiert oder gelöscht werden können. Entsprechend können Akteure einer sozialen Struktur beitreten, z.B. indem sie ihre gesellschaftliche Position ändern und die Normen einer neuen Gesellschaftsschicht befolgen (Agent betritt die Plattform). Beim Verlassen der Gruppe, zu der sie früher gehört haben, werden Akteure die Normen und Regeln dieser Gruppe nicht mehr befolgen müssen (Agent verlässt die Plattform).²

In Bezug auf die Emotionsmodellierung sind insbesondere solche sozialen Strukturen von Interesse, die soziale / emotionale Wechselbeziehungen zwischen Akteuren / Agenten festlegen. Diese sozialen Strukturen üben Einfluss auf die Emotionen der einzelnen Akteure aus und werden ihrerseits durch Emotionen der Akteure beeinflusst und zum Teil konstituiert. Sie sind für alle Akteure / Agenten bindend / maßgebend, die einer natürlichen / künstlichen Gesellschaft angehören, in der sich die jeweiligen sozialen Strukturen etabliert haben. Beispiele solcher "sozial-emotionalen" Strukturen sind etwa die Regeln der Emotionsexpression ("feeling rules", vgl. Kapitel 6.3.2), emotionsbasierte Etablierung und Aufrechterhaltung sozialer Normen (siehe Kapitel 6.3.3) oder Gruppenemotionen.

Ferner wird die Interaktion der Akteure durch die gegebene soziale Struktur bestimmt. Die interne Kommunikation zwischen den Mitgliedern einer sozialen Struktur (Gruppe, Schicht etc.) wird durch diese Struktur festgelegt, während die externe Kommunikation zwischen Mitgliedern verschiedener sozialer Strukturen durch die Kommunikations- und Migrationsinfrastruktur der darüber liegenden Gesellschaft(sebene) determiniert ist. Bei dem Referenzmodell bezieht sich die interne und externe Kommunikation insbesondere auf die Emotionskommunikation und Emotionswahrnehmung durch die anderen Akteure. Die entsprechenden Interaktions- und Kommunikationsmuster werden durch die (Struktur

²Dieses Erweitern von Normen bzw. das Nicht-Befolgen von Normen stellt einen schwierigen Übergang dar. Im Normalfall werden die Akteure die jeweils internalisierten Regeln und Vorgaben einer sozialen Struktur (Gesellschaft(sschicht), die sie verlassen haben, wenn überhaupt, nicht sofort ablegen können. Darunter fällt beispielsweise der Neue in einem Feld (siehe [Bou70]) oder das Ausgrenzen von Fremden (siehe [ES90]).

8. Referenzmodell für Emotionsmodellierung

der) Plattform definiert und gelten für alle Akteure / Agenten, die sich auf der Plattform befinden.

8.2.3. Akteurs- / Agentenebene

Diese Ebene spezifiziert den internen Aufbau eines (emotionalen) Akteurs / Agenten. Dieser Aufbau umfasst die Anbindung an die sozialen Strukturen, in die der Akteur / Agent eingebettet ist, die interne Organisation des Akteurs (bzw. die Agentenarchitektur) und die Art und Weise, wie die internen Emotionen des Akteurs / Agenten eingegliedert sind. Jeder Akteur / Agent muss in mindestens eine soziale Struktur (Plattform) eingebettet sein. Diese bestimmt, welche Normen, Rollen usw. ihm zur Verfügung stehen bzw. sein Verhalten einschränken.

Das beste Beispiel für einen Akteur / Agenten ist ein Mensch. Andere soziale Einheiten, z.B. Gruppen, lassen sich ebenfalls als Agenten auffassen. Dieser allgemeine Betrachtung wird weiter unten genauer erläutert (vgl. Kapitel 8.4).

Auf der Akteurs- / Agentenebene wird das Emotionsmodell eines Agenten abgebildet. Das interne Emotionsmodell der Agenten- / Akteurebene legt fest, auf welche Weise Emotionen generiert werden. Wie bei allen anderen Ebenen des Referenzmodells kann das abgebildete Emotionsmodell entsprechend der jeweils anzuwendenden bzw. zu modellierenden Emotionstheorie ausgetauscht werden.³

Aus emotionspsychologischer Sicht kann das Emotionsmodell in zwei grundsätzliche Komponenten aufgeteilt werden (vgl. Kapitel 5): zum einen das Modell der *Emotionsgenerierung*, das sogenannte *Einschätzungsmodell*, das interne Mechanismen spezifiziert, die der Entstehung der Emotionen zugrunde liegen (z.B. Einschätzungsmechanismen); zum anderen das Modell der *emotionsbasierten Steuerung* (oder *Emotionssteuerung*), das die Auswirkungen der Emotionen auf Aktionswahl, Planen, Informationsverarbeitung usw. spezifiziert. Eine ähnliche Aufteilung findet man auch bei anderen Emotionsarchitekturen (vgl. z.B. Kapitel 7.2.1).

In Kapitel 8.3 wird die Abbildung dieser zwei grundsätzlichen Modellkomponenten - Emotionsgenerierung und emotionsbasierte Steuerung - auf der E-MULAN-Agentenebene illustriert.

Um das Modell der Emotionsgenerierung in E-MULAN abzubilden, muss gezeigt werden, wie die in Kapitel 6.1 beschriebenen Mechanismen der Emotionseinschätzung und die in Kapitel 5 dargestellten Emotionsklassen mit E-MULAN-Konzepten modelliert werden können. Im Allgemeinen können zwei grundsätzliche Arten der Emotionsgenerierung unterschieden werden (vgl. Kapitel 5.6):

- **Reaktive Emotionsgenerierung:** Eine Emotion wird durch den Agenten als Ergebnis der Bewertung von Umweltereignissen generiert. Die Bewertung kann dabei sehr rudimentär sein, z. B. wenn im Falle der Basisemotionen die spezifischen Eigenschaften der Umweltsituation, bzw. im Falle der sozialen Emotionen die vorgegebenen

³Das Emotionsmodell auf der Agentenebene kann z.B. entsprechend der Emotionstheorie von Ortony et al. [OCC88], Scherer [Sch93] oder beliebiger anderer psychologischer Emotionstheorie gestaltet sein.

sozialen Verhaltensnormen (vgl. Ebene der sozialen Struktur), automatisch emotionale Reaktionen auslösen (vgl. Kapitel 5.2.1).

Der Bewertungsprozess kann aber auch beliebig komplex sein und höhere Kognition erfordern, z.B. bei der Bewertung der Relevanz des Ereignisses für die Ziele des Agenten bzw. seine Kompatibilität mit aktuellen Überzeugungen (Beliefs) (vgl. Kapitel 6.1.2). Ein gemeinsames Merkmal der reaktiven Emotionsgenerierung ist jedoch immer ein externes Ereignis in der Umwelt des Agenten, das entweder zur automatischen Emotionsgenerierung oder zur Auslösung des Einschätzungs- / Bewertungsprozesses führt. Die Umweltereignisse werden hierbei in gleicher Weise wie im konventionellen MULAN-Modell durch ankommende Nachrichten repräsentiert.

Die durch reaktive Emotionsgenerierung erzeugten Emotionen haben wir als *reaktive Emotionen* bezeichnet (vgl. Kapitel 5.6). Entsprechend unserer Definition von reaktiver Emotionsgenerierung handelt es sich bei den reaktiven Emotionen um zwei strukturell ungleiche Emotionsarten: zum einen um unmittelbare automatische Reaktionen auf bestimmte (kritische) Ereignisse in der dynamischen Umgebung des Agenten (in der Literatur je nach Emotionsdefinition als primäre, elementare oder Basisemotionen bezeichnet, vgl. Kapitel 5); zum anderen um komplexe deliberative Prozesse der Einschätzung der Agentenumwelt in Relation zu seinen Beliefs, Desires, Intentions, Umgebungsbeobachtungen etc.

- **Proaktive Emotionsgenerierung:** Emotionen entstehen als Ergebnis der Deliberationsprozesse, die in einem Agenten entweder kontinuierlich ablaufen oder z.B. durch die interne Veränderung der aktuellen Abbildung der Umwelt (etwa infolge von Wissensbasisaktualisierung / -veränderung) aktiviert werden. Für die Emotionsgenerierung ist in diesem Falle kein externer Auslöser erforderlich. Auf dem Wege der proaktiven Emotionsgenerierung erzeugte Emotionen haben wir als *proaktive Emotionen* bezeichnet.

Zumindest diese beiden Arten der Emotionsgenerierung sollen mit Hilfe des Referenzmodells abgebildet werden können, sofern das Modell für die Repräsentation psychologischer und sozialwissenschaftlicher Emotionstheorien eingesetzt werden soll. Die Abbildung geschieht mit Hilfe der sogenannten Einschätzungsprotokolle, die die Prozesse der Emotionseinschätzung/-generierung spezifizieren. Die Realisierung des Modells der Emotionsgenerierung durch die Einschätzungsprotokolle wird in Kapitel 8.3 beschrieben.

Agenteninterne Steuerungsfunktionen der Emotionen, d.h. die vielfältigen Auswirkungen der Emotion auf die akteurs- / agenteninternen Verarbeitungsprozesse, werden ebenfalls auf der Akteurs- / Agentenebene spezifiziert. In Bezug auf Agenten wird hier von internen, verfestigten Ordnungen innerhalb der Verarbeitung und des Wissens gesprochen. Generelle Modelle der Funktionalität der Emotion bei der Entscheidungsfindung, Aktionswahl, Informationsverarbeitung usw. werden auf der Akteurs- / Agentenebene verankert und in die Agentenarchitektur integriert. Durch die Verlinkung zum kognitiven Agentenmodell (MULAN) können die Einflüsse der Emotion auf die Deliberation (Verhalten, Aktionswahl, Planen usw.) spezifiziert und gezielt untersucht werden, so weit diese Einflüsse in der zu modellierenden Emotionstheorie betrachtet werden.

8. Referenzmodell für Emotionsmodellierung

Die Managementfunktion von Emotionen besteht insbesondere in der emotionsbasierten Steuerung des Agentenverhaltens, wobei unter "Verhalten" verschiedenartige kognitive Prozesse wie Aktionswahl, Planen, Entscheidungsfindung, Informationsverarbeitung usw. zusammengefasst werden. Diese emotionsbasierte Steuerung kann auf viele verschiedenen Weisen in E-MULANrepräsentiert werden, z.B. durch spezielle Emotionsprotokolle (emotionale Steuerungsprotokolle, siehe Kapitel 8.2.4), durch Festlegung der Relation zwischen Emotion und Verhalten in der Agentenwissensbasis oder auf der Ebene der Decision Components (siehe Kapitel 8.3.1). Die verschiedenen Modellierungsmöglichkeiten der Steuerungsfunktion der Emotion werden in Kapitel 8.4.2 diskutiert und in Kapitel 9.4 an einer Beispielanwendung illustriert.

Insbesondere kann die Steuerung des rationalen Modells durch emotionale Steuerungsprotokolle erfolgen. Nach den theoretischen Ansätzen von [BDD00], [FS87], [Sch88] und [Laz91] (vgl. Kapitel 6) wird der Entscheidungsprozess bzw. die Handlungswahl eines Akteurs in hohem Maße durch seine persönlichen emotionalen Erfahrungen gesteuert. Dabei werden soziale Normen (spezifiziert auf der Ebene der sozialen Struktur) integriert und damit tradiertes / überliefertes Wissen um Gruppennormen und Verhaltensvorschriften verinnerlicht. Diese aus der Gruppennorm übernommenen (internalisierten) Anteile bilden ein wichtiges Entscheidungskriterium und schränken die Handlungsmöglichkeiten eines Akteurs dahingehend ein, dass nur von der Gruppe akzeptierte Handlungen (spezifiziert auf der Protokollebene) zur Ausführung kommen (vgl. Kapitel 6.3.3).

So kann ein Steuerungsprotokoll den Ablauf eines rationalen Protokolls beeinflussen, beispielsweise die Aktivierung eines bestimmten Verhaltensprotokolls verhindern oder herbeiführen. Z.B. kann ein Emotionsprotokoll für Abscheu, Verachtung oder Wut (gegen einen anderen Agenten) die Auslösung der rationalen Protokolle für kooperatives Verhalten verhindern (siehe hierzu auch Kapitel 6.3.3).

Durch Spezifikation der Emotion auf der Agentenebene können nicht nur die Wechselwirkungen zwischen emotionalen und deliberativen Prozessen auf eben dieser Architekturebene analysiert werden, sondern auch die Wechselwirkungen der Emotionen mit sozialen Strukturen und Gesellschaft, also mit den höheren Hierarchieebenen untersucht werden (s. hierzu auch Abschnitt 8.4.2).

8.2.4. Emotionsebene

Auf der vierten Modellierungsebene werden die akteurs-internen Emotionen repräsentiert. Wir haben in Kapitel 5 Emotionen als komplexe natürliche Mechanismen definiert, die in einem natürlichen intelligenten System als Resultat der Einschätzung seiner Umweltsituation entstehen und das System in einen Zustand versetzen, in dem es auf seine Umwelt bestmöglich reagieren kann (*emotionaler Zustand*).

Auf der Emotionsebene sollen also sowohl die emotionalen *Prozesse*, die in einem (emotionalen) Agenten ablaufen, als auch die emotionalen *Zustände*, die ein Akteur / Agent einnimmt, modelliert werden können (die grundlegende Dichotomie emotionaler Zustand vs. Emotionsprozess wird in Kapitel 5.2.3 diskutiert).

Um diese emotionalen Prozesse und Zustände der Emotionsebene genauer zu spezifizieren, greifen wir unsere Emotionsdefinition aus Kapitel 5 erneut auf. Danach entsteht eine

8.2. Vorstellung des allgemeinen Referenzmodells

Emotion als Ergebnis der Emotionseinschätzung und sie versetzt das System in einen emotionalen Zustand, der diverse interne Verarbeitungsprozesse, insbesondere das Verhalten, das Handeln, die Wahrnehmung usw. beeinflusst.

Alle Arten von Einheiten, die auf der Emotionsebene modelliert werden, weisen folgendes charakteristisches Merkmal auf: Sowohl emotionale Zustände als auch emotionale Prozesse werden agentenintern aufgebaut und sind nicht transferierbar, d.h. sie sind im Agenten gekapselt und können nur durch die Schnittstelle auf der Agentenebene nach außen gelangen. Wenn durch die Agentenebene ein Mensch modelliert wird, entspricht dies der Vorstellung von LeDoux [LeD98], dass menschliche emotionale Zustände und Prozesse für einen Außenstehenden nicht direkt greifbar / beobachtbar sind, sondern nur indirekt - etwa durch Gestik, Mimik, spezifisches Verhalten usw. - sichtbar werden können.

Die unter dieser Voraussetzung zu modellierenden Emotionen sind strukturell und konzeptionell äquivalent zu den dynamisch generierten Verhaltensweisen / Aktionen eines Agenten. Die Verhaltensweisen werden in dem konventionellen MULAN-Modell durch "Protokolle" modelliert, die Zustände werden durch die Wissensbasis abgebildet. Der Protokollbegriff wird nun auch auf die emotionalen Prozesse und Zustände der Emotionsebene ausgeweitet, d.h. dass wir die verschiedenartigen Einheiten der Emotionsebene zusammenfassend als *Emotionsprotokolle* bezeichnen werden.

Durch Emotionsprotokolle werden sowohl die emotionalen Zustände als auch die Prozesse der Emotionsgenerierung und der emotionsbasierten Verhaltenssteuerung, d. h. des Einflusses der generierten Emotionen auf die Verhaltens-, Planungs-, Entscheidungsprozesse usw. modelliert. Es werden im Folgenden drei Arten von Emotionsprotokollen unterschieden, wobei sich diese Klassifizierung nicht auf die strukturellen Eigenschaften der Protokolle, sondern auf ihre semantische Bedeutung, d.h. ihre Zweckbestimmung bzw. ihre Funktion im Modell bezieht:

- **Emotionszugriffsprotokolle:** Durch Emotionsprotokolle dieser Art wird der emotionale Zustand eines Agenten abgefragt, gesetzt oder geändert. Jede Änderung des emotionalen Zustandes eines Agenten durch ein Emotionszugriffsprotokoll entspricht der Generierung einer Emotion im einschätzungstheoretischen Sinne.
- **Einschätzungsprotokolle:** Durch diese Emotionsprotokolle werden die in verschiedenen Emotionstheorien beschriebenen Prozesse der Emotionsgenerierung modelliert. Insbesondere in Einschätzungstheorien wird Emotionsgenerierung als Bewertung der Umweltsituation bezüglich ihrer Konformität mit den internen Agentenzielen und -absichten – dem sogenannten Einschätzungsprozess – beschrieben (daher auch der Name dieser Emotionsprotokollart).
- **Emotionale Steuerungsprotokolle:** Diese Emotionsprotokolle repräsentieren die emotionale Steuerung verschiedener Aspekte des Agentenverhaltens, d.h. sie modellieren den Emotionseinfluss auf die Wahrnehmung, die Handlungswahl, die Informationsverarbeitung, die Entscheidungs- und Planungsprozesse usw., was wir oben zusammenfassend als Agentenverhalten bezeichnet haben. Wir nennen solche Emotionsprotokolle deshalb auch *Verhaltenssteuerungsprotokolle*. Die Beziehung zwischen

8. Referenzmodell für Emotionsmodellierung

den Einschätzungsprotokollen und den emotionalen Steuerungsprotokollen muss genau spezifiziert werden, d.h. es muss festgelegt werden, welche Verhaltenssteuerungsprotokolle durch welche Einschätzungsprotokolle ausgelöst bzw. beeinflusst werden. Hier kommen insbesondere psychologische Einschätzungstheorien in Frage (z.B. [Sch99a], [FKT89], [Laz91]), die beschreiben, wie sich eine eingeschätzte Emotion auf Planung, Aktionswahl, Perzeption, Coping etc. auswirkt.

In der MULAN-Architektur stellen Protokolle die *Ablaufschemas* dar, die Struktur von Verhalten modellieren und ein Muster für die Generierung der aktiven *Protokollinstanzen* - in MULAN *Konversationen* genannt - bilden. In unserem Referenzmodell für Emotionsmodellierung verstehen wir Emotionsprotokolle ebenfalls als Templates, die statische *Schemata* festlegen, nach denen dynamisch generierte *Emotionsprotokollinstanzen* gebildet werden. Jedoch müssen wir zusätzlich zwischen *Zustandsschemata* bzw. -*instanzen* für die emotionalen Zustände und *Prozessschemata* bzw. -*instanzen* für die emotionalen Prozesse unterscheiden.

Bei den emotionalen Prozessen geben die Einschätzungs- bzw. Verhaltenssteuerungsprotokolle die Struktur vor, nach der ihre entsprechenden Protokollinstanzen generiert und abgearbeitet werden. So geben die Einschätzungsprotokolle "Einschätzungsmuster" vor, die den Ablauf der Einschätzungsprozesse, modelliert durch die dynamisch generierten Einschätzungsprotokollinstanzen, definieren.

Bei den emotionalen Zuständen entspricht die Initialmarkierung bzw. die vordefinierte Datenstruktur in der Agentenwissensbasis einem Zustandsschema. Die aktuelle Markierung / Variablenbelegung stellt die Instanziierung des Zustandsschemas dar (siehe Kapitel 9 für eine ausführliche Darstellung).

Im Falle des emotionalen Agenten werden also nicht nur seine Aktivitäten, sondern auch seine Emotionen mit Protokollen modelliert. Dazu wird sein Handlungsrepertoire (seine Wissensbasis) um Emotionszugriffsprotokolle, Einschätzungsprotokolle, die die Emotionsgenerierung spezifizieren, sowie emotionale Steuerungsprotokolle, die den Einfluss der Agentenemotionen auf sein Verhalten repräsentieren, ergänzt. D.h., netztechnisch unterscheiden sich die neu eingeführten emotionalen Protokolle nicht grundsätzlich von den konventionellen "rationalen" Protokollen, da in beiden Fällen mit dem Protokoll eine Aktionssequenz vorgegeben wird, die zur Laufzeit abgearbeitet werden muss, ganz gleich, ob damit ein "rationaler" oder "emotionaler" Prozess modelliert wird.

Durch die Shadow-Struktur des E-MULAN-Modells ist es möglich, die emotionalen und die kognitiven Prozesse konzeptionell zu trennen, um z.B. die Beziehung zwischen den beiden Modellkomponenten explizit zu betrachten oder das Modellverhalten mit und ohne die emotionale Agentenkomponente zu vergleichen und auszuwerten. Dazu werden die Prozesse der Emotionsgenerierung und ggf. die emotionsbasierte Steuerung in der emotionalen Shadow-Komponente des Agenten gekapselt. Dies wird im Folgenden deshalb oft *emotionale Agentenkomponente* bzw. vereinfacht *emotionaler Agent* genannt, wobei die letzte Bezeichnung formal nicht korrekt ist, weil zu einem emotionalen Agenten im herkömmlichen Sinne eigentlich beide Shadow-Komponenten der Agentenebene gehören, sowohl "rationale" als auch "emotionale" (vgl. Kapitel 1.5).

Im nächsten Kapitel wird die Modellbildung mit Hilfe der E-MULAN-Referenzarchitektur an einigen Beispielen illustriert und anschließend diskutiert.

8.3. Modellierung mit E-Mulan

In diesem Abschnitt zeigen wir, wie die in Kapitel 6.1 beschriebenen emotionalen Prozesse, insbesondere die kognitive Emotionseinschätzung und die emotionsbasierte Verhaltenssteuerung, mit dem E-MULAN-Referenzmodell modelliert werden können.

In den Netzmodellen dieses Abschnittes werden zuerst nur strukturelle Abläufe ohne technische Details dargestellt. Dazu wird ein stark vereinfachtes, aber ausführbares MULAN-Modell verwendet, das in [Ree10], S. 96ff., vorgestellt wurde. Mit dem Modell lassen sich die komplexen Interaktionsabläufe zwischen Modellelementen auf verschiedenen Architekturebenen besser nachvollziehen als mit dem konventionellen MULAN-Modell, das durch die vielen technischen Erweiterungen nicht mehr transparent genug ist. Ferner können die internen Prozesse zur Laufzeit mit einem Kausalnetz veranschaulicht werden (vgl. [Ree10], S. 100), so dass sie visualisiert und der Systemzustand zu einem beliebigen Simulationszeitpunkt beobachtet werden kann.

Das konventionelle MULAN-Modell wird später in Kapitel 9 in Verbindung mit dem PAOSE-Modellierungsansatz angewendet. Dabei findet die prozessorientierte Anwendungsentwicklung in Form von Interaktionsprotokollen zeitgleich mit der strukturorientierten Anwendungsentwicklung in Form von Agenten zur Entwurfszeit statt. Mit dem PAOSE-Ansatz werden z.T. die gleichen Anwendungsszenarien modelliert, die in diesem Abschnitt betrachtet werden, aber auch einige weitere Anwendungen hinzugenommen.

Durch die Faltung / Entfaltung kann das Agentenmodell in E-MULAN jederzeit in rationale und emotionale Modellkomponenten aufgeteilt werden (siehe Abb. 8.2). Je nach Modellierungsgegenstand findet entweder das aufgefaltete Modell Anwendung, bei dem rationale und emotionale Komponenten als zwei separate Agenten betrachtet werden (die Anwendungsmodellierung in den Kapiteln 9.2, 9.3 und 9.4), oder das zusammengefaltete Modell, in dem das rationale und das emotionale Modell in einem Agenten zusammengefasst sind (vgl. Kapitel 9.5).

Die in Kapitel 4.3 beschriebenen Grundelemente der konventionellen MULAN-Architektur wurden in Kapitel 8.2 verwendet, um die grundlegenden Konzepte des E-MULAN-Referenzmodells zu definieren. In diesem Abschnitt wird ferner folgende Terminologie verwendet:

- Als *Emotionsprotokollen* werden die drei grundlegenden Arten der Protokolle der E-MULAN-Emotionsebene - Einschätzungsprotokolle, Emotionszugriffsprotokolle und emotionale Steuerungsprotokolle, - zusammenfassend bezeichnet (vgl. Kapitel 8.2.4). Letztere werden an einigen Stellen auch Verhaltenssteuerungsprotokolle bzw. einfach nur Steuerungsprotokolle genannt.
- Die beiden Komponenten der E-MULAN-Agentenebene werden oft *rationale Agentenkomponente* und *emotionale Agentenkomponente* (auch Shadow-Agentenkomponente) genannt (vgl. Abb. 8.2). Bei der Beschreibung der Interaktionen der beiden Agentenkomponenten in Zusammenhang mit der Anwendungsmodellierung in den Kapiteln 9.2 und 9.3 wird auch vereinfacht von einem rationalen Agent und einem emotionalen Agenten gesprochen. Wenn die Akteurs- / Agentenebene der E-MULAN-Architektur

8. Referenzmodell für Emotionsmodellierung

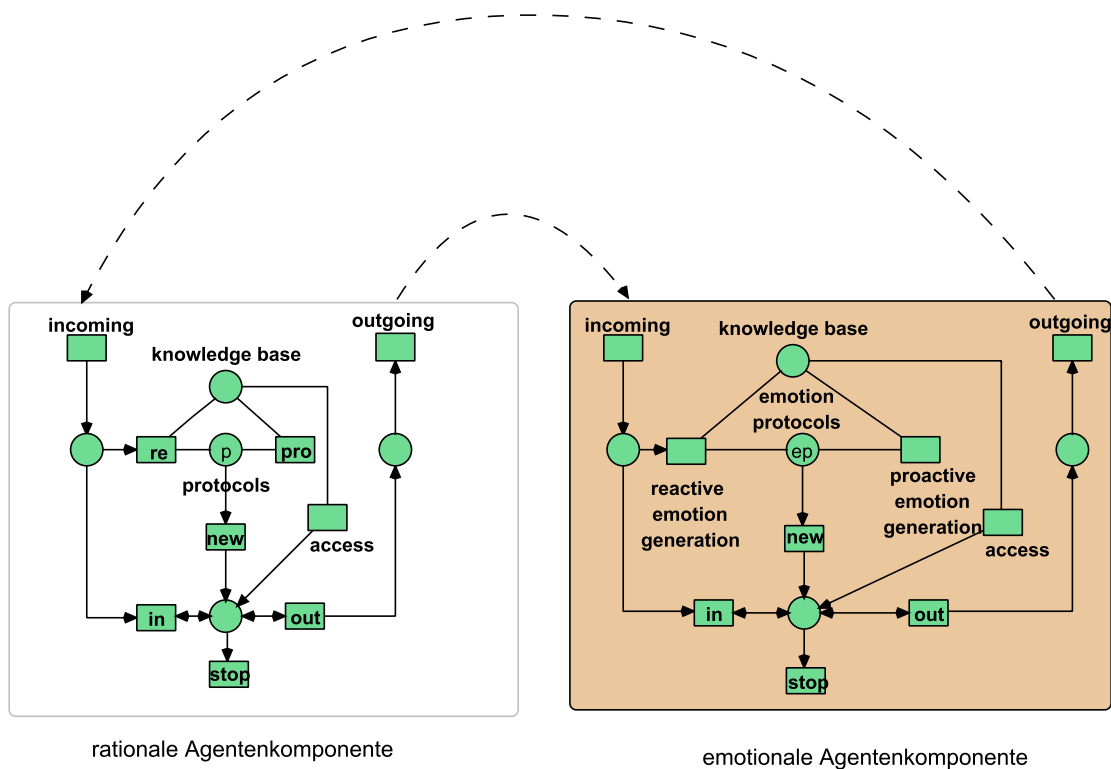


Abbildung 8.2.: MULAN-Agent und seine emotionale Shadow-Komponente

als eine Gesamtheit, d.h. als Zusammenfassung der beiden Teilkomponenten, angesprochen wird, wird diese oft einfach nur als Agent bezeichnet.

Zuerst werden wir zeigen, wie das Modell der Emotionsgenerierung in E-MULAN abgebildet werden kann (Kapitel 8.3.1). Das Einschätzungsmodell wird schrittweise auf Basis der einfachen Petrietz-Darstellungen in Kapitel 5 ausgebaut, indem es für die Generierung verschiedener Emotionsklassen angepasst und erweitert wird. Ziel ist es hierbei, ein generisches Emotionsgenerierungsmodell zu erhalten, mit dem die verschiedenen Mechanismen der Emotionsgenerierung aus den betrachteten Emotionstheorien (vgl. Kapitel 5 und 6) weitestgehend abgedeckt werden. Anschließend wird in Kapitel 8.3.2 ein Modell der emotionsbasierten (Verhaltens-)Steuerung vorgestellt.

8.3.1. Das Modell der Emotionseinschätzung

In Kapitel 5.6 wurde aus mehreren einschätzungstheoretischen Emotionsdefinitionen ein allgemeines Grundmodell der Emotionseinschätzung abgeleitet (vgl. Abb. 8.3). Dieses Modell bildet den Startpunkt für die Modellierung.

Nach dem einschätzungstheoretischen Grundmodell wird eine Emotion als Reaktion auf die von Agenten wahrgenommenen Umweltinformationen generiert (vgl. Kapitel 6). Die

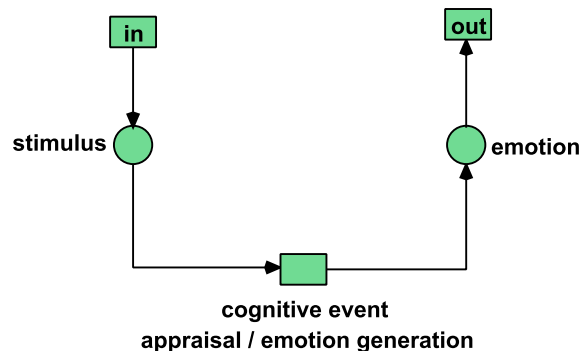


Abbildung 8.3.: Grundmodell der Emotionseinschätzung

Emotionsgenerierung wird in den meisten Einschätzungstheorien als sequenzieller Prozess betrachtet, bei dem der aktuelle Input aus der Umgebung ggf. in mehreren Schritten in Bezug auf den internen (Wissensbasis-)Zustand des Agenten bewertet wird. Aufgrund von vorgegebenen differenzierten Einschätzungsdimensionen, die je nach Theorie variieren können, wird eine Emotion erzeugt.

Aus dem Grundmodell der Emotionseinschätzung lässt sich das Grundmodell eines emotionalen Agenten ableiten (vgl. Abb. 8.4). Dieser Agent verarbeitet eine Umgebungsinformation und generiert aufgrund dieser Verarbeitung eine Emotion. Auslöser dieses Verarbeitungsprozesses kann ein beliebiger Input aus der Umgebung sein (z.B. ein Objekt, ein Ereignis, eine Situation). Die interne Verarbeitung (Transition *cognitive event appraisal / emotion generation*) besteht in der kognitiven Einschätzung der Stimuluseigenschaften in Bezug auf bestimmte Kriterien, die von der jeweils verwendeten Emotionstheorie abhängen. Durch die kognitive Einschätzung des Stimulus wird bestimmt, ob und ggf. welche Emotion generiert wird. Die Kopplung des Umgebungsinputs mit der generierten Emotion erfolgt nach diesem Modell asynchron über die Wissensbasis des Agenten (Stelle *emotional WB*). Dort sind u.a. die Einschätzungskriterien, die zu verwendende Einschätzungsstrategie sowie die funktionale Zuordnung zu den zu generierenden Emotionen hinterlegt.

Obwohl die meisten psychologischen Emotionstheorien ausschließlich sequenzielle Aussagen über die Ausgestaltung des Emotionsgenerierungsprozesses machen, geben einige einschätzungstheoretische Beschreibungen Hinweise darauf, dass diese sequenzielle Betrachtung nur deshalb gewählt wurde, weil es bei der herkömmlichen textuellen Beschreibung keine andere Darstellungsmöglichkeit für die einzelnen Prozessschritte gibt (z.B. [Fri07], [OCC88], [Laz91]).

So ist in der OCC-Theorie die Forderung nach einer sequenziellen Ausführung der Einschätzungen nicht explizit enthalten, sondern sie kann lediglich aus der Aufteilung der Stimuli in drei Klassen bezüglich der Einschätzung und aus einigen sprachlichen Formulierungen gefolgert werden (vgl. [BodAHN03]). Die Einschätzungsprozesse für verschiedene Stimuli könnten demnach auch parallel zueinander laufen (vgl. Kapitel 6.1.2).

Ebenso geht Lazarus in seiner Einschätzungstheorie nicht von sequenziell ablaufenden

8. Referenzmodell für Emotionsmodellierung

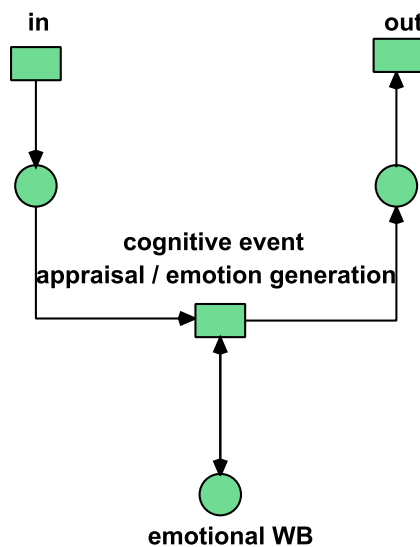


Abbildung 8.4.: Grundmodell des emotionalen Agenten.

Einschätzungsschritten aus, sondern von sehr schnellen, unwillkürlichen und kausal nicht vorhersagbar verlaufenden Prozessen (vgl. Kapitel 6.1.3).

In der Theorie von Frijda wird eine parallele Einschätzung der Stimulusrelevanz, die unabhängig von anderen Verarbeitungsprozessen stattfinden kann, sogar explizit zugelassen (vgl. Kapitel 6.1.2).

Ebenso ist die Möglichkeit einer synchronen Modellierung für die Umsetzung von mehreren Emotionstheorien innerhalb eines Modells bzw. für die Parallelisierung von Einschätzungen im Emotionsgenerierungsprozess interessant. Eine mögliche technische Erweiterung des Grundmodells des emotionalen Agenten, die auch eine synchrone Modellierung zulässt, ist in Abb. 8.5 dargestellt. Je nach Modellierungsperspektive (Weltausschnitt, Granularität, Domäne) können mit diesem Modell asynchrone und synchrone Aktivitäten /Handlungen (kognitive oder emotionale) modelliert werden.

Auf der technischen Ebene lässt sich die Synchronisation über eine Erweiterung des MULAN-Modells realisieren. Bisher lag dem MULAN-Agentenmodell eine asynchrone Kommunikation zugrunde. In [FDM11] wurde die Möglichkeit der Synchronisierung auf Agentenebene vorgestellt, die in Kapitel 8.4.3 beschrieben wird.

Die Umsetzung der Möglichkeit zur synchronen Modellierung in konkreten Simulationen bzw. eine Parallelisierung von Einschätzungen im Emotionsgenerierungsprozess ist lediglich im Rahmen einer differenzierten Betrachtung der betreffenden Theorien möglich und kann daher nur auf der Seite der Psychologie erfolgen. Im Rahmen dieser Arbeit wird diese (fachliche) Fragestellung explizit ausgeschlossen, da nur die Modellierung der bestehenden etablierten theoretischen Emotionsgrundlagen angestrebt wird und an dieser Stelle ohne Mitwirkung von fachlichen Experten keine Umschreibung / Weiterentwicklung von Emotionstheorien auf Basis der verwendeten Modellierungstechnik betrieben werden soll.

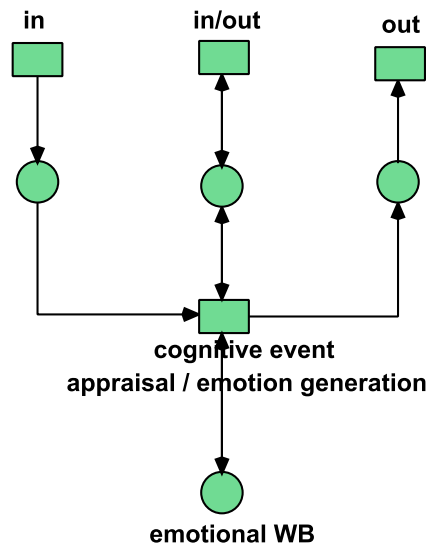


Abbildung 8.5.: Technische Erweiterung des Grundmodells für synchrone Modellierung

Da bei der Betrachtung der psychologischen Emotionstheorien in Kapitel 6 bisher nur sequenzielle Beschreibungen des Einschätzungsprozesses vorgefunden wurden, erfordert die Modellierung dieser Prozesse grundsätzlich keine synchrone Kopplung innerhalb des Einschätzungsmodells: Der Einschätzungsprozess kann mit dem Grundmodell aus Abb. 8.3 durch die asynchrone Kopplung über die Wissensbasis des Agenten modelliert und simuliert werden. Deshalb wird diesem Abschnitt das asynchrone Grundmodell aus Abb. 8.3 zugrunde gelegt. Die technische Realisierung der Synchronisation über eine Erweiterung des MULAN-Modells wird in Kapitel 8.4.3 beschrieben.

Anforderungen an das Einschätzungsmodell

Aus der Betrachtung der Einschätzungstheorien in Kapitel 6.1.2 lassen sich die folgenden grundlegenden Anforderungen an das Einschätzungsmodell ableiten (vgl. die Kapitel 5.6 und 8.2):

- Das Einschätzungsmodell muss die Mechanismen festlegen, durch die Emotionen aufgrund von bestimmten Ereignissen in der Agentenumgebung generiert und weiterverarbeitet werden.⁴
- Es müssen zum einen automatische emotionale Reaktionen (primäre oder Basisemotionen, vgl. Kapitel 5.3 modelliert werden können, - z.B. als unmittelbare Reaktionen auf bestimmte (kritische) Ereignisse in der dynamischen Umgebung des Agenten -, zum anderen aber auch die komplexeren Prozesse der Emotionseinschätzung, d.h. die

⁴Die Umweltereignisse können hierbei in der gleichen Weise wie im konventionellen MULAN-Modell durch ankommende Nachrichten repräsentiert werden

8. Referenzmodell für Emotionsmodellierung

Bewertung der aktuellen Wahrnehmungen des Agenten anhand seiner BDIs (vgl. Kapitel 5.4).

- Des Weiteren muss das Einschätzungsmodell die Möglichkeit bieten, Emotionen als Ergebnisse der Deliberationsprozesse zu modellieren, welche in einem Agenten entweder kontinuierlich ablaufen oder z.B. durch Änderung der aktuellen Umweltaabbildung (Wissensbasisaktualisierung / -veränderung) aktiviert werden. Für die Emotionsgenerierung ist in diesem Falle kein externer Auslöser erforderlich (siehe Kapitel 6.1.2).

Unter Berücksichtigung dieser Anforderungen zeigen wir im Folgenden, auf welche Weise die Generierung der verschiedenen Klassen von Emotionen, die in Kapitel 5 besprochen wurden, in einem E-MULAN-Agenten realisiert werden können.

Die Emotionsklassen, die bei der Betrachtung der emotionstheoretischen Literatur in den Kapiteln 5 und 6 identifiziert wurden, werden entsprechend ihrer Erzeugungsart jeweils entweder der Gruppe der reaktiv generierten oder der Gruppe der proaktiv generierten Emotionen zugeordnet. Die reaktive Emotionsgenerierung wird in Kapitel 8.3.1, die proaktive Emotionsmodellierung in Kapitel 8.3.1 diskutiert und modelliert.

Reaktive Emotionsgenerierung

Als reaktive Emotionen haben wir oben solche Emotionen bezeichnet, die von außen - z.B. durch ein Umweltereignis bzw. eine Umweltveränderung - angestoßen werden (vgl. Kapitel 5.6). Unter diesen Begriff fallen sowohl unmittelbare emotionale Reaktionen auf einen Umweltstimulus, die keiner Einschätzung bedürfen (z.B. Basisemotionen, elementare Affektzustände - siehe Kapitel 5.3), als auch solche Emotionen, die durch einen komplexen Einschätzungsprozess generiert werden (siehe Kapitel 5.4). Gemeinsames Merkmal dieser Emotionsklassen ist das externe Auslöseereignis, das zur Emotionsgenerierung führt.

In Kapitel 5.6 wurde in Abb. 5.11 ein allgemeines Petrinetz-Modell der reaktiven Emotionsgenerierung vorgestellt. Mit diesem Modell wurde die Generierung von primären / elementaren Emotionen (bzw. Hintergrundemotionen [Dam03]) und von Basisemotionen beschrieben. Mit dem gleichen Modell war auch die Darstellung der Emotionsgenerierung im einschätzungstheoretischen Sinne möglich, welche die komplexen kognitiven Prozesse der Stimulusbewertung in Abhängigkeit von den individuellen Bedürfnissen und Zielen einschließt.

Dieses Modell dient nun als Grundlage für die Implementierung der reaktiven Emotionsgenerierung in E-MULAN.

Grundmodell der Generierung von Basisemotionen Nach unserer Definition von reaktiven Emotionen stellen Basisemotionen eine spezielle Unterklasse dieser dar, bei denen die Emotionsgenerierung automatisch erfolgt und keine kognitive Einschätzung erfordert.

Eine Einschätzung bezeichnet im emotionstheoretischen Sinne die *kognitive* Auswertung eines Stimulus und erfordert deshalb immer einen Schlussfolgerungsmechanismus, um die Bedeutung bestimmter Stimuluseigenschaften des Agenten für individuelle Ziele, Absichten,

Pläne usw. zu bewerten. Bei Basisemotionen geht es hingegen um die automatische Generierung einer Emotion, die lediglich von den spezifischen Eigenschaften eines Auslösestimulus abhängt.

Die Generierung einer Basisemotion erfolgt dementsprechend so, dass eine Emotion aus einer vorgegebenen Grundausrüstung an Basisemotionen in Abhängigkeit von dem Auslösestimulus nach einer vordefinierten Relation ausgewählt wird. Abb. 8.6 zeigt einen möglichen Ablauf der Generierung von Basisemotionen in einem Agenten.

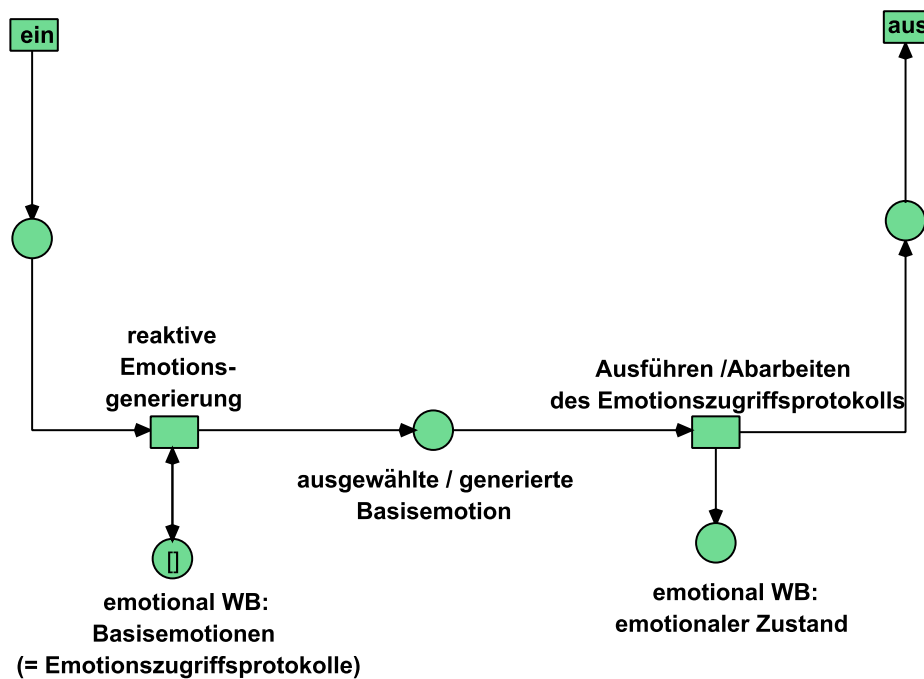


Abbildung 8.6.: Generierung von Basisemotionen.

Der in Abb. 8.6 abgebildete emotionale Agent empfängt über die Transition **ein** eine Nachricht, die Informationen über die Umgebung (bzw. über Veränderungen in der Umgebung) enthält. Der Agent besitzt eine Grundausrüstung an Basisemotionen, die in seiner Wissensbasis hinterlegt sind und hier als Marken auf der Stelle **emotional WB: Basisemotionen** repräsentiert werden. Implementationstechnisch handelt es sich bei den Basisemotionen um eine spezielle Art von Emotionsprotokollen, nämlich um *Emotionszugriffsprotokolle*, die den emotionalen Zustand eines Agenten setzen und ändern (vgl. Kapitel 8.2.4).

Die entsprechende Basisemotion wird als Marke auf der Stelle **generierte Basisemotion** abgelegt. Sie gilt als erfüllte Vorbedingung für die Auslösung eines *Emotionszugriffsprotokolls*, das mit dieser Basisemotion verknüpft ist. Die Aktivierung / Generierung einer Basisemotion besteht somit in der (rein reaktiven) Auswahl / Auslösung eines *Emotionszugriffsprotokolls*, das mit der emotionsauslösenden Nachricht verknüpft ist. Das Emotionszugriffsprotokoll greift auf die Agentenwissensbasis zu (Transition **Ausführen / Abarbeiten des Emotionszu-**

8. Referenzmodell für Emotionsmodellierung

griffsprotokolls) und setzt oder ändert die internen Variablen, die den emotionalen Zustand eines Agenten repräsentieren (Stelle emotional WB: emotionaler Zustand. Dadurch wird ein Agent in einen bestimmten emotionalen Zustand versetzt.

Anschließend kann ggf. ein sogenanntes *emotionale Steuerungsprotokoll* ausgelöst werden, das beschreibt, wie sich die ausgelöste Emotion im Verhalten des Agenten manifestiert (vgl. Kapitel 8.2.4). Die reaktive Generierung von Basisemotionen, die unmittelbar von der automatischen Auslösung eines Verhaltenssteuerungsprotokolls gefolgt wird, entspricht der Definition der Basisemotionen als fest verankerte Verhaltensprogramme, die als automatische Antworten auf einen spezifischen Reiz elementare emotionale Reaktionen auslösen (siehe Kapitel 5.3).

Sowohl alle charakteristischen Eigenschaften von Basisemotion als auch die Auslösung der Emotion als automatische Antwort auf einen spezifischen Umweltreiz im Sinne von LeDoux ([LeD96a], [LeD00]) bzw. Panksepp [Pan98] lassen sich mit dem "reaktiven" Einschätzungsmodell bzw. der reaktiven (Emotions)Protokollauswahl realisieren. Die Grundausstattung eines Agenten an Basisemotionen kann je nach verwendeter Emotionstheorie variieren (vgl. Kapitel 5).

Grundmodell der Emotionsgenerierung durch kognitive Einschätzungs- / Bewertungsprozesse Von einem allgemeinen Einschätzungsmodell haben wir die Eigenschaft gefordert, Emotionen als Ergebnisse eines beliebig komplexen kognitiven Bewertungs- / Einschätzungsprozesses modellieren zu können (vgl. Kapitel 8.3.1). In die Definition der reaktiven Emotionen wurde neben elementaren und Basisemotionen eine Gruppe der Emotionen eingeschlossen, die als Ergebnisse eines komplexen Prozesses der Einschätzung einer aktuellen Umweltsituation in Bezug auf die internen Ziele und Überzeugungen des Agenten generiert werden.

Die Generierung dieser Art von Emotionen erfordert die Modellierung eines Einschätzungsprozesses: Es muss ein spezieller Schlussfolgerungsmechanismus modelliert werden, der die Bedeutung bestimmter Stimuluseigenschaften für individuelle Ziele, Absichten, Pläne usw. des Agenten bewertet.

Das Einschätzungsmodell soll es erlauben, ohne viel zusätzlichen Aufwand verschiedene Einschätzungstheorien durch alternative Sätze von Einschätzungsprotokollen zu implementieren, z.B. um die entsprechenden Einschätzungstheorien zu simulieren bzw. zu vergleichen. Dies wird durch die Realisierung der Emotionsgenerierung auf der Ebene der Emotionsprotokolle erfüllt, wobei das allgemeine Einschätzungsmodell auf der Agentenebene nicht grundsätzlich geändert werden muss.

In Abb. 8.7 wird das Modell eines emotionalen Agenten vorgestellt, das Emotionen als Ergebnisse kognitiver Einschätzung des Auslöseereignisses generieren kann.

Der Agent in Abb. 8.7 empfängt über die Transition ein eine Nachricht mit den Umgebungsinformationen. Jeder Agent besitzt eine Menge der *Einschätzungsprotokolle* (siehe Kapitel 8.2.4), die als Marken auf der Stelle *Einschätzungsprotokolle* liegen. Die Ausstattung eines Agenten mit Einschätzungsprotokollen und der strukturelle Aufbau dieser Protokolle kann je nach verwendeter / modellierter Emotionstheorie variieren (vgl. Kapitel 6.1).

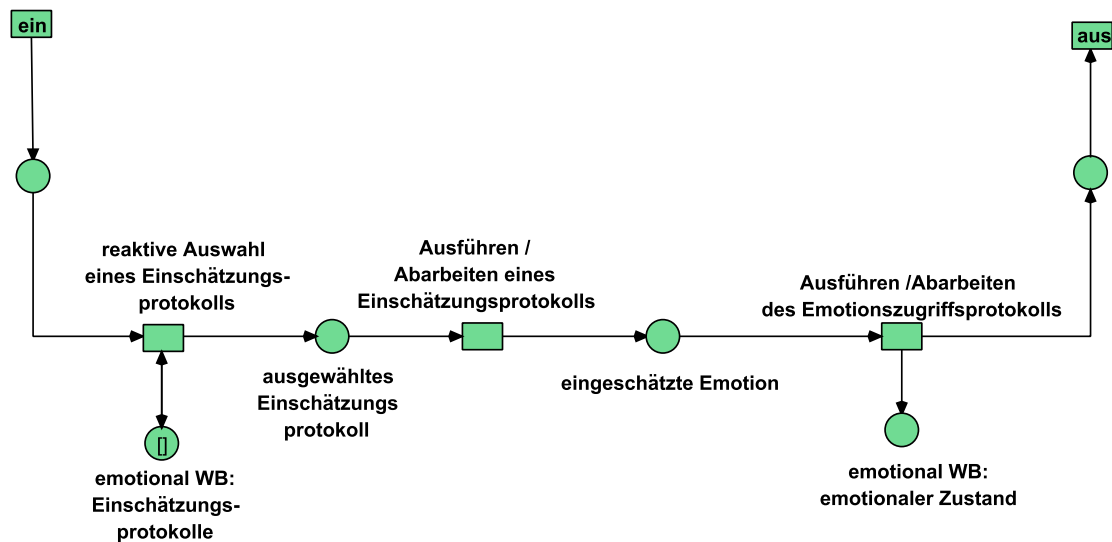


Abbildung 8.7.: Generierung von Emotion als Ergebnis der kognitiven Einschätzung.

Bei Ankunft einer Nachricht wird ein Einschätzungsprotokoll gestartet. Die Art der Relation zwischen dem durch die Nachricht repräsentierten Umgebungsstimulus und dem auszuwählenden Einschätzungsprotokoll ist durch die jeweils verwendete Einschätzungstheorie bestimmt und kann je nach Theorie sehr unterschiedlich ausgeprägt sein. Jedoch handelt es sich bei der Auswahl eines Einschätzungsprotokolls immer um eine rein reaktive Protokollauswahl, wie bei der Generierung von Basisemotionen, die durch reaktive Auswahl eines Emotionszugriffsprotokolls ausgelöst werden (siehe Kapitel 8.3.1).

Ein grundsätzlicher Unterschied besteht allerdings darin, dass im Falle der Basisemotionen ein *Emotionszugriffsprotokoll* reaktiv ausgewählt wird, was nach Definition der Emotionszugriffsprotokolle bereits der Generierung einer Basisemotion entspricht. D.h., bei Basisemotionen wird ein Agent unmittelbar nach der Ankunft einer Nachricht (durch reaktive Auswahl eines Emotionszugriffsprotokolls) in einen emotionalen Zustand versetzt. Durch diesen Ansatz werden die grundlegenden charakteristischen Eigenschaften von Basisemotionen realisiert, nämlich der automatische Einsatz der Emotion alleine aufgrund von spezifischen Eigenschaften eines Auslösestimulus, der keine höheren Kognitionen erforderlich macht.

Dagegen geht es im Falle der kognitiven Emotionseinschätzung lediglich um die reaktive Auslösung eines *Einschätzungsprotokolls*. Das Einschätzungsprotokoll an sich stellt noch keine generierte Emotion dar, es spezifiziert vielmehr den Einschätzungsprozess, der letztendlich zur Generierung einer Emotion (d.h. zur Aktivierung eines Emotionszugriffsprotokolls) führt bzw. führen kann.

Die automatische Emotionszugriffs- bzw. Einschätzungsprotokoll-Auslösung als Reaktion auf einen spezifischen Umweltstimulus ist in der internen konzeptionellen Struktur des reaktiven Protokollaufrufes der MULAN-Architektur fest verankert (vgl. Kapitel 4.3).

8. Referenzmodell für Emotionsmodellierung

Nach der reaktiven Auswahl eines Einschätzungsprotokolls (Transition **reaktive Auswahl eines Einschätzungsprotokolls**) aus dem vordefinierten Repertoire an Einschätzungsprotokollen (Stelle **emotional WB: Einschätzungsprotokolle**) wird dieses als Marke auf der Stelle **ausgewähltes Einschätzungsprotokoll** abgelegt. So kann im nächsten Schritt die Transition **Ausführen / Abarbeiten eines Einschätzungsprotokolls** schalten. Das ausgewählte Einschätzungsprotokoll wird gestartet und abgearbeitet. Im Laufe der Abarbeitung des Einschätzungsprotokolls wird aufgrund von kognitiver Bewertung bestimmter Stimuluseigenschaften eine Emotion - als Ergebnis der kognitiven Umwelteinschätzung - generiert und auf der Stelle **eingeschätzte Emotion** abgelegt. Daraufhin kann ein Emotionszugriffsprotokoll (oder ggf. mehrere Emotionszugriffsprotokolle, falls mehrere emotionale Zustände durch verschiedene Einträge in der Agentenwissensbasis repräsentiert werden können) ausgelöst werden (Transition **Ausführen / Abarbeiten des Emotionszugriffsprotokolls**). Dieses greift auf die Agentenwissensbasis zu und setzt oder ändert die internen Variablen, die den emotionalen Zustand eines Agenten darstellen (Stelle **emotional WB: emotionaler Zustand**).

Zusammengefasst, erfolgt die Generierung der Emotionen durch kognitive Einschätzungs- / Bewertungsprozesse nach dem vorgestellten Grundmodell in zwei Schritten:

- Auf die Wahrnehmung eines spezifischen Umweltereignisses folgt die automatische Auslösung eines Einschätzungsprotokolls, das mit dem Auslöseereignis verbunden ist. Durch das Einschätzungsprotokoll wird bestimmt, welche Emotion anschließend ausgelöst werden soll.
- Unmittelbar nach dem Einschätzungsprozess wird ein Emotionszugriffsprotokoll gestartet, das mit der eingeschätzten Emotion verknüpft ist. Das Emotionszugriffsprotokoll versetzt den Agenten durch die Wissensbasisänderung in einen der eingeschätzten Emotion entsprechenden emotionalen Zustand.

Schließlich kann aufgrund der eingeschätzten Emotion ein *emotionales Steuerungsprotokoll* gestartet werden, das mit dem jeweiligen Emotionswert verknüpft ist. Das Steuerungsprotokoll beschreibt, wie sich die ausgelöste Emotion im Verhalten des Agenten manifestiert (siehe Kapitel 8.2.4).

Proaktive Emotionsgenerierung

Als proaktive Emotionen haben wir weiter oben Emotionen bezeichnet, die keinen externen Auslöser erfordern, sondern als Ergebnisse der internen Deliberationsprozesse entstehen (vgl. Kapitel 5.6). Diese Deliberationsprozesse laufen in einem Agenten entweder kontinuierlich ab oder werden durch die Änderung der aktuellen Umweltabbildung (Wissensbasisaktualisierung / -veränderung) angestoßen (vgl. Kapitel 8.2).

Unter diese Definition von proaktiven Emotionen fallen insbesondere die sekundären bzw. sozialen Emotionen [Dam03], die ein Agent infolge einer Auswertung der Umweltsituation (bzw. des ihm vorliegenden Wissens über die Umweltsituation) bezüglich seiner (langfristigen) Konsequenzen für die aktuellen Ziele und Pläne, seiner Erwartungskonformität usw. generiert (vgl. Kapitel 5.6).

Die einschätzungstheoretische Emotionsdefinition kann auch eine proaktive Emotionsgenerierung implizieren, z.B. dann, wenn die kognitiven Einschätzungs- / Bewertungsprozesse nicht durch Umweltveränderungen, sondern durch Änderungen des internen Zustandes bzw. der Wissensbasis des Agenten angestoßen werden. So erfolgt z.B. die Einschätzung der sogenannten *zusammengesetzten Emotionen* nach der OCC-Theorie proaktiv (vgl. Kapitel 6.7).

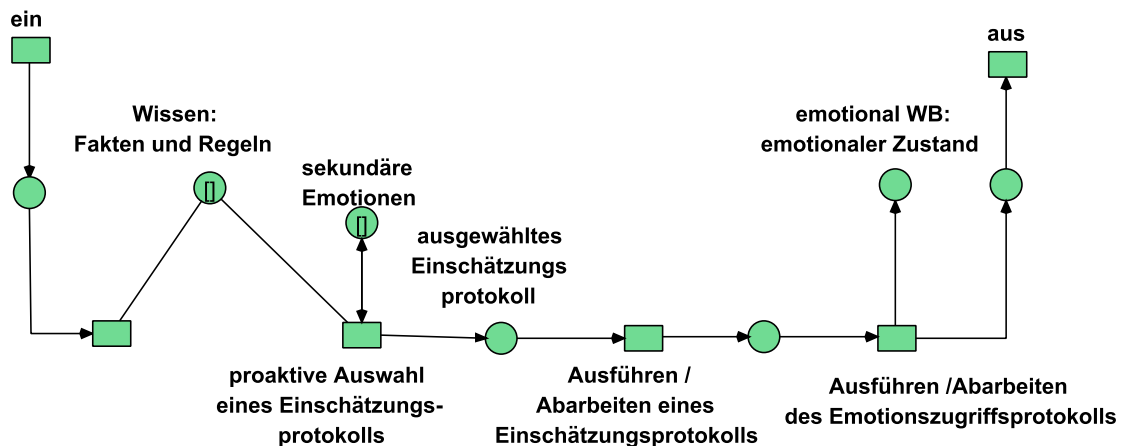


Abbildung 8.8.: Proaktive Emotionsgenerierung

Der Unterschied von sekundärer und einfacher - bzw. primärer - Einschätzung, die wir oben als reaktive Emotionsgenerierung bezeichnet haben (siehe Kapitel 9.2), besteht somit hauptsächlich darin, dass bei der sekundären Einschätzung der Einschätzungsprozess nicht durch einen externen Auslöser (eine Nachricht), sondern durch einen internen Auslöser (z.B. eine Änderung des Wissensbasiszustandes) angestoßen wird.

Die Deliberations- bzw. Einschätzungsprozesse, die zur Emotionsgenerierung führen, können bei den beiden Einschätzungsarten durch Einschätzungsprotokolle realisiert werden. Gleichmaßen erfolgt die Aktualisierung des emotionalen Zustandes des Agenten bei beiden Einschätzungsarten durch Emotionszugriffsprotokolle. Deshalb erfordert die Modellierung der proaktiven Emotionsgenerierung keine zusätzlichen Anpassungen an das Einschätzungsmodell, außer dass die Möglichkeit des proaktiven Startens der Einschätzungsprotokolle ergänzt werden muss (vgl. Abb. 8.8).

8.3.2. Modell der emotionsbasierten Steuerung

Die Steuerungsfunktion der Emotionen, d.h. die vielfältigen Auswirkungen der Emotion auf die akteurs- / agenteninternen Verarbeitungsprozesse, soll im Folgenden beispielhaft auf Basis des E-MULAN-Referenzmodells abgebildet werden.

Betrachtet man den Agenten als Plattform für Emotionen und Aktionen, so können diese zwischeneinander dynamisch (d.h. zur Laufzeit) Informationen (Nachrichten) austauschen. Auf diese Weise ist die Modellierung der Interaktion zwischen Agentenemotionen,

8. Referenzmodell für Emotionsmodellierung

modelliert durch Emotionsprotokolle, und Aktionen, modelliert durch konventionelle MULAN-Protokolle, gegeben.

Dem folgenden Beispiel wird ein aufgefaltetes E-MULAN-Modell zugrunde gelegt, das einen Akteur durch die beiden Komponenten der E-MULAN-Agentenebene, die emotionale und die rationale Agentenkomponente, repräsentiert (vgl. Abb. 8.2). Jede Komponente wird im Folgenden als ein separater Agent betrachtet.

In der emotionalen Agentenkomponente wird proaktiv eine Emotion generiert und (asynchron) an die rationale Agentenkomponente gesendet. Diese wird ggf. sein aktuelles Verhalten abhängig von der aktuellen Emotion anpassen, die sie von der emotionalen Komponente übermittelt bekommt. Die Emotionsrepräsentation verbleibt in der Wissensbasis der emotionalen Komponente, während die relationale Zuordnung der aktuellen Emotion zu einem Verhalten bzw. einer Verhaltensanpassung in der Wissensbasis der rationalen Komponente abgelegt wird. Eine ausführliche Beschreibung des Szenarios zur emotionsbasierten Steuerung sowie zur emotionstheoretischen Motivation findet sich in Kapitel 9.4, in dem die Implementierung dieses Szenarios in einer E-MULAN-Anwendung beschrieben wird.

Hier wird zuerst wieder ein vereinfachtes MULAN-Modell nach [Ree10] verwendet, um die Interaktionen zwischen den Modellkomponenten transparenter beschreiben und nachvollziehen zu können. Das vereinfachte MULAN besteht aus einem Plattform-Referenznetz (Abb. 8.9), zwei Agenten-Referenznetzen (rationaler Agent - Abb. 8.11 und emotionaler Agent - Abb. 8.12) und einem Referenznetz für die Protokollfabrik (Abb. 8.10). In den Abb. 8.13 und 8.14 sind ferner emotionale und rationale Protokolle dargestellt.

Die Plattform hat eine zentrale Stelle für die instanziierten Agenten-Referenznetze, daran je eine Transition zum Starten und Beenden von Agenten sowie eine Transition für Plattform-internen (asynchronen) Nachrichtenaustausch. Außerdem gibt es eine Stelle, an der zwei Tupel zum Initialisieren der Agenten mit einem Namen (“A“ oder “B“) und einer rationalen oder emotionalen Wissensbasis (“rational-prot“ oder “emo-prot“) liegen. An der Transition **start** werden die Agenten-Referenznetze instanziiert und die entsprechenden Tupel, bestehend aus dem Namen und der Wissensbasis, zugewiesen.

Das Agenten-Referenznetz unterscheidet sich nicht grundsätzlich von dem in [Ree10], S. 96, vorgestellten Modell. Lediglich werden beim emotionalen Agenten die Transitionen **reactive emotion generation** bzw. **proactive emotion generation** ergänzt, weil es sich an dieser Stelle nicht um eine Aktionswahl, sondern um eine reaktive bzw. proaktive Generierung von Emotionen handelt (vgl. Kapitel 8.3.1).

Die Umsetzung des oben beschriebenen Szenarios erfolgt mit Hilfe der Protokollnetze, die in den Abb. 8.13 und 8.14 dargestellt sind. Im emotionalen Protokoll erfolgt ein Nachrichterversand an den rationalen Agenten, wobei die in der Nachricht enthaltene Emotion zuerst versendet (Transition **send**) und dann in die Wissensbasis des emotionalen Agenten geschrieben wird (Transition **update kb**), jedoch ohne dass der Inhalt der Wissensbasis als Marke dargestellt wird). Das rationale Protokoll, das bisher ein bestimmtes Verhalten ausgeführt hat (Transition **process behavior**), empfängt die Emotion über die Transition **receive**, greift auf seine Wissensbasis zu (Transition **update kb**), um das für diese spezielle Emotion vordefinierte Verhalten abzufragen, und passt das Verhalten entsprechend an (Transition **update behavior**) bzw. behält das bisherige Verhalten (Transition **retain behavior**)

bei, falls keine Verhaltensanpassung an die aktuelle Emotion in der Wissensbasis des Agenten definiert ist.

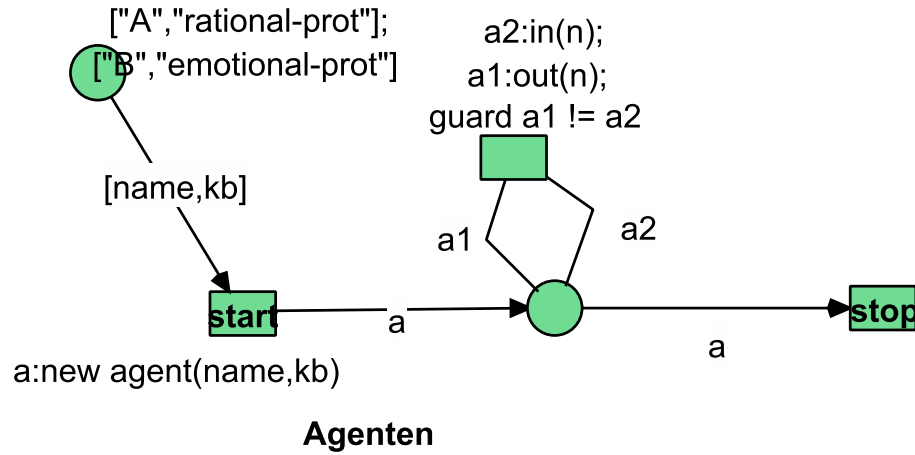


Abbildung 8.9.: Plattform-Netz

In Abb. 8.15 wird die asynchrone Kommunikation zwischen einem rationalen und einem emotionalen E-MULAN-Agenten mit einem Kausalnetz veranschaulicht. Dieses Kausalnetz ist ausführbar, man kann damit das oben beschriebene Szenario simulieren und das Verhalten der Modellkomponenten beobachten. Durch farbliche Kennzeichnung lassen sich die einzelnen Modellkomponenten und ihr Verhalten im Kausalnetz besser zuordnen. Weitere Details zur Darstellung und Schaltsemantik der Kausalnetze für Prozessdarstellungen in Mulan können [Ree10], S. 100, entnommen werden.

8. Referenzmodell für Emotionsmodellierung

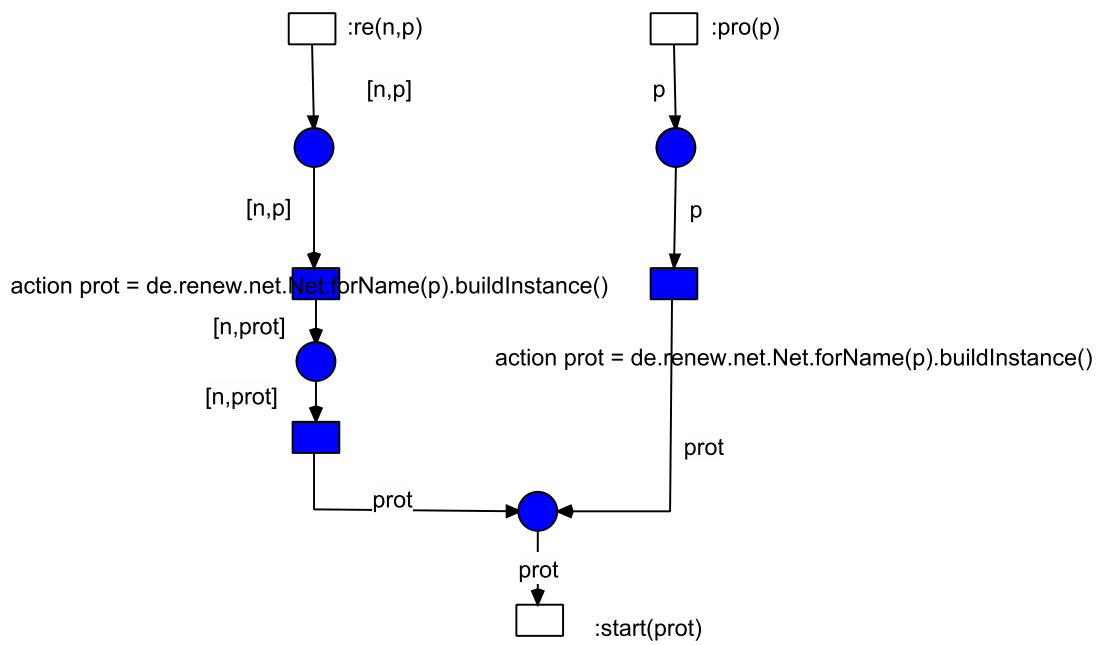


Abbildung 8.10.: Protokollfabrik

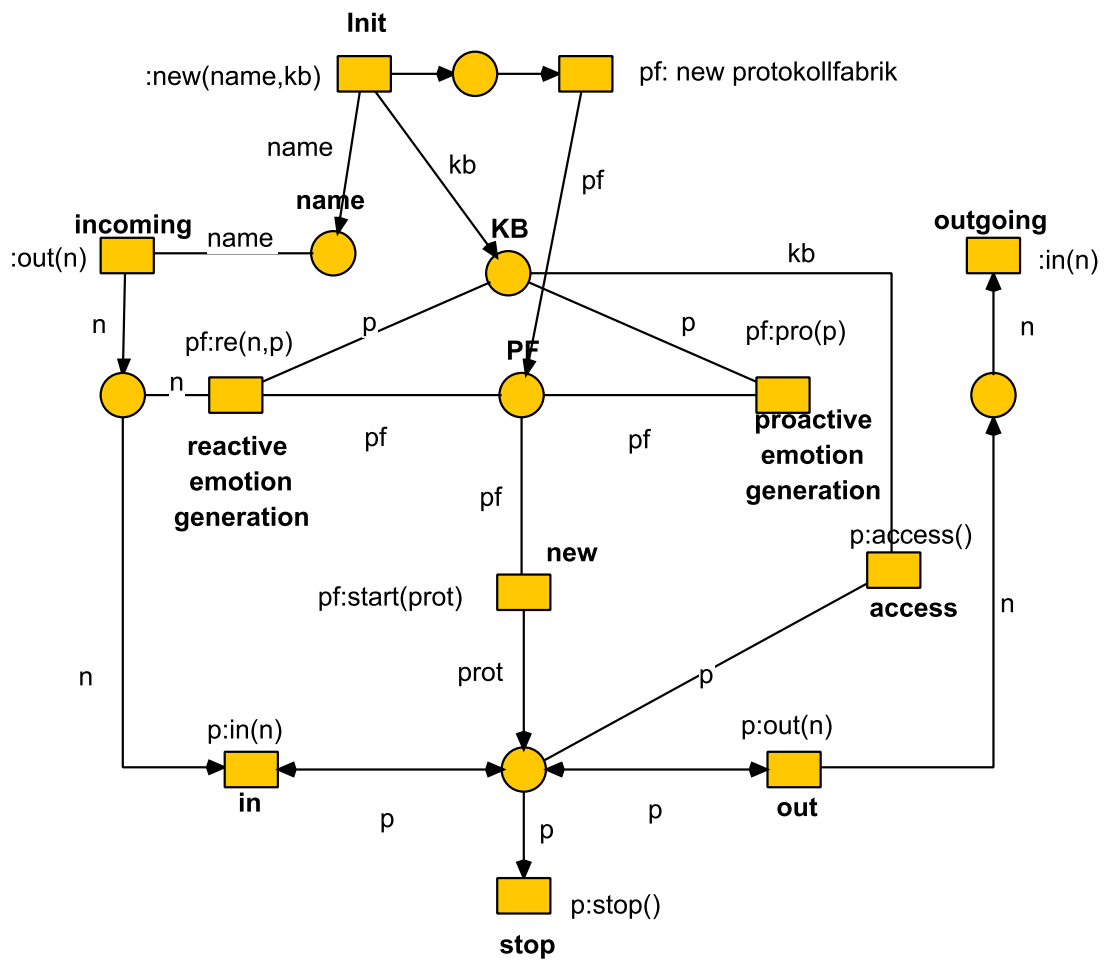


Abbildung 8.11.: Stark vereinfachtes Agentennetz - rationaler Agent.

8. Referenzmodell für Emotionsmodellierung

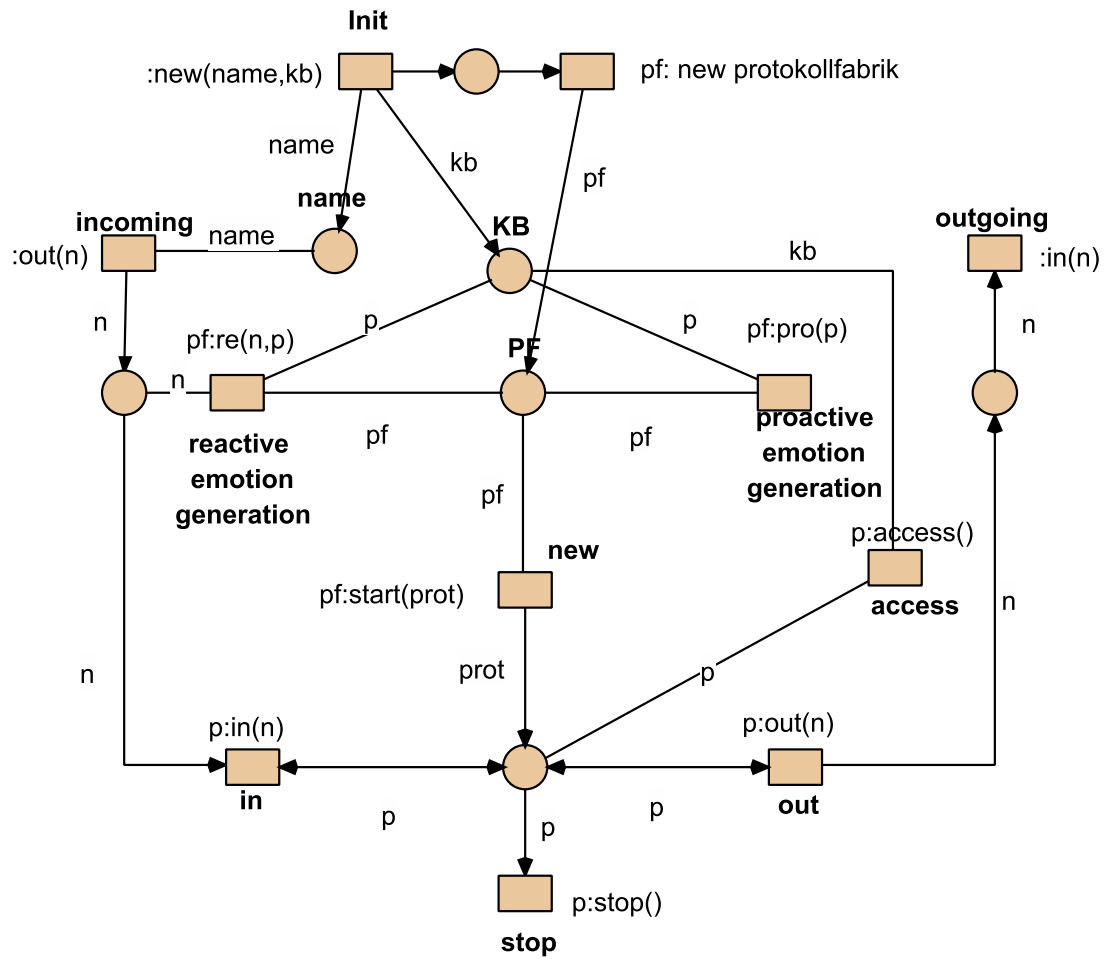


Abbildung 8.12.: Stark vereinfachtes Agentennetz - emotionaler Agent.

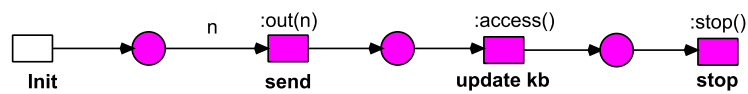


Abbildung 8.13.: Emotionaler Protokoll

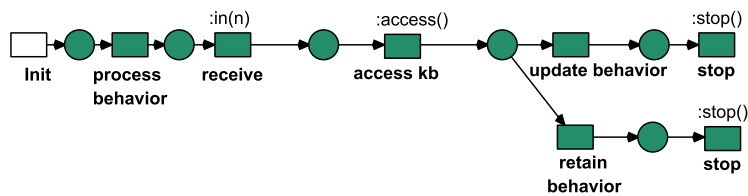


Abbildung 8.14.: Rationaler Protokoll

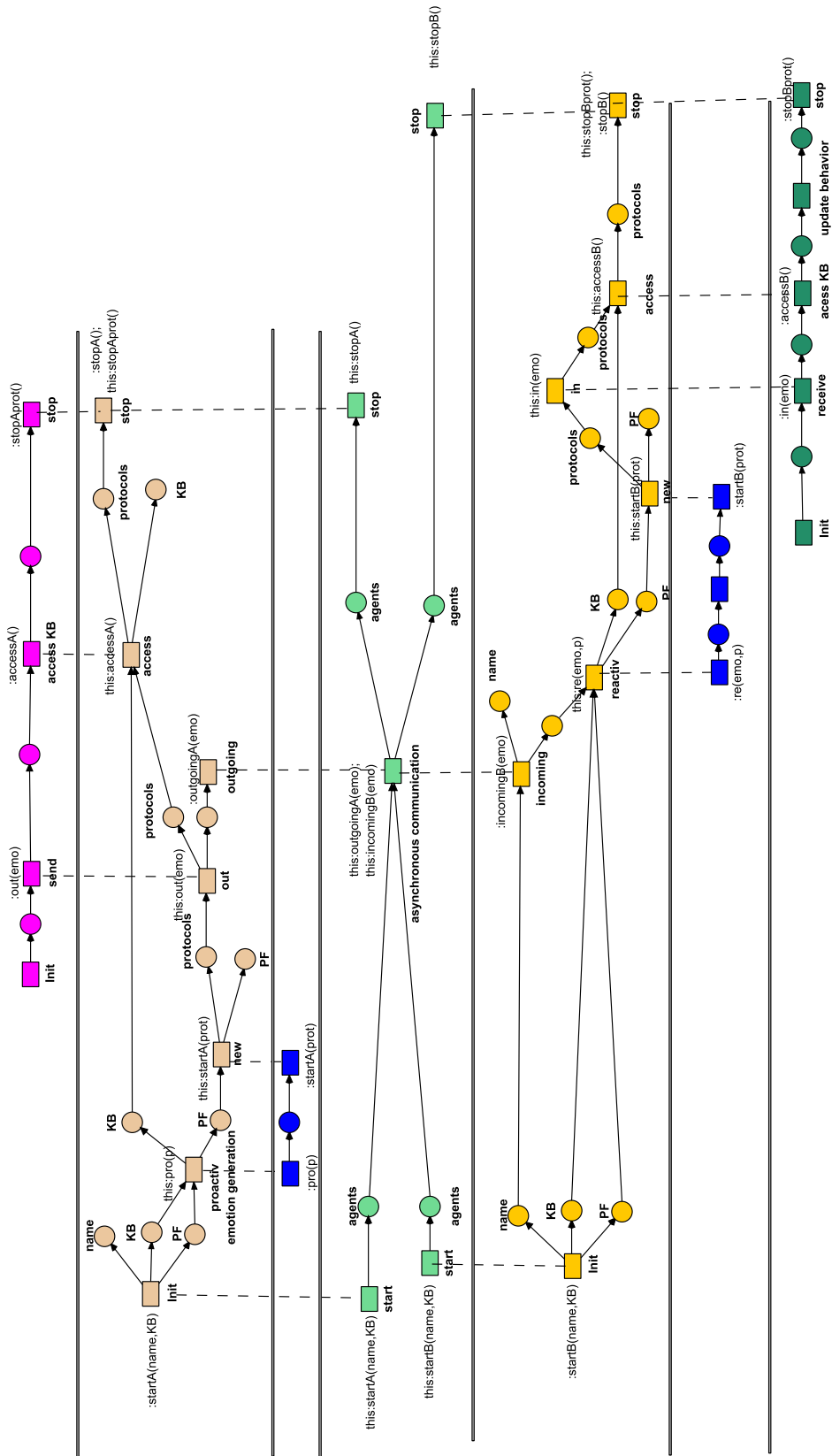


Abbildung 8.15.: Modell der emotionsbasierten Verhaltenssteuerung: Interaktion eines MULAN-Protokolls mit einem emotionalen Steuerungsprotokoll in E-MULAN.

8.4. Verknüpfung der rationalen und der emotionalen Modellkomponenten

In E-MULAN gibt es einerseits die “vertikale“ Verknüpfung zwischen den vier hierarchischen Modellebenen (vgl. Abb. 8.16), andererseits die “horizontale“ Verknüpfung der konventionellen MULAN-Elemente zu den aufgefalteten emotionalen Komponenten der Shadow-Architektur (im Bild mit doppelter roter Linie dargestellt). Diese zwei Arten von Beziehungen werden in Abb. 8.16 veranschaulicht und im folgenden Abschnitt beschrieben und diskutiert.

8.4.1. Lose vs. enge Kopplung

Die Verbindung der beiden oben vorgestellten Modellkomponenten (MULAN und E-MULAN) kann auf unterschiedliche Weise realisiert werden.

Zum einen ist eine *enge Kopplung* denkbar, die einer festen, nicht veränderbaren Verknüpfung von Modellkomponenten entspricht. Eine enge Kopplung könnte z.B. durch eine statische Verbindung der jeweiligen Modellebenen durch Reservierungs-, Test-, inhibitorische oder einfache Kanten realisiert werden.

Dabei könnte durch den Einsatz von Reservierungskanten oder entsprechender Netzelemente (z.B. Konflikt, vgl. Kapitel 4.4.2) ein Ausschluss implementiert werden, so dass die Verarbeitung entweder nur in der rationalen (MULAN)- oder nur in der emotionalen (E-MULAN)-Komponente des Referenzmodells erfolgt. Dies würde jedoch der generellen Zielsetzung widersprechen, das Zusammenspiel der emotionalen und kognitiven Verarbeitungsprozesse zu modellieren und zu untersuchen. Z.B. wäre die Darstellung von emotionalen und kognitiven Prozessen gemäß den Untersuchungsergebnissen von [LeD96b], wonach die beiden Arten von Prozessen parallel ablaufen und sich gegenseitig ergänzen (vgl. Kapitel 6.2.1), bei der engen Kopplung der “kognitiven“ und “emotionalen“ Modellkomponenten nicht möglich.

Eine lose Kopplung erlaubt dagegen eine dynamische Bindung von Modellteilen. Der Art der Abhängigkeit zwischen den rationalen und den emotionalen Modellteilen wird durch die jeweiligen Netzelemente (Kanäle, Aufrufrichtung, Kantenrichtung etc.) angedeutet.

Die Art der Anbindung der MULAN und der E-MULAN-Modellkomponente wird im Referenzmodell explizit offengehalten und ist letztendlich dem Modellierer zu überlassen. Die Anbindung über eine *lose Kopplung* weist jedoch einige wichtige Vorteile auf: Wird der wechselseitige oder unidirektionale Einfluss zwischen den Referenzmodell-Komponenten durch synchrone Kanäle (siehe Kapitel 4.4.2 bzw. 4.6) modelliert, so kann die gleichzeitige Übergabe der Eingabeinformationen an beide Modellkomponenten erfolgen und eine nachfolgende Synchronisation der entsprechenden Verarbeitungsprozesse realisiert werden.

Dank der losen Kopplung können die rationalen und emotionalen Modellkomponenten bei der Modellierung dynamisch, d.h. sehr flexibel kombiniert werden. Zudem können sie sowohl separat als auch gemeinsam - abhängig vom Modellgegenstand - verwendet werden. Es bleibt dem Modellierer überlassen, je nach Modellgegenstand bzw. modellierter Emotionstheorie die Beziehung zwischen den rationalen und emotionalen Elementen durch eine explizite

8.4. Verknüpfung der rationalen und der emotionalen Modellkomponenten

Abgrenzung der Emotionsprozesse zu den kognitiven / rationalen Modellkomponenten - bzw. durch den Anschluss an diese - zu definieren bzw. zu spezifizieren.

Man kann auch hierbei die kognitiven Modellkomponenten komplett weglassen, um von dem kognitiven Apparat zu abstrahieren, oder ihn als Umgebung betrachten, die einen Input für das emotionale Modell liefert. In dem einen Fall werden allein die E-MULAN-Komponenten des Referenzmodells zur Modellierung benutzt. Dies ist z. B. für die Modellierung von Emotionstheorien sinnvoll. Im anderen Fall übernimmt das rationale Modell lediglich die Rolle der Ein- und Ausgabe bzw. der Sensoren und Effektoren eines Agenten.

Will man umgekehrt von den emotionalen Komponenten abstrahieren, so kann man auf das konventionelle MULAN -Modell zurückgreifen und die E-MULAN-Komponenten ausklammern. Die Modellierung erfolgt dann wie bisher, da nicht weiter zwischen Emotion und Rationalität unterschieden wird: Die rein rationalen Modelle können mit rationalen Rahmenwerkkomponenten beschrieben werden. Hierbei können die nicht berücksichtigten emotionalen Aspekte ggf. durch nachträgliche Modellierung mit emotionalen Komponenten ergänzt werden.

Damit erreichen wir, dass wir zum einen die bisherigen Modelle beibehalten können. Nach Bedarf können Emotionen in den Modellen auf allen Ebenen ergänzt werden, wenn sich dies z.B. durch veränderte Aufgabenstellung bzw. Modellanwendung als notwendig erweist.

Zum anderen ermöglicht die vorgeschlagene Architektur insbesondere jedoch auch die Modellierung der Interaktion der kognitiven und emotionalen Prozesse und somit die Integration der beiden Systemsichten auf der Ebene der Modellierung. In den Kapiteln 8.2.4 und 8.2.3 wurde exemplarisch gezeigt, wie auf der Agentenebene bzw. auf der Emotionsebene die Interaktion der beiden Modellkomponenten MULAN und E-MULAN modelliert werden kann. Speziell wurde die Modellierung der emotionsbasierten Steuerung betrachtet, bei der die E-MULAN -Komponenten als Steuerungselemente für MULAN-Komponenten auftreten: Durch die Emotionsprotokolle (speziell: emotionale Steuerungsprotokolle), die parallel zu den konventionellen MULAN-Protokollen ablaufen und mit diesen interagieren, ist die Möglichkeit der Modellierung des Einflusses der Emotion auf das Agentenverhalten gegeben (siehe Kapitel 9).

8.4.2. E-Mulan als Steuerungsmodell

Die wichtige Steuerungsfunktion der Emotion für das Verhalten und die kognitiven Verarbeitungsprozesse wurde bereits in Kapitel 5 im Rahmen der interdisziplinären Betrachtung verschiedener Emotionsdefinitionen identifiziert.

In Kapitel 6 wurde die Rolle der Emotion bei der Steuerung kognitiver Prozesse am Beispiel der psychologischen Theorien von Scherer, Frijda und Lazarus (vgl. Kapitel 6.1.2) veranschaulicht. Schließlich wurde in Kapitel 6.3 die Steuerungsfunktion der Emotion auf der Ebene der sozialen Strukturen betrachtet. Hierbei konnte aus mehreren soziologischen Theorien der Mechanismus der emotionsgesteuerten Normetablierung abgeleitet und mit Netzmodellen verdeutlicht werden (vgl. Kapitel 6.3.3).

In Kapitel 8.3 wurde die Möglichkeit der Realisierung der Steuerungsfunktion der Emotion auf der Agentenebene aufgezeigt (vgl. Kapitel 8.3.2). Betrachtet man Agenten als Plattform für Emotionen und Aktionen, so können diese zwischeneinander dynamisch

8. Referenzmodell für Emotionsmodellierung

(d.h. zur Laufzeit) Informationen (Nachrichten) austauschen. Auf diese Weise ist die Modellierung der Interaktion zwischen Agentenemotionen (modelliert durch emotionale Verhaltenssteuerungs-Protokolle) und Aktionen (modelliert durch konventionelle MULAN-Protokolle) gegeben.

Das gleiche Prinzip – das Herausziehen einer emotionalen Steuerungseinheit, die auf einer Ebene die Steuerung rationaler Prozesse übernimmt – kann auch auf die anderen Modellebenen übertragen werden.

Ein ähnliches Vorgehen wird in [WEKBM08] vorgeschlagen: Jede Modell-Ebene einer hierarchischen Referenzarchitektur ORGAN wird durch eine universelle Systemeinheit repräsentiert, welche durch jeweils drei unterschiedliche Arten von internen Prozessen / Einheiten – Operations-, Integrations- und Verwaltungseinheiten – charakterisiert ist (vgl. [WEKBM08], S. 4). Durch die Festlegung einer universellen Systemeinheit wurde ein einheitliches Strukturierungsmerkmal für alle Ebenen definiert.

Als konzeptionelle Grundlage für ein solches Vorgehen kann die Einheitentheorie (siehe [TM03], [Hew10]) hinzugezogen werden. Hiernach kann jede Systemeinheit jeweils als Protokoll, Agent, Plattform oder MAS in die Umgebung eingebettet werden, wobei die Art der Einbettung durch die Umgebungseigenschaften bzw. die gewünschte Sicht auf die zu modellierenden Einheiten und ihre Prozesse definiert wird.

Entsprechend ist die unterschiedliche Einbettung eines Modellgegenstandes auf verschiedenen Referenzmodell-Ebenen denkbar: Jede Einheit kann je nach Modellierungsinteresse aus der Protokoll-, Agent-, Plattform- oder MAS-Perspektive betrachtet werden und auf den entsprechenden Modellebenen beschrieben werden. Z.B. kann eine Emotion einerseits als ein Prozess gesehen werden, der innerhalb eines Agenten / Akteurs abläuft und deshalb als Protokoll modelliert wird (wie in Kapitel 8.2.4 vorgeschlagen). Andererseits ist auch eine Modellierung auf der Agentenebene denkbar, wenn etwa die Interaktion zwischen verschiedenen Emotionen eines Agenten den Hauptgegenstand des Modells bildet. Gleichzeitig kann ein Agent selbst als Plattform für seine internen Verhaltensdeterminanten (Motive, Pläne, Emotionen etc.) betrachtet werden.

Hierdurch erhalten wir einen rekursiven generischen Modellierungsansatz, mit dem die Wechselwirkungen zwischen Emotion und Kognition auf verschiedenen Ebenen des E-MULAN-Referenzmodells untersucht und modelliert werden können. Auf diese Weise kann der Einfluss der Emotion auf Kognition auf jeder Ebene des Referenzmodells beschrieben werden: Emotion kann so als Kontrollstruktur auf jeder Modellebene gesehen werden. In diesem Zusammenhang kann die emotionale Modellkomponente allgemein als Steuerungsmodell innerhalb der E-MULAN -Architektur betrachtet werden.

8.4. Verknüpfung der rationalen und der emotionalen Modellkomponenten

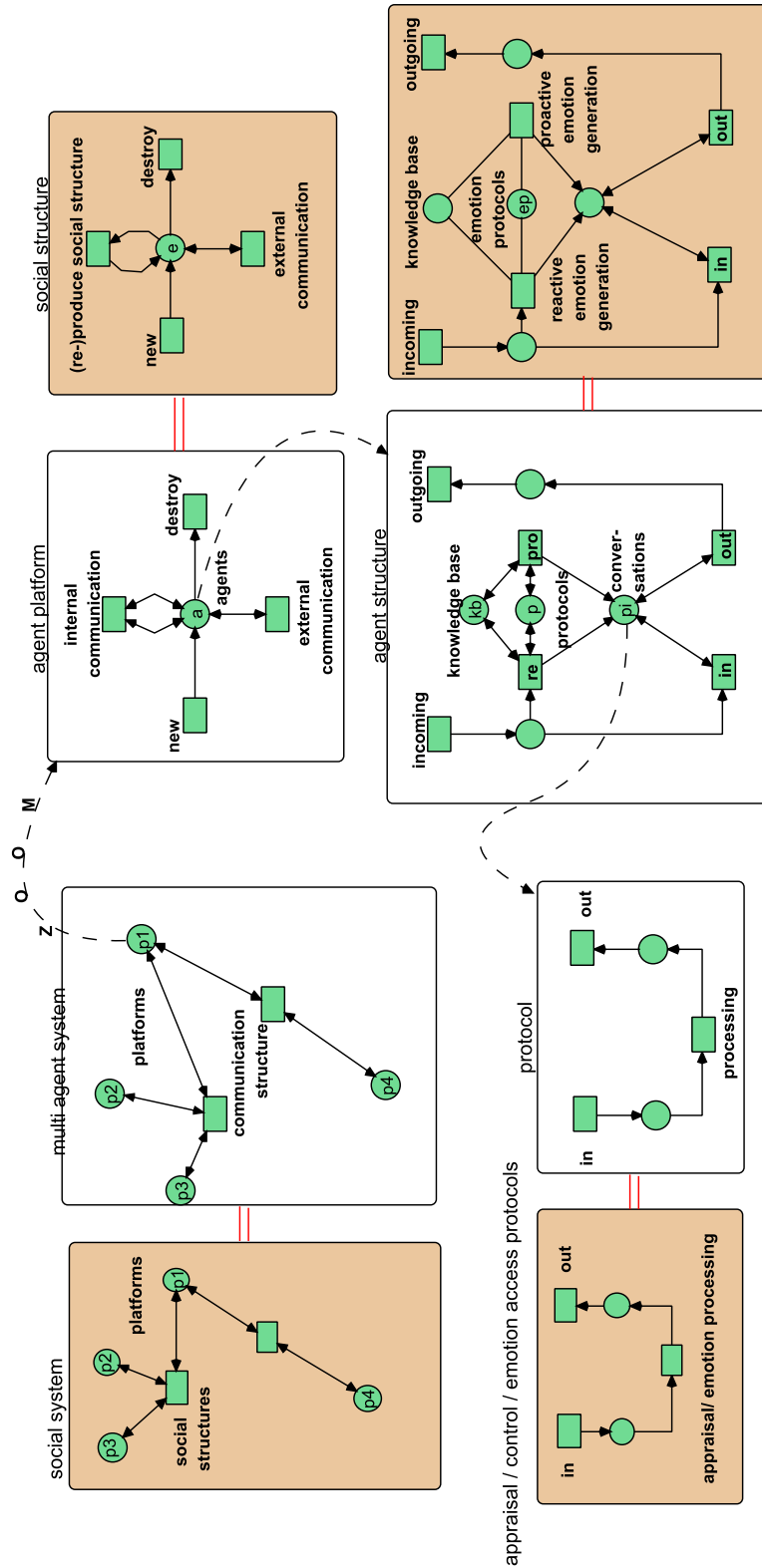


Abbildung 8.16.: Horizontale und vertikale Verknüpfung zwischen E-MULAN-Ebenen

8.4.3. Synchronisationsmechanismus für emotionale und rationale Modellkomponenten

Eine weitere Möglichkeit der Kopplung von rationalen und emotionalen Bestandteilen des E-MULAN-Modells besteht in der Erweiterung der MULAN-Architektur um einen Synchronisationsmechanismus, der es erlaubt, Agenten innerhalb einer Plattform zu synchronisieren. Die bisherige MULAN-Architektur sah in der konkreten Umsetzung mittels CAPA lediglich eine asynchrone Kommunikation vor. Die Erweiterung von MULAN um einen Synchronisationsmechanismus wurde in [FDM11] vorgestellt. Sie wird in diesem Abschnitt allgemein erläutert und in Kapitel 8.4.4 durch ein konkretes Anwendungsbeispiel illustriert.

Durch die Auffaltung der MULAN-Architektur im E-MULAN-Referenzmodell entsteht eine Schatten-Architektur zu MULAN, so dass jede Architekturebene nun konzeptionell durch zwei getrennte symmetrische Agentenmodelle dargestellt wird (vgl. Abb. 8.1). Damit diese ausdrucksstark synchronisiert werden können, ist das Referenzmodell entsprechend zu erweitern. des Agentenverhaltens (insbesondere von emotionalen und rationalen Agentenkomponenten) im Fokus steht, werden wir die Synchronisation nicht auf allen Ebenen zugleich ergänzen müssen, sondern wir können uns auf die synchrone Kommunikation auf der Agentenebene begrenzen.

Damit wird eine synchrone Kommunikation für Agenten durch die Erweiterung der MULAN-Agentenebene, die Bereitstellung durch die Plattform sowie die Steuerung auf der Protokollebene möglich (siehe Abb. 8.17).

Die grundlegende Idee ist dabei, die Schnittstelle des Agenten zur Plattform, welche bisher aus zwei für asynchrone Kommunikation genutzten Kanälen besteht, um einen weiteren Kanal für synchrone Kommunikation zu ergänzen (Transition **Synchronous communication** in Abb. 8.17 bzw. **sync** in Abb. 8.18). Auf der Plattformebene wird die Dienstleistung „Interne Kommunikation“ an die erweiterte Schnittstelle angepasst, so dass sie sowohl synchrone als auch asynchrone Kommunikation unterstützt. Dabei ist anzumerken, dass die Transition **Interne Kommunikation** des herkömmlichen MULAN-Modells de facto schon immer eine synchrone Verschmelzung der Ein- und Ausgabekanäle beider Kommunikationspartner war. Die Asynchronizität wurde auf Agentenseite sichergestellt. In der technischen CAPA-Plattform wurde die interne Kommunikation hingegen tatsächlich asynchron implementiert. Hier besteht die Ergänzung darin, die synchrone Transition des MULAN-Modells zusätzlich zu übernehmen.

Auf Agentenseite wird eine dritte Schnittstellentransition für den synchronen Kommunikationskanal in das Modell integriert. Da die Verhaltenssteuerung des Agenten im Wesentlichen auf der Ebene der Protokolle geschieht, muss dieser Kanal zu den Protokollinstanzen durchgereicht werden. Die neuen Transitionen und ihre Verbindung sind in 8.18 im Agentennetz farblich hervorgehoben.

8.4. Verknüpfung der rationalen und der emotionalen Modellkomponenten

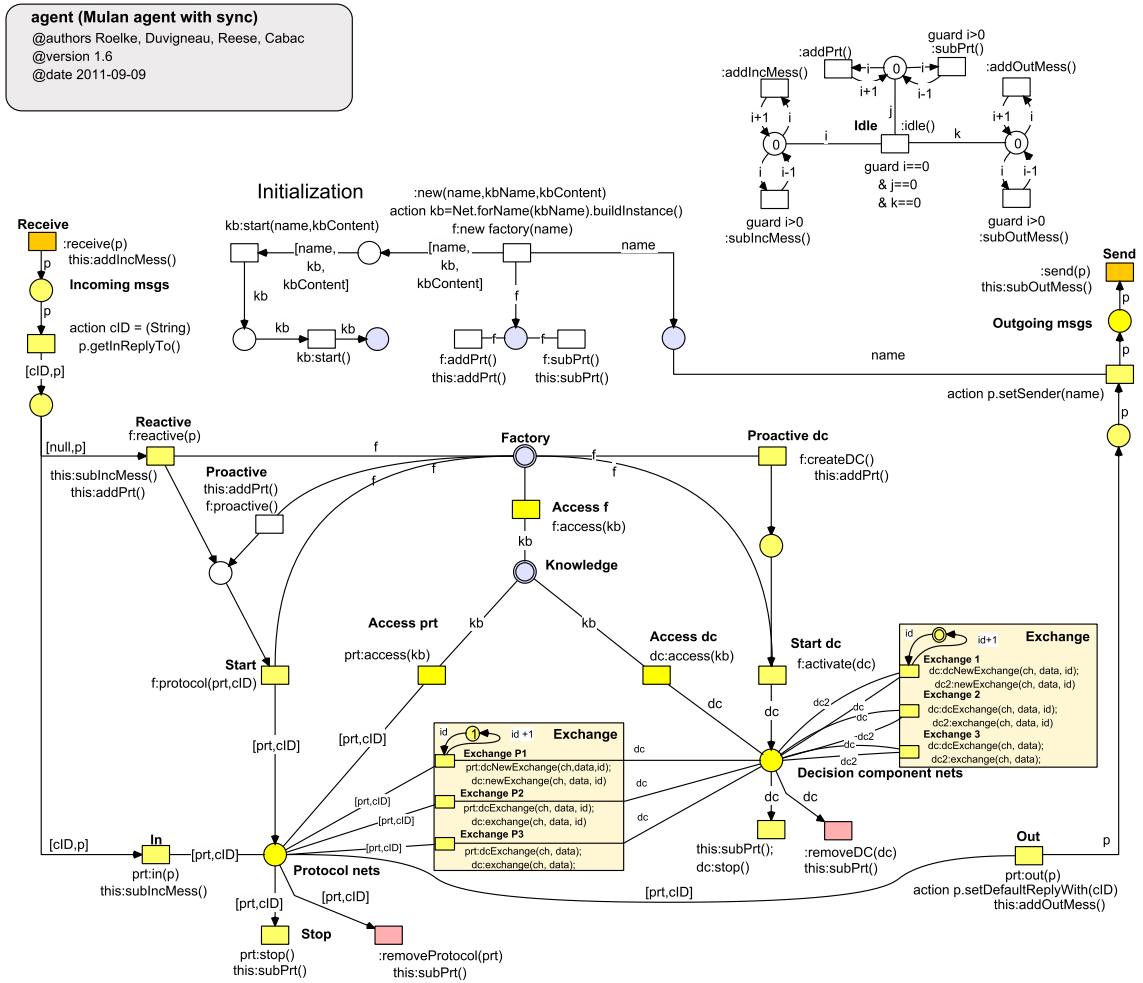


Abbildung 8.17.: Agentenmodell (aus [FDM11]).

8. Referenzmodell für Emotionsmodellierung

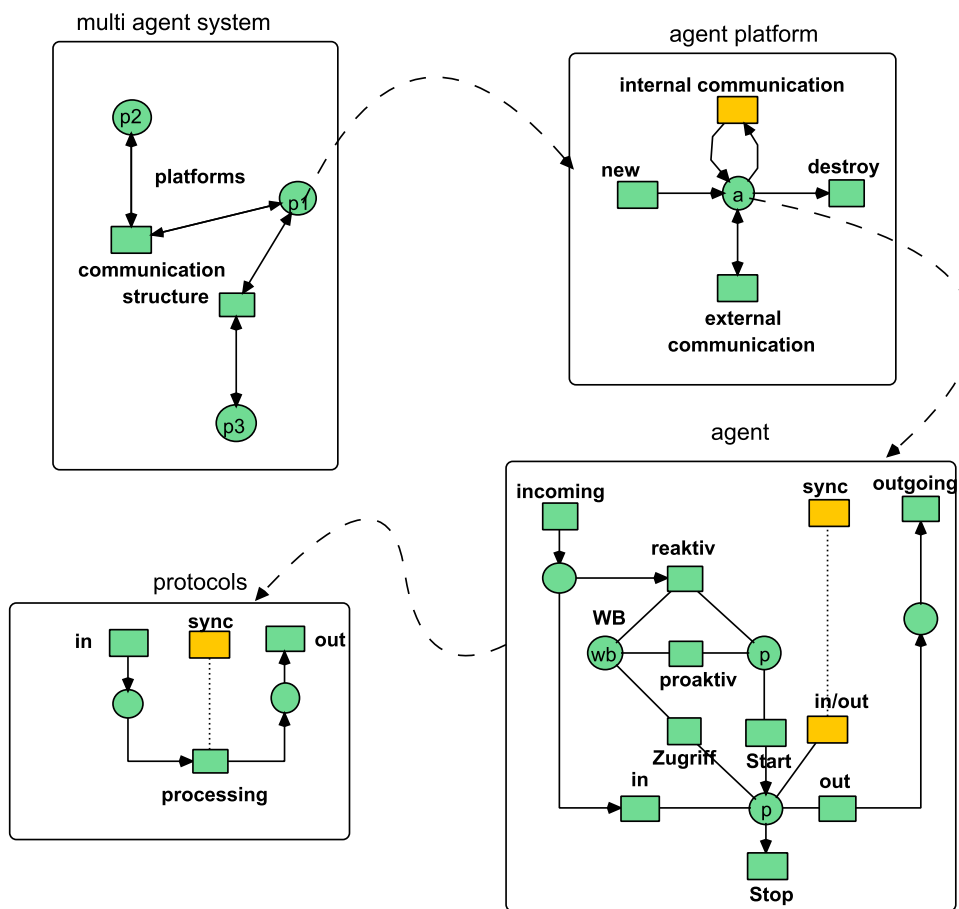


Abbildung 8.18.: MULAN aus [KLR03], modifiziert mit den zur Bereitstellung des Synchronisationsmechanismus vorgenommenen Erweiterungen bzw. Anpassungen (farblich hervorgehoben)

8.4. Verknüpfung der rationalen und der emotionalen Modellkomponenten

Die gepunktete Verbindung zwischen den beiden beteiligten Transitionen symbolisiert zunächst einen fest verdrahteten synchronen Kanal (quasi eine permanente Verschmelzung der beiden Transitionen). Diese Verbindung kann in einer verfeinerten Versionen des Modells durch ein Unternetz weiter ausgestaltet werden, um dem Agentennetz die Kontrolle über die Kommunikationsmöglichkeiten seiner Protokolle zu geben.

Auf der Ebene der Protokollnetze wird ebenfalls ein neuer Kanal in die Schnittstelle eingefügt. Neben **start** und **stop** (Protokoll-Lebenszyklussteuerung), **in** und **out** (asynchrone Kommunikation) sowie **access** und **exchange** (agenteninterner Informationsfluss) steht nun ein Kanal **in/out** für synchrone Kommunikation zur Verfügung. Dieser Kanal nimmt zum einen Meta-Informationen über die angestrebte Interaktion auf, insbesondere Angaben zum intendierten Partner. Zum anderen bietet er einen frei unifizierbaren Parameter, über welchen die kommunizierenden Protokolle beliebige Daten in beide Richtungen austauschen können. Die Synchronisation von Kontrollfluss und Information zwischen den Interaktionspartnern wird somit vollständig durch die Bindungssuche des zugrunde liegenden Petrinetz-Simulators hergestellt.

Der hier für E-MULAN präsentierte Synchronisationsmechanismus erweitert auch MULAN um eine synchrone Schnittstelle: Er ist nicht auf die Anwendungsdomäne der Emotionstheorien beschränkt. Jegliche zu synchronisierenden Agenten können mittels dieses Verfahrens modelliert werden. Insbesondere wird damit eine technische Umsetzung für die CAPA-Umgebung vorgestellt. Inwieweit andere Agentenmodelle von dem hier vorgestellten Verfahren Anregungen für eine synchrone Architekturweiterung übernehmen können, bleibt in weiterführenden Arbeiten zu klären.

Während Agenten ihre Autonomie dadurch behalten, dass sie die Protokolle selbstständig starten können, wird wie bisher für die asynchronen Protokolle eine statische Realisierung auf der Agentenebene implementiert. Hier könnte eine weitere Flexibilisierung durch die Einführung einer vermittelnden Stelle erreicht werden. Die bisher in Abb. 8.18 dargestellte gestrichelte Linie würde dann durch eine Stelle mit dem vermittelnden Netz, das die jeweiligen Zugriffe auf die synchronen Protokolle auf spezifische Art erlaubt, ausgestattet werden. Die Behandlung von Parametern ist ein weiterer Aspekt, der bzgl. der Synchronisation von Bedeutung ist. Durch die bi-direktionale Kommunikation, die zudem durch die Schaltregel bzw. den Simulator realisiert wird, können beliebig komplexe (synchrone) Funktionen realisiert werden. Mit der hier vorgestellten Lösung wird jedoch nicht die Plattform übergreifende Synchronisation gelöst. Dafür ist eine wesentlich komplexere Lösung erforderlich, wobei es für die aktuell möglichen Implementierungen von Modellen keine realistische Lösung für den verteilten Fall gibt. Örtlich verteilte Plattformen müssen also mit einer anderen Art der Synchronisation ausgestattet werden. Die hier vorgestellte Lösung liefert aber für die E-MULAN und MULAN / CAPA Umgebungen eine ausdrucksstarke und leistungsfähige Erweiterung.

8.4.4. Realisierung der emotionsbasierten Steuerung unter Verwendung der synchronen Kommunikation

In den E-MULAN-Modellen in Kapitel 8.3 haben wir noch nicht von der Möglichkeit der synchronen Kommunikation für MULAN- bzw. E-MULAN-Agenten Gebrauch gemacht -

8. Referenzmodell für Emotionsmodellierung

die beiden Agentenkomponenten von E-MULAN haben ausschließlich über asynchronen Nachrichtenaustausch interagiert.

Nun soll exemplarisch gezeigt werden, wie sich das oben dargestellte Szenario der emotionsbasierten Steuerung (vgl. Kapitel 8.3) unter Verwendung des hier eingeführten Synchronisationsmechanismus auf eine einfachere Weise realisieren lässt, als es mit einem herkömmlichen asynchronen MULAN-Modell möglich war.

Um dies zu illustrieren, wird hier wiederum ein stark vereinfachtes, aber ausführbares MULAN-Modell verwendet. Das Modell aus Kapitel 8.3.2 wird hier um den Synchronisationsmechanismus ergänzt. Dazu wird, wie oben beschrieben, eine Transition **synchronous communication** auf der Plattform-Ebene eingeführt sowie die korrespondierenden Transitionen **sync** und **inout** auf Agenten- bzw. Protokollebene.

Mit diesem erweiterten Modell wird in Abb. 8.19 exemplarisch eine synchrone Kommunikation zwischen einem rationalen MULAN-Protokoll und einem emotionalen Steuerungsprotokoll abgewickelt und mit einem Kausalnetz veranschaulicht.

Die dazugehörigen Netze sind in den Abb. 8.20 (Agentennetz), 8.21 und 8.22 (die beiden Protokollnetze), 8.23 (Plattform) und 8.24 (Protokollfabrik) dargestellt. Beim Vergleich des Kausalnetzes 8.19 mit einem entsprechenden Kausalnetz in Kapitel 8.3.2 (Abb. 8.15) sieht man, dass sich die Agenteninteraktionen durch die Gestattung synchroner Kommunikation wesentlich einfacher gestalten und beschreiben lassen.

Mit dem ausführbaren Kausalnetz aus Abb. 8.19 kann man das oben beschriebene Szenario simulieren und das Verhalten der Modellkomponenten beobachten. Bei einer entsprechenden Erweiterung der Netze wird das Kausalnetz mit den vier oben dargestellten Architekturkomponenten (Plattform, Protokollfabrik, Agent, Protokoll) verknüpft und synchronisiert. Damit kann man das Verhalten der Netze dynamisch zur Laufzeit beobachten und den aktuellen Zustand des Systems zu einem beliebigen Zeitpunkt der Simulation nachverfolgen. Die notwendigen Erweiterungen für diese Synchronisation wurden mit der blauen Farbe der Netzelemente bzw. der Netzanschriften gekennzeichnet und werden in Abb. 8.25 bis 8.30 gezeigt.

Die Wechselwirkungen von emotionalen und rationalen Architekturkomponenten werden im nächsten Kapitel erneut aufgegriffen und in weiteren Anwendungsszenarien im Zusammenhang mit dem PAOSE-Modellierungsansatz analysiert (vgl. Kapitel 9.4).

8.5. Diskussion des Referenzmodells

Unter der Annahme, dass Emotion einen wichtigen Einfluss auf menschliche intelligente Prozesse ausübt (wie Entscheiden, Verhalten, Planen etc.), wurde in diesem Kapitel, ausgehend von den oben vorgestellten emotionstheoretischen Grundlagen und der eher rationalitätsorientierten Multiagentensystemarchitektur MULAN, ein Referenzmodell vorgeschlagen, das neben der Modellierung der kognitiven auch eine explizite Modellierung der emotionalen Prozesse erlaubt.

In diesem Abschnitt diskutieren wir das dargestellte Referenzmodell, gehen auf offene Fragen ein und zeigen prinzipielle Anwendungsmöglichkeiten und Grenzen des Modells auf.

- Der Erklärungsanspruch des E-MULAN-Modells gilt den allgemeinen sozialen Zusammenhängen in einer (künstlichen oder natürlichen) Gesellschaft mit einer speziellen Ausrichtung auf die Funktion bzw. Bedeutung der Emotion darin. Den Startpunkt stellen psychologische und soziologische Emotionsgrundlagen dar, prinzipiell soll sich mit dem vorgeschlagenen Referenzmodell jedoch eine beliebige Emotionstheorie nachbilden lassen. U.a. sollte in zukünftigen Arbeiten die Möglichkeit der Nachbildung von neurowissenschaftlichen, biologischen, philosophischen und anderen Theorien der Emotion untersucht werden.
- E-MULAN stellt ein Referenzmodell dar, in dem man die blinden Flecken der Emotionstheorien erkennen kann. So kann durch die Nachbildung einer Emotionstheorie in E-MULAN erkannt werden, welche Prozesse von der Theorie nicht (vollständig) beschrieben werden. Man würde auf diese Weise etwa das Fehlen einer soziologischen Fundierung einer Theorie durch das Fehlen der Beschreibung der Prozesse / Konzepte der entsprechenden Modellebenen (Gesellschaftsebene bzw. Ebene der sozialen Struktur) sofort erkennen können. Ebenso würde die fehlende Spezifizierung der Auswirkungen der Emotion auf das Handeln dadurch erkannt werden, dass die Prozesse der emotionsbasierten Verhaltenssteuerung in der jeweiligen Emotionstheorie nicht beschrieben sind und folglich nicht modelliert werden können.

Das Referenzmodell kann deshalb zur Identifizierung der Grenzen von Emotionstheorien eingesetzt werden, um zu erkennen, welche Emotionstheorien für bestimmte informatische Anwendungen besonders geeignet bzw. welche Modelle in Bezug auf die geforderten Eigenschaften eines emotionalen Systems weniger aussagekräftig sind. So kann eine Vorauswahl getroffen werden und die Modellierungs- / Implementierungsaktivitäten können gezielt auf die zur geplanten Anwendung passenden Emotionstheorien ausgerichtet werden.

- Die Betrachtung der soziologischen Theorien in Kapitel 6.3 hat gezeigt, dass Emotionen maßgeblich an der Aufrechterhaltung und Reproduktion sozialer Strukturen bzw. sozialer Ordnung beteiligt sind. Diese Sichtweise wird verstärkt, je intensiver man Emotionstheorien aus angrenzenden Disziplinen wie der Psychologie und den Neurowissenschaften berücksichtigt. Bislang fehlt jedoch eine umfassende Emotionstheorie, die sich dieser spezifischen Funktion von Emotionen widmet und dabei die weit reichenden gesellschaftstheoretischen Implikationen beleuchtet.

Das E-MULAN-Referenzmodell bietet die Möglichkeit zur gleichzeitigen Modellierung mehrerer Emotionstheorien auf verschiedenen Architekturebenen. Es kann deshalb die emotionstheoretischen Ansätze aus verschiedenen Disziplinen integrieren. So ließen sich mit der vorgeschlagenen Architektur sowohl emotionale Abläufe auf der Makroebene soziologischer Strukturen als auch psychologische und neurobiologische Prozesse auf der Mikroebene darstellen. Ferner bietet sie die Möglichkeit zur formalen Spezifikation der emotionalen Abläufe und ihrer Abbildung in einem informatischen System.

- Durch die Darstellung der Wechselwirkungen von individuellem Handeln und sozialen Strukturen in dem E-MULAN-Referenzmodell ließe sich die Funktion der Emotio-

8. Referenzmodell für Emotionsmodellierung

nen explizit darstellen und untersuchen. Ferner erlaubt sie es grundsätzlich, logische Inkonsistenzen in den erstellten Theoriemodellen aufzudecken und abweichendes Modellverhalten zu erkennen. Somit könnte das E-MULAN-Referenzmodell Unterstützung bei der Formulierung von sozialwissenschaftlichen Theorien bieten.

- Ferner bietet das Referenzmodell Repräsentationsmöglichkeiten für verschiedene emotionale Konzepte / Prozesse und ermöglicht einen Übergang zur operationalen Darstellung durch die Überführung der konzeptionellen Modelle in ausführbare Systemarchitekturen. Als Erweiterung von MULAN stellt es grundsätzlich ein ausführbares Modell dar und könnte bei entsprechender Anpassung zur Simulation sozialwissenschaftlicher Theorien bzw. Theorienaspekte eingesetzt werden.
- Das E-MULAN-Referenzmodell kann zur Nachbildung der existierenden informatischen Emotionsmodelle / emotionalen Agentenarchitekturen eingesetzt werden, um sie zueinander in Beziehung zu setzen und miteinander zu vergleichen. Angesichts der starken Divergenz der zugrunde liegenden Emotionsdefinitionen und Vorannahmen in verschiedenen Modellen könnte dies eine Basis für eine bessere Vergleichbarkeit verschiedener Modellierungsansätze, mehr Transparenz und weniger Fragmentierung in der informatischen Emotionsforschung sein.
- Vor dem Hintergrund der Forschungsfragmentierung im Bereich der Emotionserforschung und -modellierung ist die Möglichkeit der Integration verschiedener emotionstheoretischer Ansätze in einem Modell besonders interessant.
- Als eine der Anforderungen an das Referenzmodell wurde die Möglichkeit der konzeptionellen Trennung bei der Repräsentation “rationaler“ / kognitiver und emotionaler Prozesse definiert. Diese Anforderung basiert auf neurophysiologischen Untersuchungsergebnissen, die besagen, dass emotionale und kognitive Verarbeitungsprozesse nach der Wahrnehmung eines Stimulus parallel ablaufen und sich gegenseitig ergänzen [LeD00]. Die explizite Trennung in rationale und emotionale Modellkomponenten soll im Referenzmodell in erster Linie Forschungszwecken dienen: Zum einen ermöglicht sie die explizite Darstellung und Erforschung der Wechselwirkungen dieser zwei Mechanismen; zum anderen ist dadurch die Auswertung des Potenzials der Emotion für die Optimierung der (zurzeit meist rationalen) (Multi-) Agentenarchitekturen möglich. Es handelt sich also um eine rein analytische Trennung, die vor allem auf die konzeptuelle Abgrenzung und Analyse der zugrundeliegenden komplementären Prozesse abzielt.
- Um rationale und emotionale Modellkomponenten des E-MULAN-Referenzmodells wieder zusammenzuführen, muss das Referenzmodell um einen Synchronisationsmechanismus erweitert werden (vgl. Kapitel 8.4). Dies lässt sich für einfache Abläufe oft durch entsprechende Interaktionen, basierend auf Nachrichtenaustausch, realisieren. Komplexe Interaktionen, die wechselseitige Abhängigkeiten beinhalten und die von außen als eine gemeinsame Aktion betrachtet werden sollen, erfordern das ausdrucksstärkere Konzept der Synchronisation.

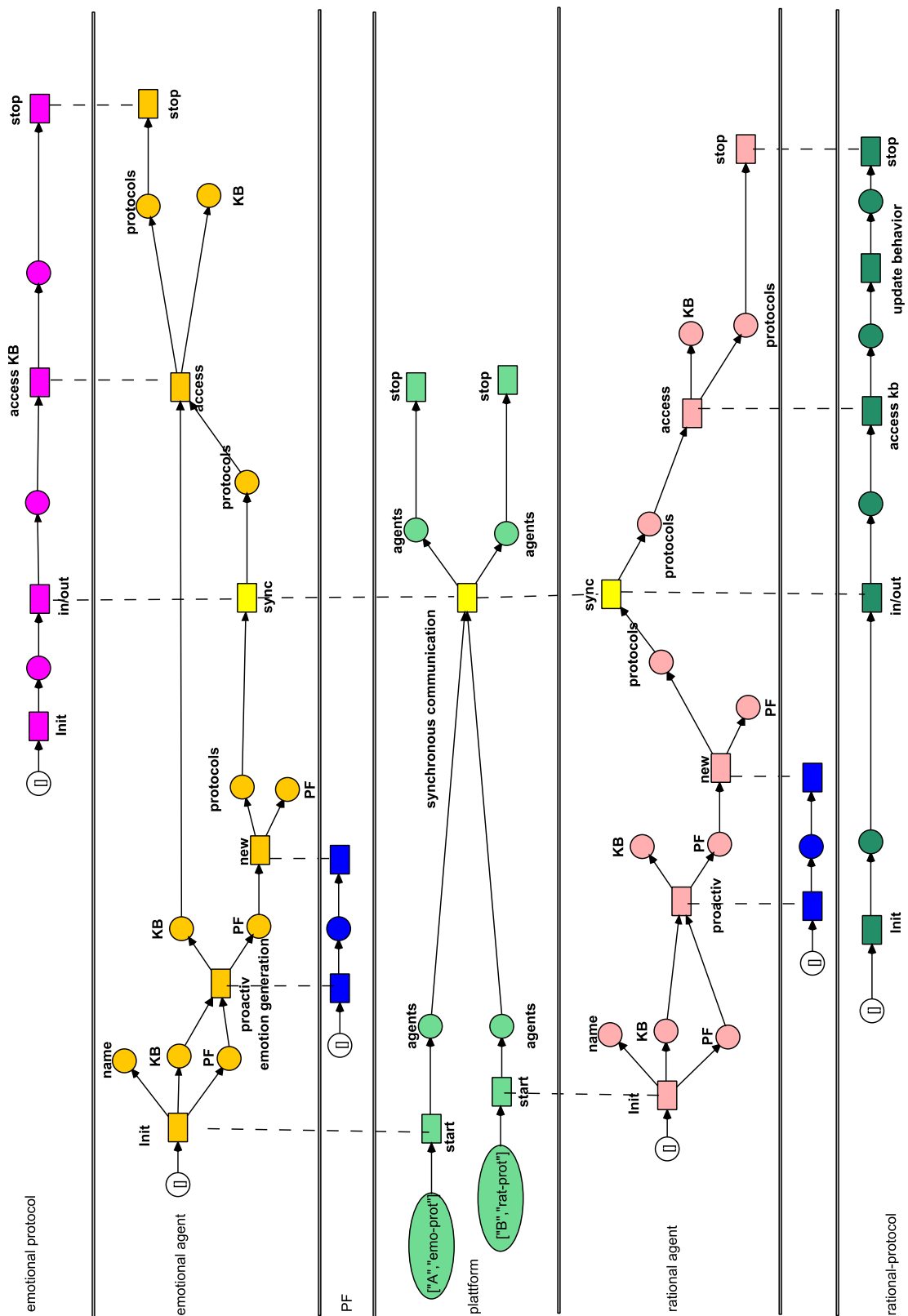
In Kapitel 8.4.3 wurde ein Synchronisationsmechanismus für die engere Kopplung von Modellen dargestellt und seine vollständige operationale Semantik mittels Referenznetzen für den MULAN-Kontext aufgezeigt. Damit lassen sich die in Kapitel 8.2 zuvor analytisch getrennten rationalen Bestandteile mit den emotionalen Bestandteilen der E-MULAN-Architektur wieder vollständig und effizient integrieren.

Das E-MULAN-Referenzmodell soll sowohl als Modellierungsrahmen zum Entwurf einer konkreten ausführbaren emotionalen (Multi-)Agentenarchitektur als auch als konzeptuelles und terminologisches Referenzmodell zur Klassifizierung und zum Vergleich sozialwissenschaftlicher Emotionstheorien und informatischer Systemarchitekturen eingesetzt werden können. In dieser zweiten Funktion kann E-MULAN einen Denkraum und eine Repräsentationssprache für zukünftige Systementwürfe agentenorientierter Emotionsarchitekturen bilden. Zudem liegt aufgrund der operationalen Semantik durch die Petrinetze eine Simulation bzw. in Verbindung mit MULAN, CAPA und dem PAOSE-Ansatz die direkte Implementierung nahe.

Die Grenzen des vorgeschlagenen Referenzmodells sind in erster Linie dadurch bedingt, dass es vor allem ein Modellierungswerkzeug für die Repräsentation sozialwissenschaftlicher Theorien zur Verfügung stellt, ohne selbst eine Emotionstheorie darzustellen. Petrinetze erlauben die Darstellung aller Arten von Prozesse. E-MULAN spezifiziert die Strukturen zur Gestaltung der Beziehungen der beteiligten Einheiten.

Das E-MULAN-Referenzmodell liefert somit eine um Emotion verbesserte Architektur für Agentensysteme, ohne jedoch einen Bezug zu speziellen Anwendungsbereichen herzustellen, d.h. es handelt sich um eine allgemeine und weitgehend universell einsetzbare Architektur. Die weiterführende Untersuchung soll klären, für welche Anwendungsfelder innerhalb des breiten Forschungsgebiets "Affective Computing" die E-MULAN-Architektur bzw. die auf Basis von E-MULAN entwickelten Agentenarchitekturen besonders geeignet sind.

8. Referenzmodell für Emotionsmodellierung



342 Abbildung 8.19.: Modell der emotionsbasierten Verhaltenssteuerung: Interaktion eines MULAN-Protokolls mit einem emotionalen Steuerungsprotokoll in E-MULAN.

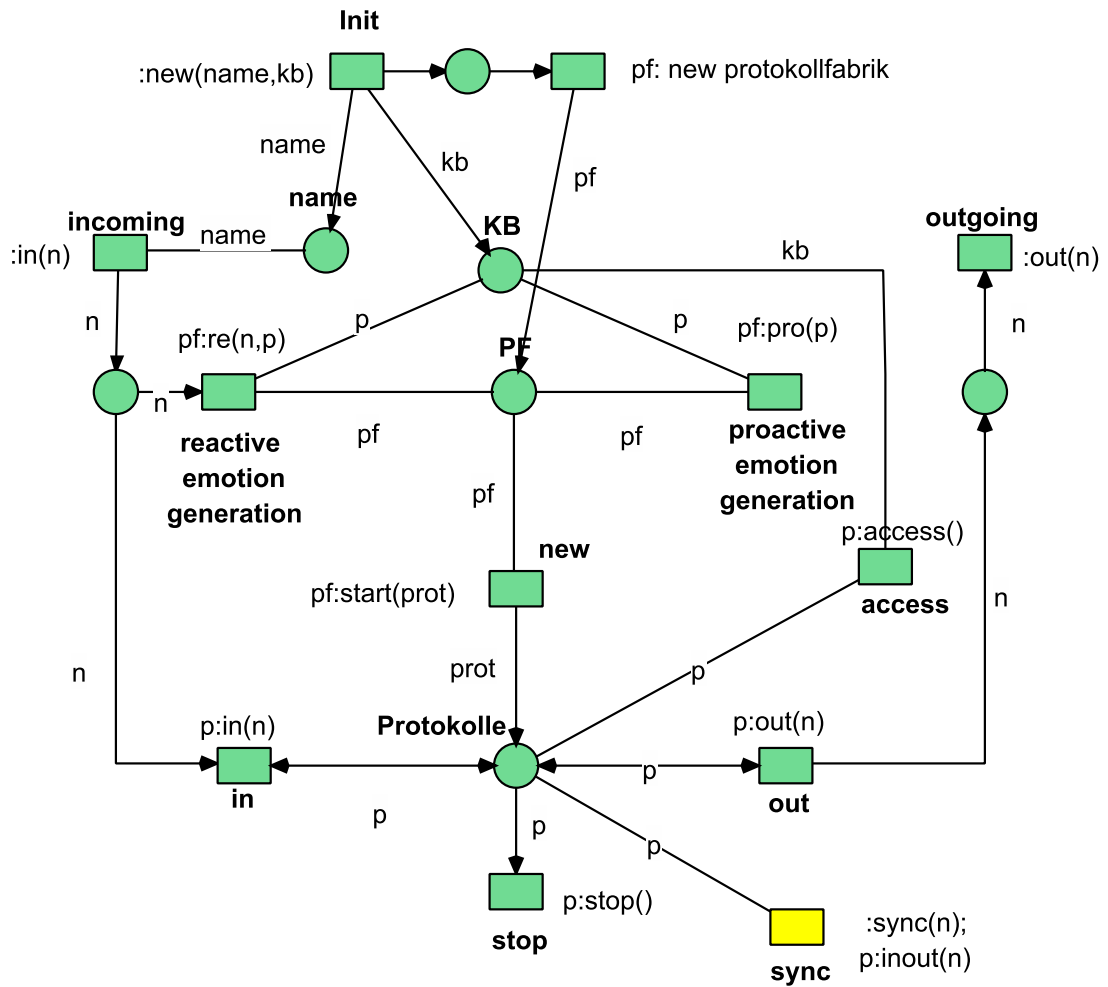


Abbildung 8.20.: Stark vereinfachtes Agentennetz.

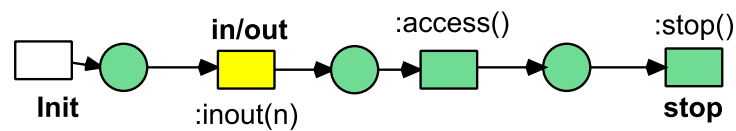


Abbildung 8.21.: Emotionales Protokoll

8. Referenzmodell für Emotionsmodellierung

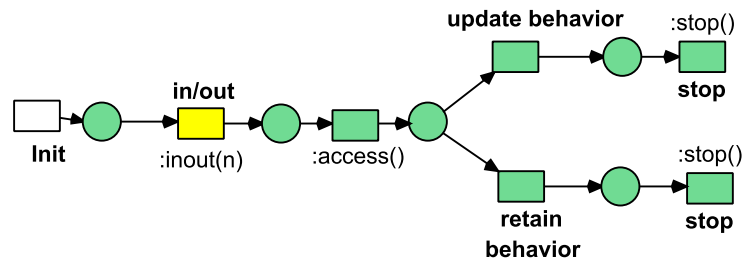


Abbildung 8.22.: Rationales Protokoll

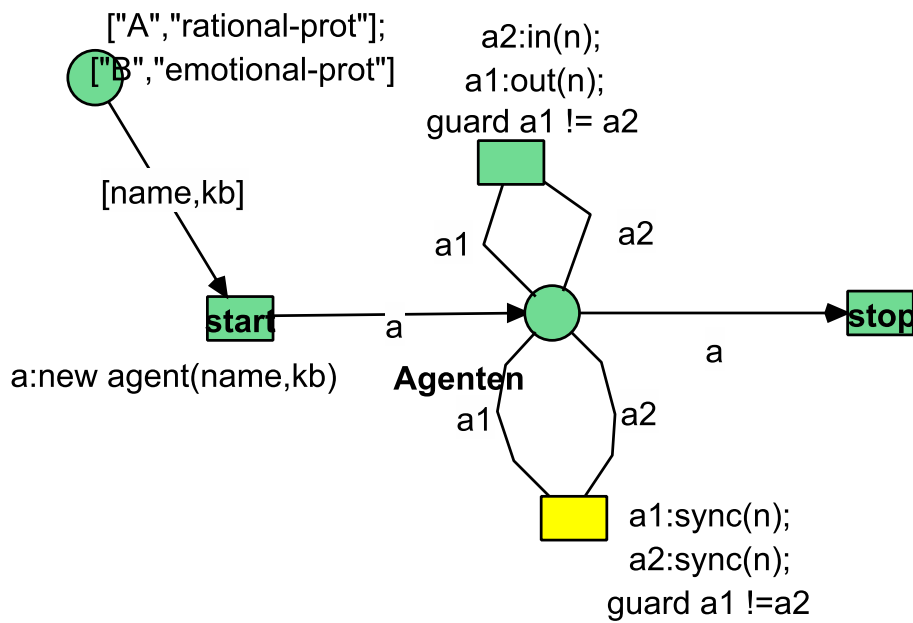


Abbildung 8.23.: Plattform-Netz

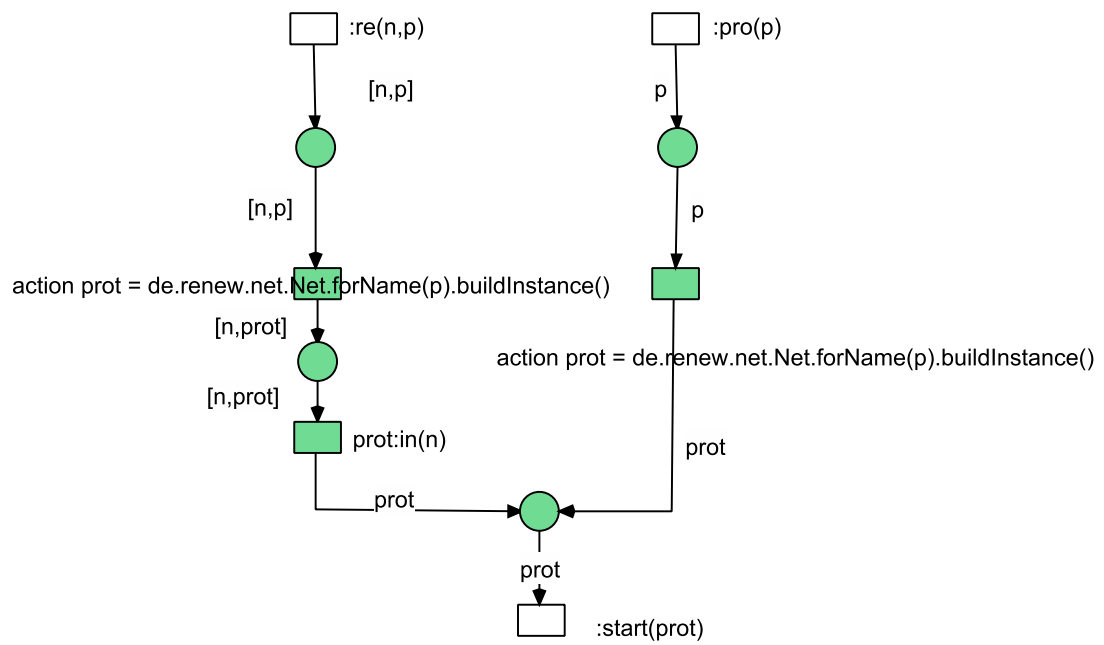
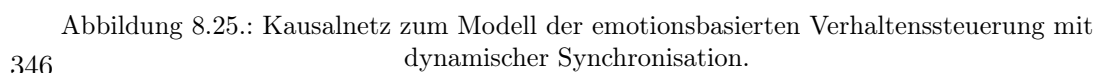


Abbildung 8.24.: Protokollfabrik

346



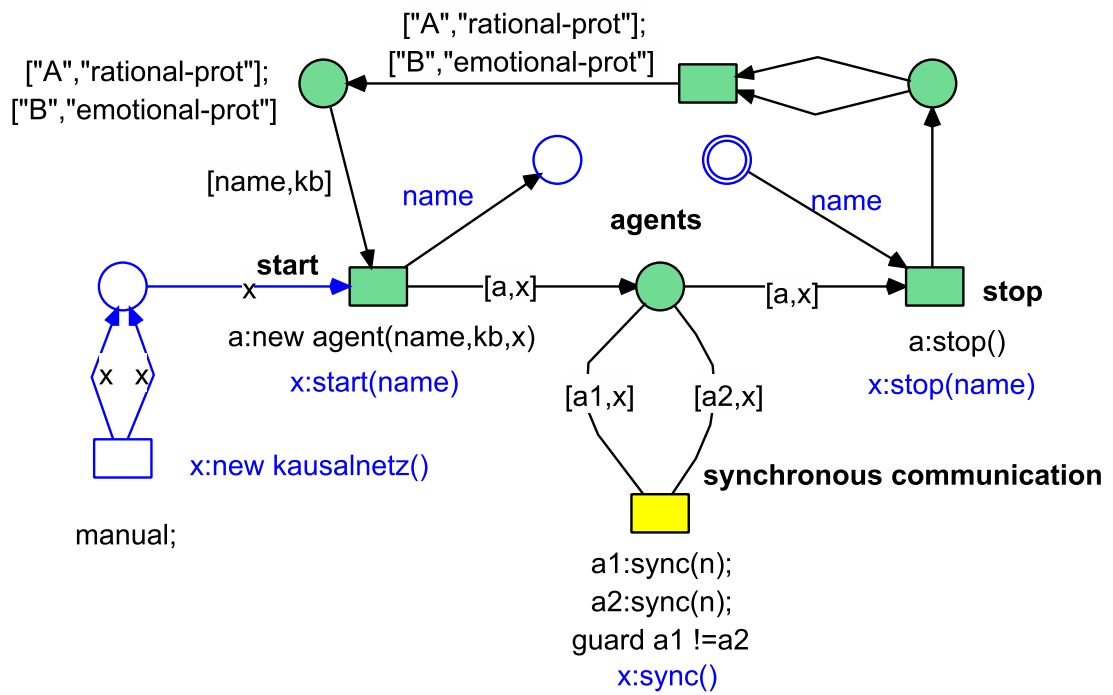


Abbildung 8.26.: Plattform-Netz, erweitert für die Synchronisierung mit dem Kausalnetz aus Abb. 8.25.

8. Referenzmodell für Emotionsmodellierung

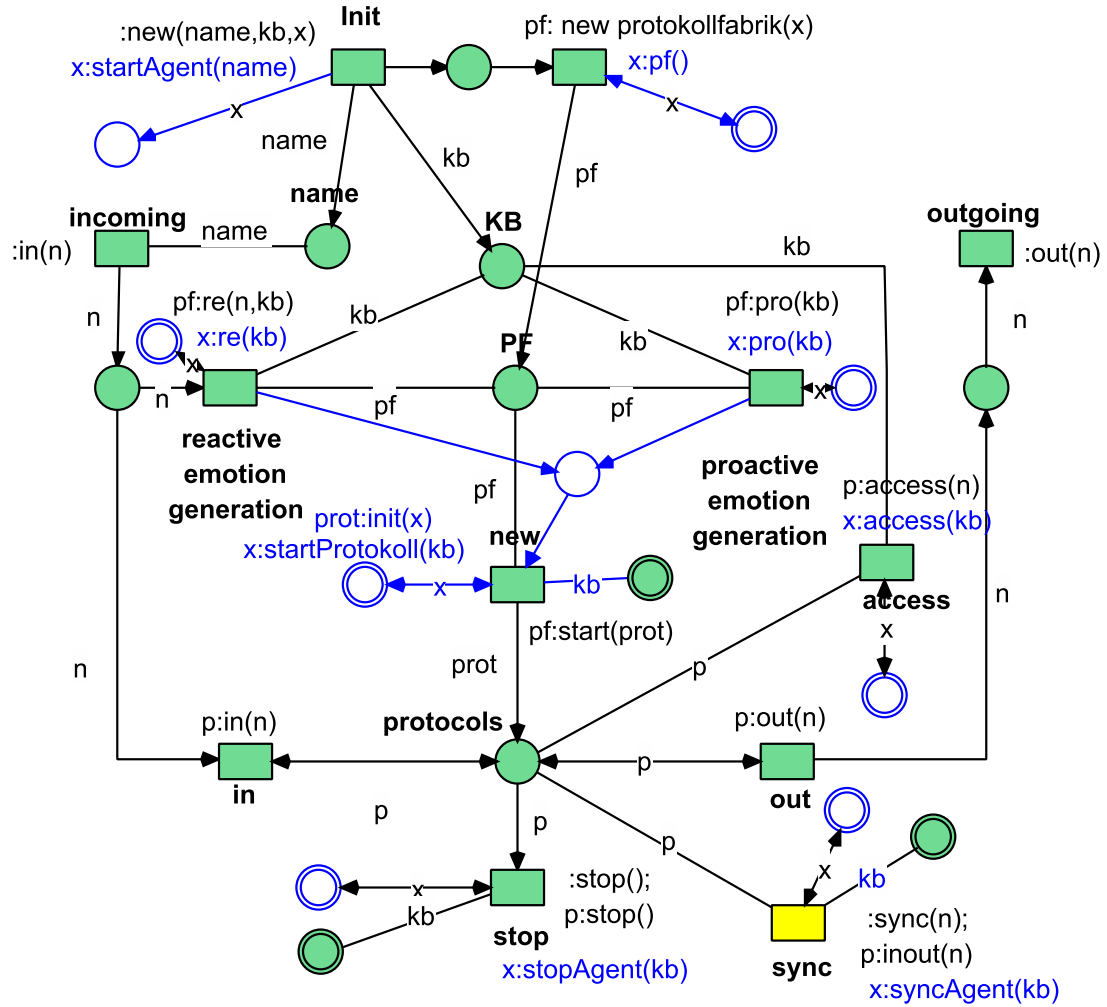


Abbildung 8.27.: Agentennetz, erweitert für die Synchronisierung mit dem Kausalnetz aus Abb. 8.25.

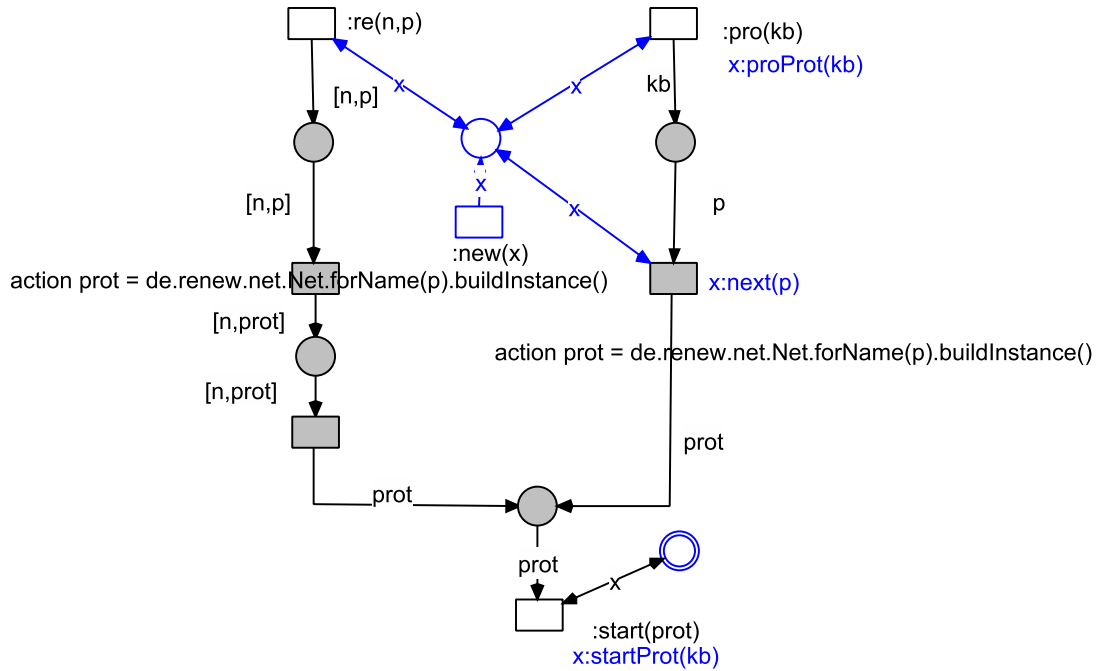


Abbildung 8.28.: Protokollfabrik, erweitert für die Synchronisierung mit dem Kausalnetz aus Abb. 8.25.

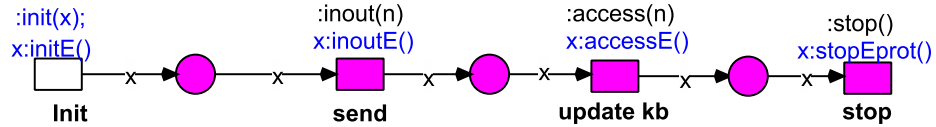


Abbildung 8.29.: Emotionales Protokoll, erweitert für die Synchronisierung mit dem Kausalnetz aus Abb. 8.25.

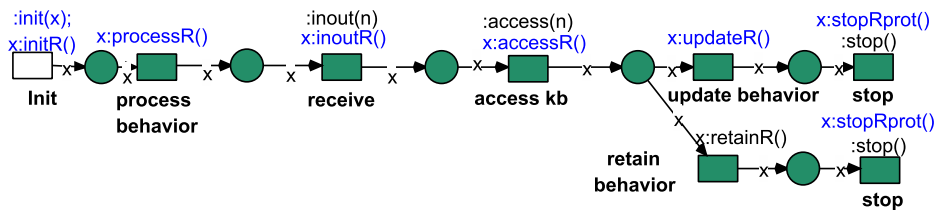


Abbildung 8.30.: Rationales Protokoll, erweitert für die Synchronisierung mit dem Kausalnetz aus Abb. 8.25.

9. Anwendungsbeispiele

Nach einer Übersichtsdarstellung und der Veranschaulichung des E-MULAN-Referenzmodells an einigen Anwendungsbeispielen mit Hilfe der vereinfachten MULAN-Architektur im vorigen Abschnitt werden wir in diesem Abschnitt die E-MULAN-Referenzarchitektur in Verbindung mit dem PAOSE-Entwicklungsansatz zur Modellierung von z.T. denselben Szenarien, jedoch auf einer höheren Abstraktionsebene anwenden.

9.1. Grundsätzliche Annahmen und Modellierungsansatz

Bei der Konzeption der E-MULAN-Architektur wurde eine Modellierungsbasis für die Darstellung emotionaler Prozesse in Agentenanwendungen (vgl. Kapitel 8.1) angestrebt. Im Folgenden werden mit E-MULAN exemplarisch ausgewählte emotionale Prozesse modelliert - sowohl solche, die innerhalb eines Agenten als interne Verarbeitungsprozesse ablaufen, als auch solche, die bei der Interaktion mehrerer Agenten stattfinden. Als Anwendungsbeispiele werden vorzugsweise Szenarien gewählt, die bereits im theoretischen Teil der Arbeit ausführlich diskutiert wurden: Insbesondere wird die Anwendung des Referenzmodells E-MULAN zur Modellierung der reaktiven (Kapitel 9.2) und proaktiven (Kapitel 9.3) Emotionsgenerierung, zur Modellierung der emotionsbasierten Verhaltenssteuerung (Kapitel 9.4) sowie zur Etablierung von normkonformen Verhalten (Kapitel 9.5) illustriert.

In Kapitel 8.1 wurde des Weiteren gefordert, dass E-MULAN als generisches Modell eine explizite Repräsentation sowohl rationaler / kognitiver als auch emotionaler Prozesse innerhalb desselben Rahmenwerks ermöglichen sowie insbesondere die Wechselwirkungen zwischen diesen Prozessen darstellen soll. In Kapitel 8 wurden im Rahmen der Vorstellung der Referenzmodell-Architektur die analytische Trennung in rationale und emotionale Modellkomponenten und die verschiedenen Verknüpfungsmöglichkeiten dieser Komponenten diskutiert. In diesem Kapitel wird die konkrete Umsetzung beschrieben und an verschiedenen Anwendungsbeispielen illustriert.

Modellierungstechnisch findet der PAOSE-Modellierungsansatz (siehe Kapitel 4.3) sowie das Referenznetz-Modellierungswerkzeug RENEW mit den entsprechenden Plugins Anwendung. Eine ausführliche Beschreibung der grundlegenden Konzepte und der Vorgehensweise bei der Anwendungsmodellierung mit PAOSE findet sich z.B. in [Mol06], in [Cab10] oder in [Ree10].

In dieser Arbeit werden nur ausgewählte Elemente des PAOSE-Ansatzes für die Anwendungsentwicklung genutzt, wobei allgemein bei den vorgestellten Anwendungsmodellierungen wie folgt vorgegangen wird: Zuerst werden die jeweiligen emotionstheoretischen Grundlagen zusammengefasst und ggf. mit Petrinetz-Modellen illustriert. Im nächsten Schritt werden die Interaktionen zwischen den rationalen und emotionalen E-MULAN-

9. Anwendungsbeispiele

Komponenten mit Hilfe der Interaktionsprotokolle dargestellt. Die Interaktionsprotokolle finden im Rahmen des PAOSE-Ansatzes Anwendung: Sie können in RENEW als Agenteninteraktionsprotokolle erstellt (AIP-Diagramme) und dann mit Hilfe eines AIP-Plugins [Cab10] automatisch in die lauffähigen Netzstrukturen für Protokollnetze überführt werden. Diese Netzstrukturen werden während der Implementierung mit Anschriften vervollständigt, damit sie zu ausführbaren Protokollnetzen werden, die das Agentenverhalten bzw. die Emotionsgenerierungsprozesse implementieren. Weitere Details zu den notwendigen Implementierungsschritten bei der Erstellung einer PAOSE-Anwendung finden sich in den Kapitel 4.1.4 und 4.3 oder können in [Cab10], S. 115 ff., oder [Ree10], S. 84 ff., nachgelesen werden.

Die für diesen Abschnitt wichtigen emotionstheoretischen Grundlagen sind bereits in den Kapiteln 6.1, 6.2, 6.3 und 7 erläutert worden, indem dort ausgewählte Emotionstheorien und Agentenarchitekturen mit Petrinetz-Modellen dargestellt wurden. Auf diese Modelle kann nun bei der Darstellung emotionaler Konzepte und Prozesse mit E-MULAN zurückgegriffen werden.

9.2. Modell der reaktiven Emotionsgenerierung in E-Mulan

In Kapitel 8.1 wurde gefordert, dass E-MULAN als generisches Modell eine explizite Repräsentation sowohl rationaler / kognitiver als auch emotionaler Prozesse innerhalb des selben Rahmenwerks ermöglichen sowie insbesondere die Wechselwirkungen zwischen diesen Prozessen darstellen soll. In Kapitel 8 wurde die analytische Trennung in rationale und emotionale Modellkomponenten diskutiert, insbesondere die Möglichkeit, alle emotionalen Prozesse eines Agenten in einer Shadow-Komponente (einer sogenannten emotionalen Agentenkomponente) zu kapseln. Dies hat den Vorteil, dass man bei der Modellierung die emotionalen Modellkomponenten zu einem beliebigen Zeitpunkt ausblenden und ohne viel Aufwand zu einem konventionellen MULAN-Modell zurückkehren kann. Diese Modellierungsmöglichkeit wird hier für den Fall einer reaktiven Emotionsgenerierung ausführlicher dargestellt.

Die Agentenebene wird durch Auffaltung in ihre beiden Bestandteile zerlegt: den MULAN-Agenten und seine emotionale Shadow-Komponente (vgl. Abb. 8.2). Die beiden Komponenten sind wie herkömmliche MULAN-Agenten aufgebaut und können mit Hilfe der Plattform Nachrichten austauschen, Dienste anfordern etc. Gleichzeitig stellen die beiden Modellkomponenten im emotionstheoretischen Sinne die internen Verarbeitungsprozesse eines einzelnen Akteurs dar und können jederzeit zu einem einzigen Agenten zusammengefaltet werden.

Da es sich bei den beiden Komponenten modellierungstechnisch um MULAN-Agenten handelt, können die PAOSE-Techniken und -Werkzeuge bei der Entwicklung der Modelle ohne Einschränkung angewendet werden. Zuerst wird das zu modellierende Szenario in einem Interaktionsdiagramm dargestellt und unter Verwendung des AIP-Diagramm-Plugins von RENEW (vgl. [Cab10], S. 187) in die MULAN-Protokolle überführt. Anschließend werden die Wissensbasen der Agenten erstellt und die beiden Agentenkomponenten auf einer CAPA-Plattform gestartet. In der Simulation kann man die Interaktion der beiden

Agentenkomponenten bei der Erzeugung von reaktiven Emotionen auf verschiedenen Ebenen betrachten.

Szenario: Reaktive Emotionsgenerierung Das Szenario der reaktiven Emotionsgenerierung für einen einzelnen Agenten wurde bereits oben erläutert (siehe Kapitel 8.3.1 und 8.3.1) und wird hier zusätzlich als eine Interaktion zwischen den beiden Komponenten der E-MULAN-Agentenebene interpretiert: der rationalen Komponente ("rationaler Agent") und der emotionalen Shadow-Komponente ("emotionaler Agent").

Im ersten Schritt erhält der Agent einen Input aus der Umgebung, auf den er ggf. emotional reagieren soll. Dieser Input ("Stimulus") wird von der rationalen Agentenkomponente an die emotionale Shadow-Komponente zur weiteren Auswertung weitergeleitet. In der emotionalen Agentenkomponente wird geprüft, ob der Stimulus "s" bekannt ist, d.h. ob dieser Stimulus in der Wissensbasis (der emotionalen Agentenkomponente) als Schlüssel hinterlegt ist. Wenn dies der Fall ist, wird die emotionale Agentenkomponente den in der Wissensbasis zum Stimulus "s" definierten Wert (Emotion "emotion") abrufen und als Nachricht an die rationale Agentenkomponente verschicken. Dieses Wert wird in der Wissensbasis der rationalen Agentenkomponente der Variable "Emotionszustand" zugewiesen, der Agent wird damit in den emotionalen Zustand "emotion" versetzt. Wenn in der Wissensbasis der emotionalen Agentenkomponente kein Emotionswert zum Schlüssel "s" hinterlegt ist, wird die Nachricht "no emotion" an die rationale Agentenkomponente geschickt.

Interaktionsdiagramm Das Interaktionsdiagramm in Abb. 9.1 zeigt eine mögliche Interaktion zwischen einer rationalen und einer emotionalen Agenten-Komponente bei der reaktiven Emotionsgenerierung. Diese Interaktion wird von dem rationalen Agenten proaktiv gestartet. Als Erstes schickt dieser dem emotionalen Agenten den aktuellen Stimulus-Wert *s*, um ggf. eine emotionale Reaktion zu *s* zu erhalten. Falls in der Wissensbasis des emotionalen Agenten eine bestimmte Emotion zum Stimulus *s* definiert ist, wird diese abgerufen und der entsprechende Wert an den rationalen Agenten geschickt (Nachricht *emotion*). Andernfalls wird ein rationaler Agent darüber informiert, dass für den Stimulus *s* keine emotionale Reaktion definiert ist (Nachricht "no emotion"). Daraufhin aktualisiert der rationaler Agent die Variable "meine Emotion" in seiner Wissensbasis mit dem aktuellen Wert von *emotion*.

Abhängig von dem Wert des emotionalen Zustandes könnten im nächsten Schritt Verhaltensprotokolle definiert werden, die mit der aktuellen Emotion korrespondieren. Auf diese Weise kann die emotionsbasierte Steuerung des Agentenverhaltens realisiert werden. Diese ist jedoch nicht Gegenstand des aktuellen Szenarios, sondern sie wird in Kapitel 9.4 behandelt.

In der Simulation wurden exemplarisch die emotionalen Zustände "frightened", "happy", "disgust" betrachtet, die durch die Auslöseereignisse "danger", "pleasant", "unpleasant" resp. hervorgerufen werden. Dieses Beispiel dient lediglich der Veranschaulichung der prinzipiellen Vorgehensweise bei der Modellierung der reaktiven Emotionen und bildet keinen konkreten emotionstheoretischen Ansatz ab. Nach diesem Muster könnte jedoch die Generierung einer beliebigen Emotionstaxonomie - etwa die Generierung von Basisemotionen - abgebildet

9. Anwendungsbeispiele

werden, die je nach Emotionstheorie unterschiedliche Emotionen und Auslösebedingungen erhalten kann.

Eine emotionstheoretisch korrekte Modellierung steht an dieser Stelle nicht im Mittelpunkt und sollte in weiterführenden Arbeiten betrachtet werden. Hier liegt vielmehr der Fokus auf modellierungstechnischen Aspekten, insbesondere auf der Veranschaulichung des gewählten Modellierungsansatzes.

Protokollnetze Die Agentenprotokolle wurden mit Hilfe des AIP-Diagramm-Plugins aus den Interaktionsdiagrammen erzeugt und angepasst. Die Protokollnetze zum Interaktionsprotokoll aus Abb. 9.1 sind in Abb. 9.2 für die emotionale und in Abb. 9.3 für die rationale Agentenkomponente dargestellt.

Wissensbasen Zur Repräsentation des Faktenwissens eines Agenten wird ein älteres einfaches Wissensbasisformat (.wis-Datei) verwendet. Die Wissensbasis wird manuell erstellt und dem Agenten bei der Initialisierung zugewiesen. Die Wissensbasis eines rationalen Agenten ist in 9.4, die eines emotionalen Agenten in Abb. 9.5 dargestellt.

Abb. 9.6 zeigt die beschriebene Beispielanwendung zur Laufzeit. Zu sehen ist die im MULANVIEWER visualisierte Wissensbasis des rationalen Agenten mit dem aktuellen emotionalen Zustand "frightened".

Modellierung mit DC Alternativ zu dem oben dargestellten Modellierungsansatz, bei dem alle Agenteninteraktionen und -aktionen direkt in den Protokollnetzen beschrieben werden, besteht die Möglichkeit, die Entscheidungsprozesse eines Agenten in die sogenannte Entscheidungskomponente (*Decision Component* - *DC*) auszulagern (vgl. Kapitel 4.2). In unserem Anwendungsbeispiel werden hierzu die vordefinierten Stimuli, die zu einer Emotion führen können, in die Entscheidungskomponente des rationalen Agenten ausgelagert und von dort aus vom Emotionsgenerierungsprotokoll abgerufen (vgl. Abb. 9.7).

9.2. Modell der reaktiven Emotionsgenerierung in E-MULAN

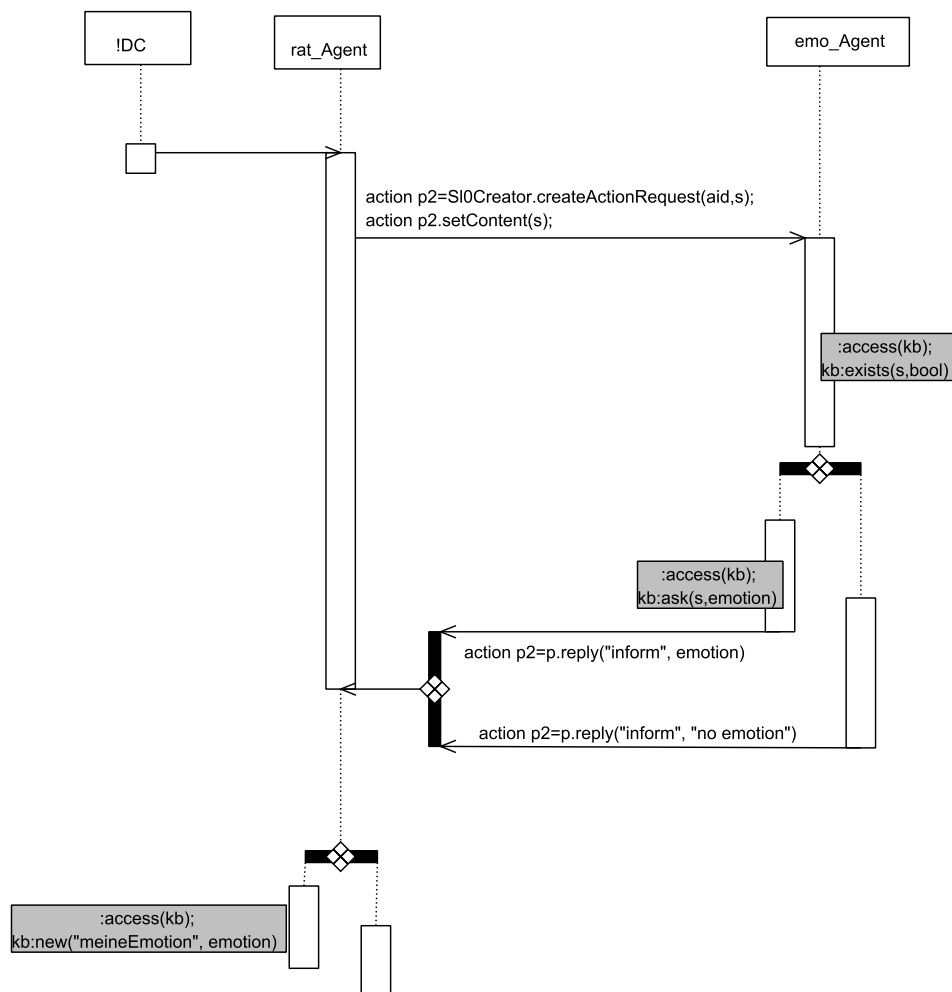
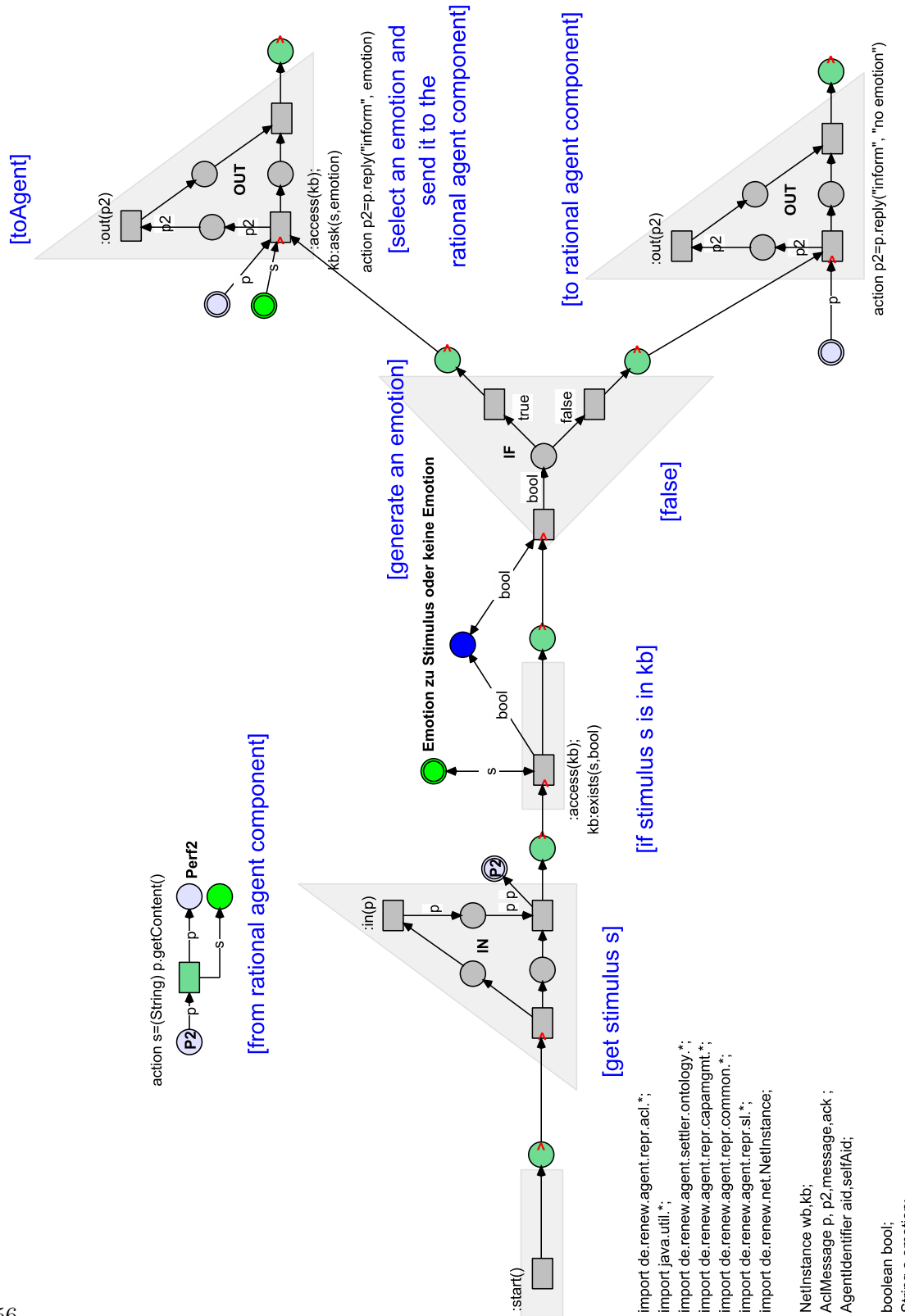


Abbildung 9.1.: Interaktionsdiagramm: Reaktive Emotionsgenerierung

9. Anwendungsbeispiele



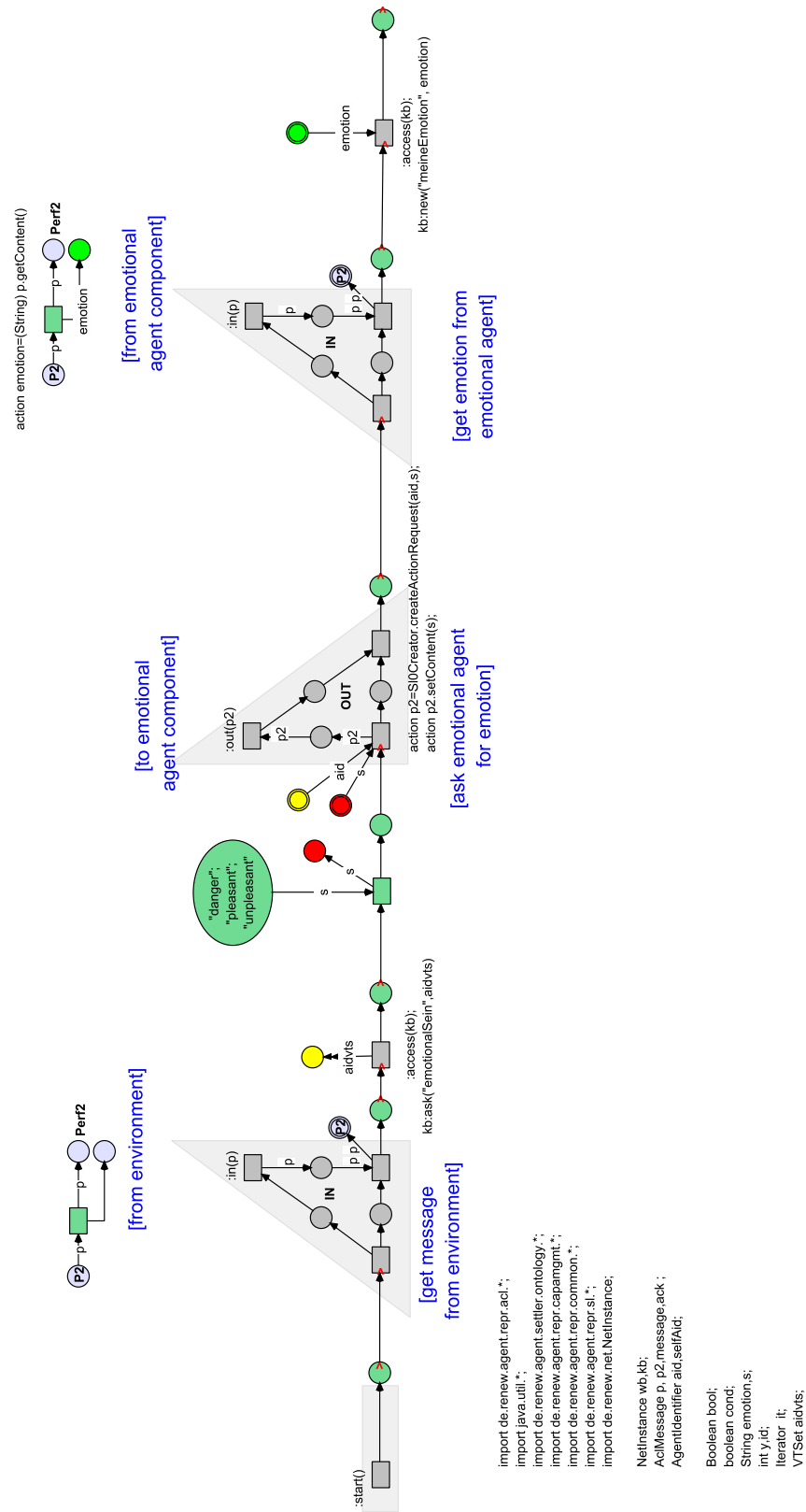


Abbildung 9.3.: Reaktive Emotionsgenerierung: Protokollnetz der rationalen Agentenkomponente

9. Anwendungsbeispiele

```
%%%.  
requiredServices=[(service-description :name "emotionalSein")]  
%%%  
protocol ams_move=(request :language FIPA-SLO  
:content "((action (agent-identifier) (move)))")  
%%%  
agentDesc=class de.renew.agent.repr.common.VTIndex  
%%%  
decisionComponent=rationalDCreaktiv
```

Abbildung 9.4.: Reaktive Emotionsgenerierung: Wissensbasis der rationalen Agentenkomponente

```
proaktiv=[generalAgentSetup]  
%%%  
serviceDesc=[(service-description :name "emotionalSein")]  
%%%  
requiredServices=[]  
%%%  
protocol emo_Agent_reaktiveEmotionsgenerierung=(request :language FIPA-SLO  
:content "danger")  
%%%  
protocol emo_Agent_reaktiveEmotionsgenerierung=(request :language FIPA-SLO  
:content "pleasant")  
%%%  
protocol emo_Agent_reaktiveEmotionsgenerierung=(request :language FIPA-SLO  
:content "unpleasant")  
%%%  
agentDesc=class de.renew.agent.repr.common.VTIndex  
%%%  
decisionComponent=emotionalDC  
%%%  
pleasant=joy  
%%%  
unpleasant=sad  
%%%
```

Abbildung 9.5.: Reaktive Emotionsgenerierung: Wissensbasis der emotionalen Agentenkomponente

9.2. Modell der reaktiven Emotionsgenerierung in E-MULAN

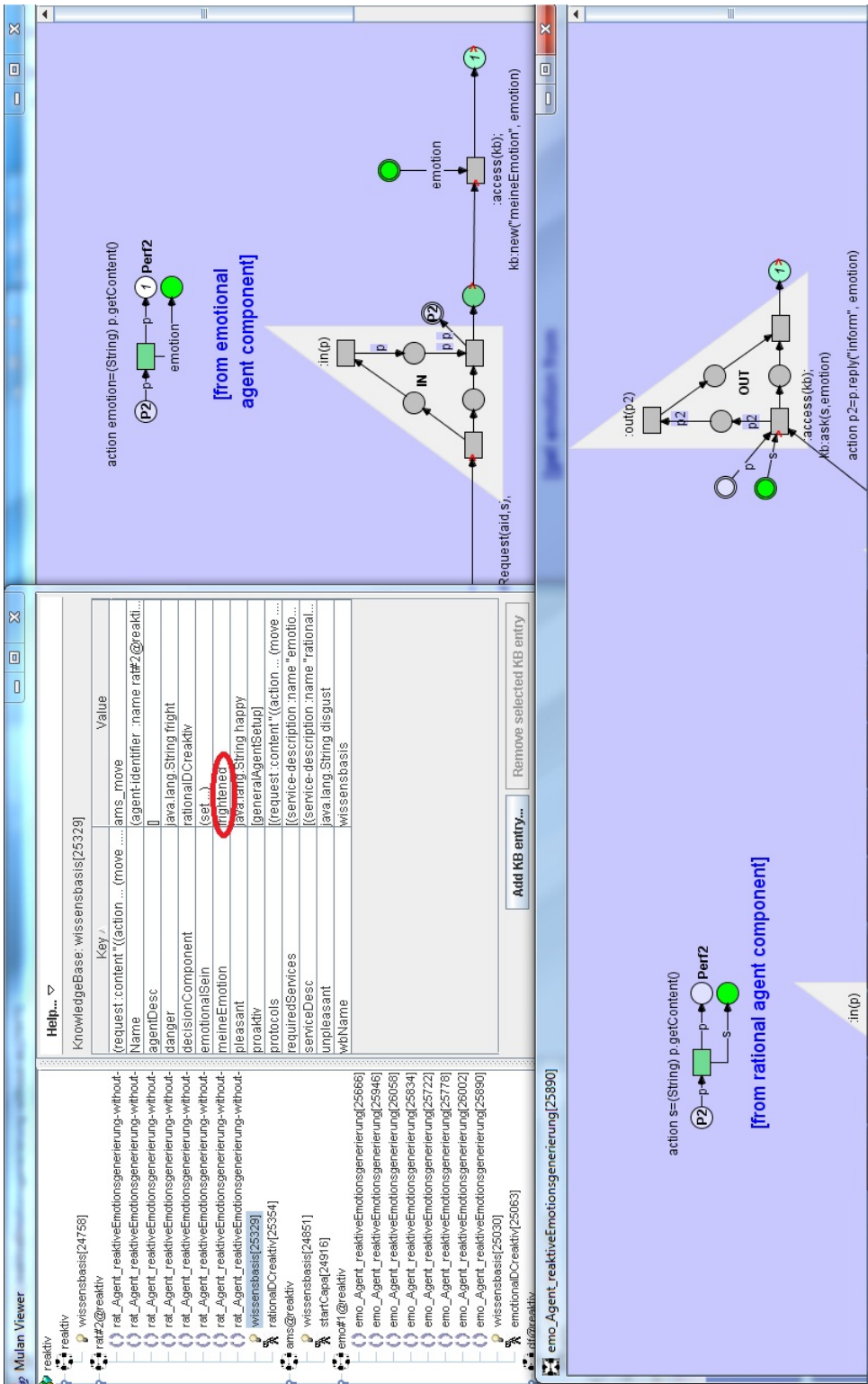


Abbildung 9.6.: Wissensbasis der rationalen Agentenkomponente in MULANVIEWER zur Laufzeit.

360

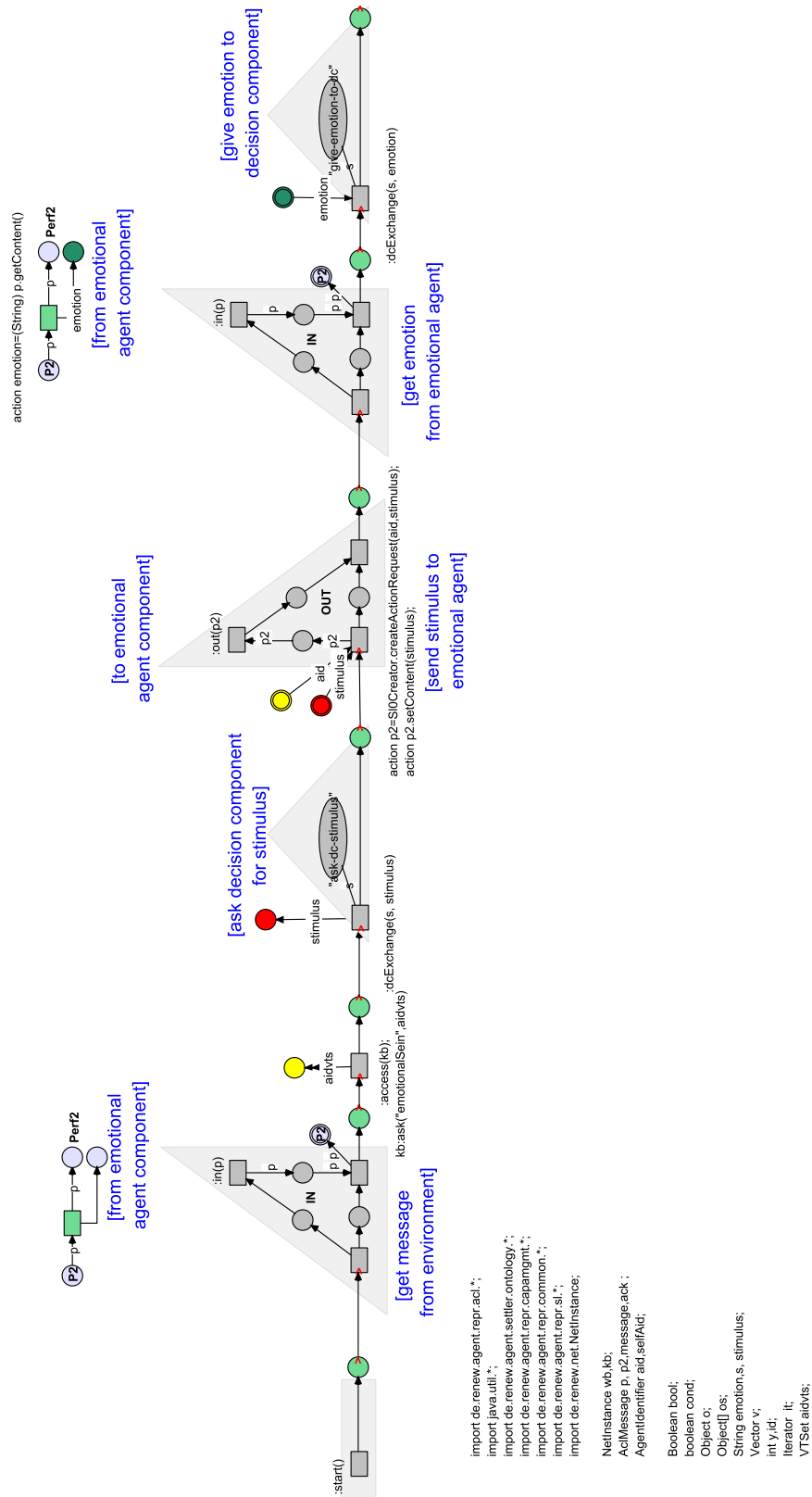


Abbildung 9.7.: Reaktive Emotionsgenerierung: Protokollnetz des rationalen Agenten mit Decision Komponente (DC)

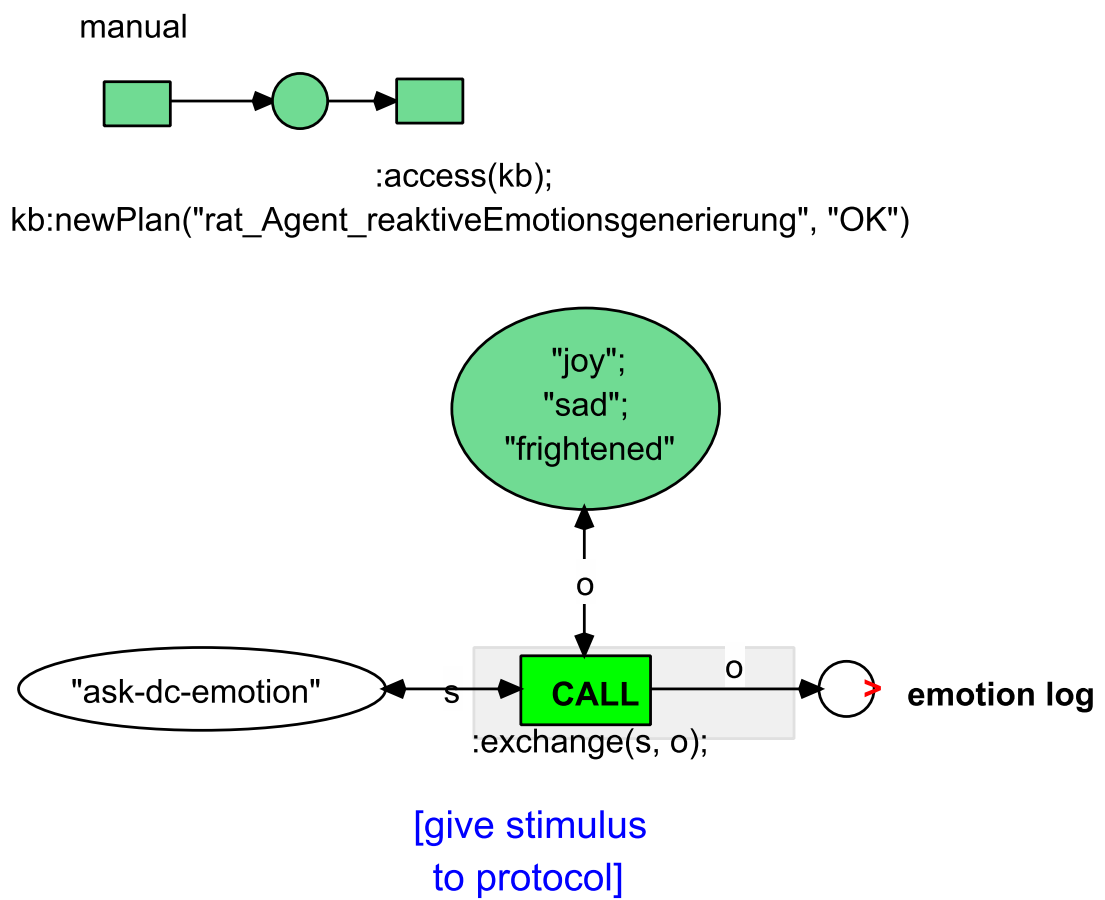


Abbildung 9.8.: Reaktive Emotionsgenerierung: Decision Komponente (DC) eines rationalen Agenten

9.3. Modell der proaktiven Emotionsgenerierung in E-Mulan

Wie im Falle der reaktiven Emotionsgenerierung werden bei der proaktiven Emotionsmodellierung wiederum rationale und emotionale Agentenkomponenten als zwei separate Agenten betrachtet, wobei alle emotionalen Prozesse eines Agenten in der emotionalen Agentenkomponente gekapselt sind. Dies erfolgt durch Auffaltung der E-MULAN-Agentenebene (vgl. Abb. 8.2). Es findet also wieder das aufgefaltete E-MULAN-Modell Anwendung.

Eine mögliche Realisierung der proaktiven Emotionsgenerierung in einem vereinfachten E-MULAN-Modell wurde bereits in Kapitel 8.3 diskutiert. Hier wird dieses Szenario als eine Interaktion zwischen den beiden Komponenten der E-MULAN-Agentenebene interpretiert: einer rationalen Komponente ("rationaler Agent") und einer emotionalen Shadow-Komponente ("emotionaler Agent").

Szenario: Proaktive Emotionsgenerierung Bei der proaktiven Emotionsgenerierung erfolgt eine emotionale Reaktion unabhängig von dem Input aus der Umgebung, sie wird stattdessen aufgrund der internen Verarbeitungsprozesse eines Agenten erzeugt. Die Interaktion wird von der emotionalen Shadow-Komponente des Agenten angestoßen (siehe Interaktionsdiagramm in Abb. 9.9). Die emotionale Agentenkomponente erzeugt proaktiv eine Emotion "e" und verschickt diese an die rationale Agentenkomponente. Die Emotion "e" wird von der rationalen Agentenkomponente empfangen und der Variable "meineEmotion" in der Wissensbasis der rationalen Agentenkomponente zugewiesen, die den emotionalen Zustand eines Agenten repräsentiert. Der Agent wird somit in einen emotionalen Zustand "e" versetzt.

Abhängig von dem Wert des emotionalen Zustandes könnten im nächsten Schritt wiederum die Verhaltensprotokolle definiert werden, die zu der aktuellen Emotion korrespondieren. Auf diese Weise kann die Steuerung des Agentenverhaltens durch die emotionale Shadow-Agentenkomponente des E-MULAN-Rahmenwerks exemplarisch realisiert werden (vgl. Kapitel 9.4).

Interaktionsdiagramm Die Kommunikation der rationalen und der emotionalen Agentenkomponenten wird für den Fall der proaktiven Emotionsgenerierung in Abb. 9.9 in einem Interaktionsdiagramm dargestellt.

Protokollnetze Die Protokollnetze zu diesem Szenario sind in Abb. 9.10 für die emotionale bzw. in Abb. 9.11 für die rationale Agentenkomponente dargestellt. Sie wurden mit Hilfe des AIP-Diagramm-Plugins aus den Interaktionsdiagrammen erzeugt und angepasst. Abb. 9.12 zeigt die beschriebene Beispielanwendung zur Laufzeit im MULANVIEWER.

9.3. Modell der proaktiven Emotionsgenerierung in E-MULAN

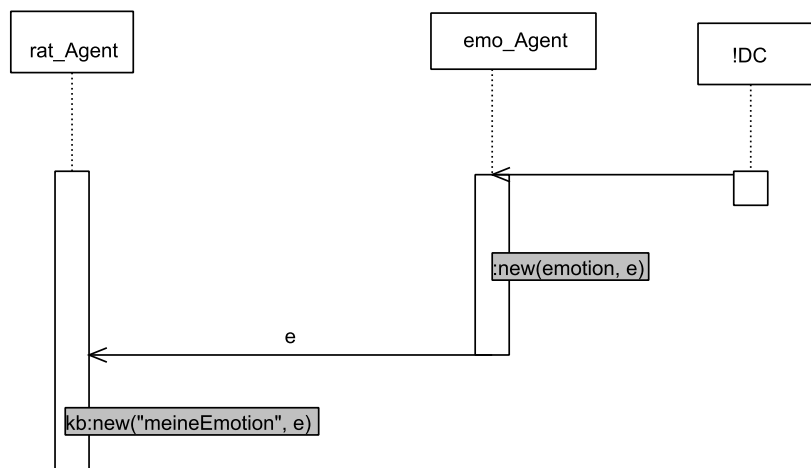


Abbildung 9.9.: Interaktionsdiagramm: Proaktive Emotionsgenerierung

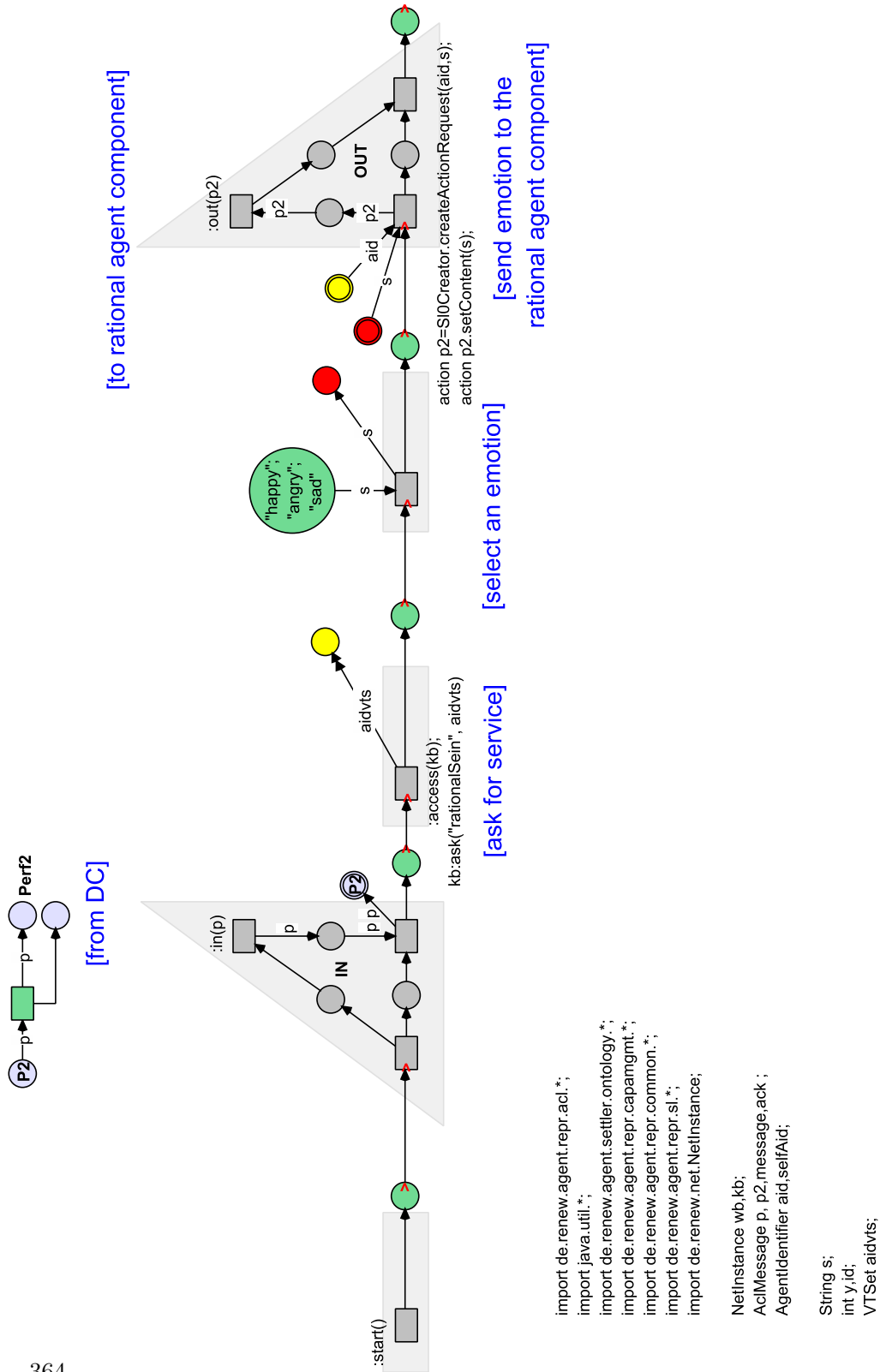


Abbildung 9.10.: Proaktive Emotionsgenerierung: Emotionaler Agent

9.3. Modell der proaktiven Emotionsgenerierung in E-MULAN

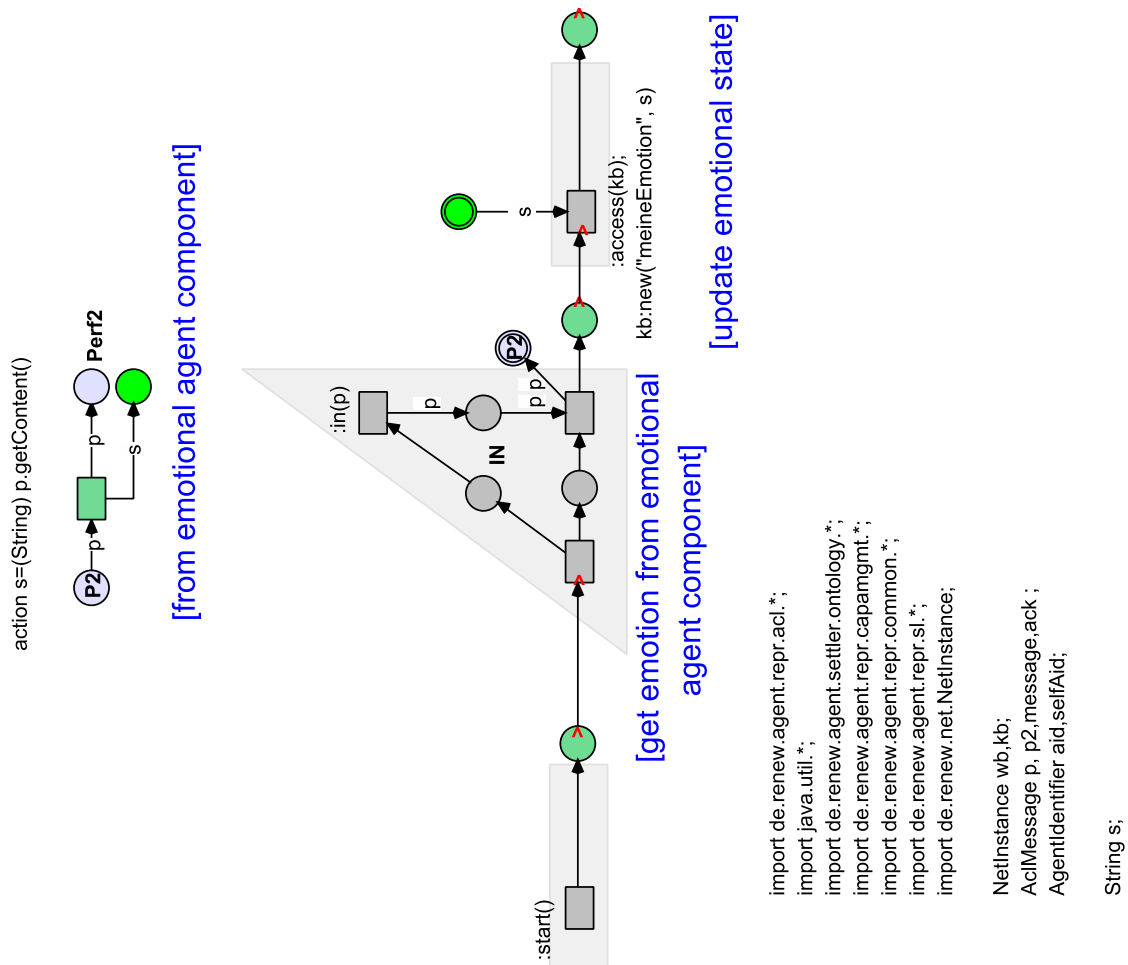


Abbildung 9.11.: Proaktive Emotionsgenerierung: Rationaler Agent

9. Anwendungsbeispiele



Abbildung 9.12.: Proaktive Emotionsgenerierung: Protokolle des emotionalen und des rationalen Agenten zur Laufzeit - Visualisierung im MULANVIEWER.

9.3. Modell der proaktiven Emotionsgenerierung in E-MULAN

Wissensbasen In Abb. 9.13 ist die Wissensbasis des emotionalen Agenten im wis-Format zu sehen. Die Wissensbasis des rationalen Agenten ist in Abb. 9.14 als wis-Datei dargestellt.

```
proaktiv=[generalAgentSetup]
%%%
requiredServices=[(service-description :name "rationalSein")]
%%%
protocol ams_move=(request :language FIPA-SLO
:content "((action (agent-identifier) (move)))")
%%%
agentDesc=class de.renew.agent.repr.common.VTIndex
%%%
decisionComponent=emotionalDC
```

Abbildung 9.13.: Proaktive Emotionsgenerierung: Wissensbasis eines emotionalen Agenten im wis-Format.

```
proaktiv=[generalAgentSetup]
%%%
serviceDesc=[(service-description :name "rationalSein")]
%%%
protocol RationalerAgent_proaktiveEmotionsgenerierung=(request :language FIPA-SLO
:content "happy")
%%%
protocol RationalerAgent_proaktiveEmotionsgenerierung=(request :language FIPA-SLO
:content "sad")
%%%
protocol RationalerAgent_proaktiveEmotionsgenerierung=(request :language FIPA-SLO
:content "angry")
%%%
agentDesc=class de.renew.agent.repr.common.VTIndex
```

Abbildung 9.14.: Proaktive Emotionsgenerierung: Wissensbasis eines rationalen Agenten im wis-Format.

9.4. Modell der emotionsbasierten Steuerung in E-Mulan

Einer

tion der Emotion innerhalb eines kognitiven Systems. Das Forschungsinteresse lag u.a. darin aufzuzeigen, wie die Funktionalität von Emotionen in informatischen Anwendungen, insbesondere in Agentensystemen, modelliert und eingesetzt werden kann. Dazu wurde in den Kapiteln 5 und 6 im Rahmen der Untersuchung der emotionstheoretischen Grundlagen speziell nach solchen Ansätzen gesucht, die die funktionale Rolle der Emotion in einem kognitiven System thematisieren (vgl. insbesondere Kapitel 5).

In Bezug auf die grundlegende Funktion einer Emotion ist es bei den meisten Emotionstheorien Konsens, dass diese in der Unterstützung der optimalen Anpassung an die Umweltsituation durch Steuerung von Verhalten besteht (vgl. exemplarisch [KH99], S. 468; siehe auch Kapitel 5.6). Durch Einnehmen eines emotionalen Zustandes wird der Organismus in die Lage versetzt, sich optimal an die aktuelle Umweltsituation anzupassen (vgl. z.B. [Laz91], [Dam03]). Der Entscheidungsprozess bzw. die Handlungswahl eines Akteurs werden in hohem Maße durch seine emotionalen Erfahrungen gesteuert (vgl. [BDD00], [FADT11], [FS87], [Sch88], [Laz91] bzw. Kapitel 6). Es wäre deshalb interessant, die Steuerungsfunktion der Emotion in einem Agenten nachzumodellieren, um den Einfluss der Emotion abzubilden, sie zu untersuchen und letztendlich zur Optimierung des Agentenverhaltens für bestimmte Anwendungen nutzen zu können (vgl. Kapitel 1.2).

In dem folgenden Anwendungsbeispiel wird gezeigt, wie der Einfluss der Emotionen auf das Agentenverhalten in E-MULAN modelliert werden kann. Die Steuerungsfunktion der emotionalen E-MULAN-Komponente auf die rationale Komponente wurde bereits in Kapitel 8.4.2 konzeptionell diskutiert. In Kapitel 8.2.4 wurden die unterschiedlichen Realisierungsmöglichkeiten beschrieben, insbesondere die emotionsbasierte Steuerung des rationalen Modells durch emotionale Steuerungsprotokolle. Um dies an einer konkreten Anwendung zu illustrieren, wird nun ein Modell der emotionsbasierten Steuerung in einer E-MULAN-Anwendung umgesetzt und simuliert. Hierzu wird das Szenario der emotionsbasierten Steuerung aus Kapitel 8.3.2 verwendet (siehe Kapitel 8.4.4).

Bei dem Anwendungsbeispiel soll primär nur die prinzipielle Realisierbarkeit einer Managementfunktion der Funktion auf Basis und mit Hilfe des E-MULAN-Rahmenwerks aufgezeigt werden. Die Modellierung der konkreten emotionstheoretischen Aussagen zur emotionsbasierten Steuerung des Agentenverhaltens, wie etwa die vielfältigen Einflüsse auf Aktionswahl, Planen, Entscheidungsfindung, Informationsverarbeitung usw., ist Aufgabe zukünftiger Forschungsarbeiten. Diese Einflüsse können grundsätzlich nur insoweit modelliert werden, wie sie in der zu modellierenden Emotionstheorie spezifiziert und in Form einer operationalisierbaren Beschreibung definiert werden. Eine Auswahl der für die Modellierung gut geeigneten theoretischen Ansätze wurde bei der Präsentation der emotionstheoretischen Grundlagen in Kapitel 6 vorgestellt, sie darf aber keinesfalls als vollständig angesehen werden.

Für die Untersuchung der Wechselwirkungen von emotionalen und rationalen Prozessen bietet das E-MULAN-Rahmenwerk eine geeignete Plattform, da es explizite Repräsentation sowohl rationaler / kognitiver als auch emotionaler Prozesse ermöglicht sowie insbesondere

eine explizite Darstellung der Wechselwirkungen zwischen diesen Prozessen unterstützt (vgl. Kapitel 8.1). Durch analytische Trennung in rationale und emotionale Modellkomponenten werden alle emotionalen Prozesse eines Agenten in einer emotionalen Agentenkomponente gekapselt, so dass eine explizite Darstellung und Untersuchung der emotionalen und rationalen Prozesse sowie ihrer Interaktionen direkt möglich wird.

Im Modell der emotionsbasierten Steuerung findet also wiederum das aufgefaltete E-MULAN-Agentenmodell Anwendung, bei dem die Agentenebene des E-MULAN-Rahmenwerks durch Auffaltung in seine beiden Bestandteile, den MULAN-Agenten und seine emotionale Shadow-Komponente, zerlegt wird. Diese beiden Komponenten werden als zwei separate Agenten betrachtet (vgl. Abb. 8.2). Der Einfluss der Emotion auf das Agentenverhalten kann über die Steuerung der rationalen Agentenkomponente durch die emotionale Agentenkomponente simuliert werden. So wird hier eine Möglichkeit aufgezeigt, die aus den Emotionstheorien bekannte Funktionalität von Emotion in natürlichen kognitiven Systemen in einem Agentensystem nachzumodellieren und zur direkten Steuerung des Agentenverhaltens zu nutzen.

Szenario: Emotionsbasierte Steuerung Die emotionsbasierte Steuerung wird also im Folgenden als eine Interaktion zwischen einer rationalen Komponente ("rationaler Agent") und einer emotionalen Shadow-Komponente ("emotionaler Agent") des E-MULAN-Rahmenwerks ausgelegt.

In der Wissensbasis der emotionalen Agentenkomponente sind Verhaltensstrategien gespeichert, die in Relation zu den vordefinierten Emotionen stehen. Diese Zuordnung kann einer geeigneten Emotionstheorie entnommen werden (z.B. der Theorie von Frijda, vgl. Kapitel 6.1.2 bzw. [Fri86], S. 88). In dem vorliegenden Beispiel verfügt die Relation "Emotion - Verhalten" jedoch über keine emotionstheoretische Grundlage. Um das Anwendungsmodell möglichst übersichtlich zu halten, wurden weitere Vereinfachungen angebracht: Zu den Emotionen "joy", "sad", "frightened" wurden die Verhaltensweisen "preserveBehavior", "UpdateBehavior" und "escape" willkürlich als optimale Verhaltensstrategien definiert. Die Emotionen, die von der emotionalen Komponente in Bezug auf das zu generierende Verhalten ausgewertet werden sollen, wurden in der rationalen Agentenkomponente einfachheitshalber als Marken an der Stelle **Emotions** dargestellt.

In einer elaborierteren Anwendung könnten natürlich die oben beschriebenen Modelle der reaktiven und proaktiven Emotionsgenerierung genutzt werden, um Emotionen dynamisch zur Laufzeit zu generieren. Ebenso könnten die Verhaltensstrategien in Abhängigkeit vom Modellgegenstand, der Anwendungsdomäne, dem Modellierungszweck und ggf. unter Berücksichtigung der emotionstheoretischen Fundierung der Emotion-Verhaltens-Relation gewählt werden. Es könnten sowohl konkrete Verhaltensweisen, als auch generelle Tendenzen - im Sinne der Handlungstendenzen nach [FS87] - angegeben werden. Um das Agentenverhalten für eine konkrete Anwendung abzubilden, wäre die Repräsentation der Verhaltensweisen durch die MULAN-Protokolle anstelle von Strings sicherlich zielführender. Auch wäre die Konstruktion und Implementierung eines vollständigen Anwendungsbeispiels, das den kompletten Emotionsprozess von der Stimulusbewertung bis zur Auswahl einer optimalen Verhaltensweise aufgrund der akkumulierten emotionalen Erfahrungen modelliert,

9. Anwendungsbeispiele

eine interessante Fragestellung für weiterführende Arbeiten. Davon muss aber an dieser Stelle abgesehen werden, da hier lediglich die Realisierbarkeit der Managementfunktion der Emotion in E-MULAN exemplarisch demonstriert werden soll.

Die rationale Agentenkomponente startet die Interaktion, indem sie an die emotionale Agentenkomponente eine Nachricht mit einer Emotion "e" sendet und nach der Verhaltensstrategie fragt, die in Bezug auf die Emotion "e" definiert ist. Die emotionale Agentenkomponente prüft, ob in der Wissensbasis zur Emotion "e" eine Verhaltensstrategie vorgegeben ist. Wenn es der Fall ist, wird das zur Emotion "e" in der Wissensbasis definierte Verhalten "b" ausgewählt, in eine Nachricht verpackt und an die rationale Agentenkomponente verschickt. Die rationale Agentenkomponente wird dann, je nach Inhalt der empfangenen Nachricht, entweder das aktuelle Verhalten entsprechend der von der emotionalen Komponente empfangenen Verhaltensstrategie anpassen oder das aktuelle Verhalten beibehalten, falls zu der ausgewerteten Emotion in der Wissensbasis der emotionalen Agentenkomponente kein Verhaltens-Update vorgeschrieben ist.

Interaktionsdiagramm Das eben beschriebene Szenario der emotionsbasierten Steuerung wird in Abb. 9.15 in einem Interaktionsdiagramm dargestellt.

Protokollnetze Die Protokollnetze der beiden Agentenkomponenten zum Interaktionsprotokoll aus Abb. 9.15 sind für die emotionale Agentenkomponente in Abb. 9.16 und für die rationale Agentenkomponente in Abb. 9.17 dargestellt.

Wissensbasen Die Wissensbasis eines emotionalen Agenten ist in Abb. 9.18 dargestellt. Darin findet man die Zuordnung der Emotionen zu den Verhaltensstrategien ("joy" = "preserveBehavior", "sad" = "UpdateBehavior", "frightened" = "escape"). Die Wissensbasis des rationalen Agenten in Abb. 9.19 entspricht weitestgehend der Wissensbasis eines rationalen Agenten in dem Anwendungsbeispiel zur reaktiven Emotionsgenerierung (vgl. Abb. 9.4).

In Abb. 9.20 wird die MULANVIEWER-Visualisierung des Modells der emotionsbasierten Steuerung zur Laufzeit gezeigt. In der Wissensbasis eines rationalen Agenten kann man die Änderung der aktuellen Verhaltensstrategie ("updateBehavior") aufgrund des aktuellen emotionalen Zustandes ("sad") beobachten.

9.4. Modell der emotionsbasierten Steuerung in E-MULAN

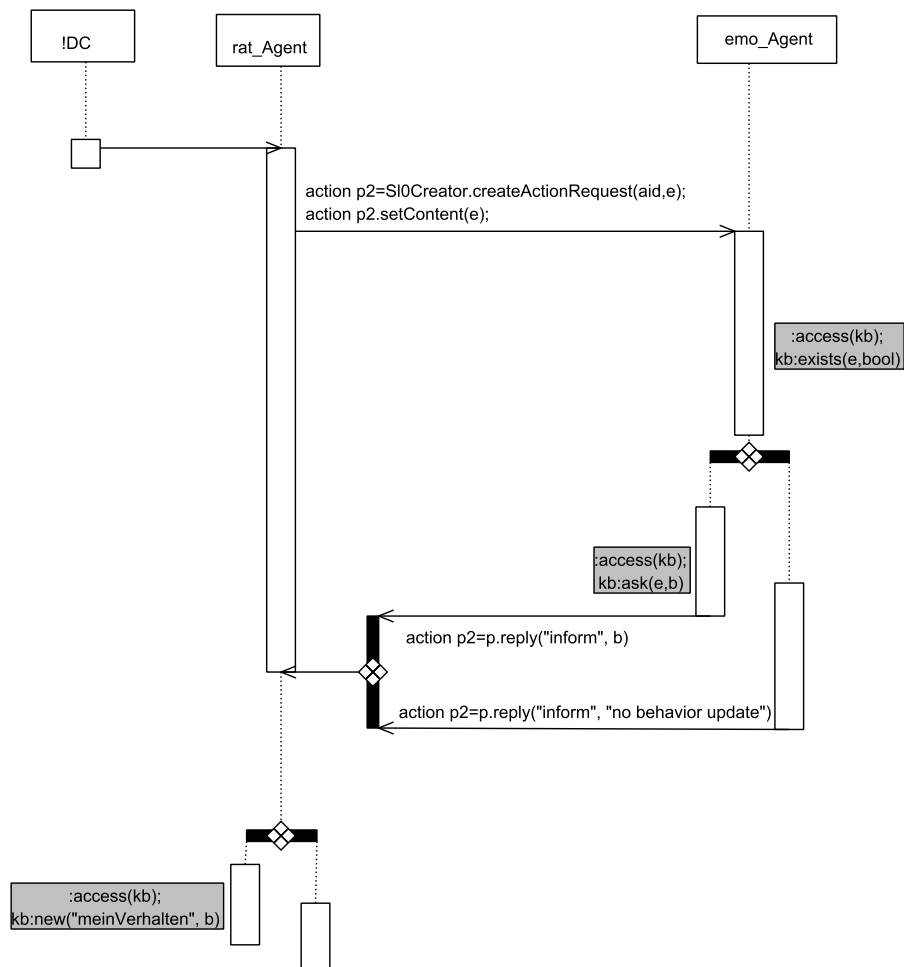


Abbildung 9.15.: Interaktionsdiagramm: Emotionsbasierte Steuerung

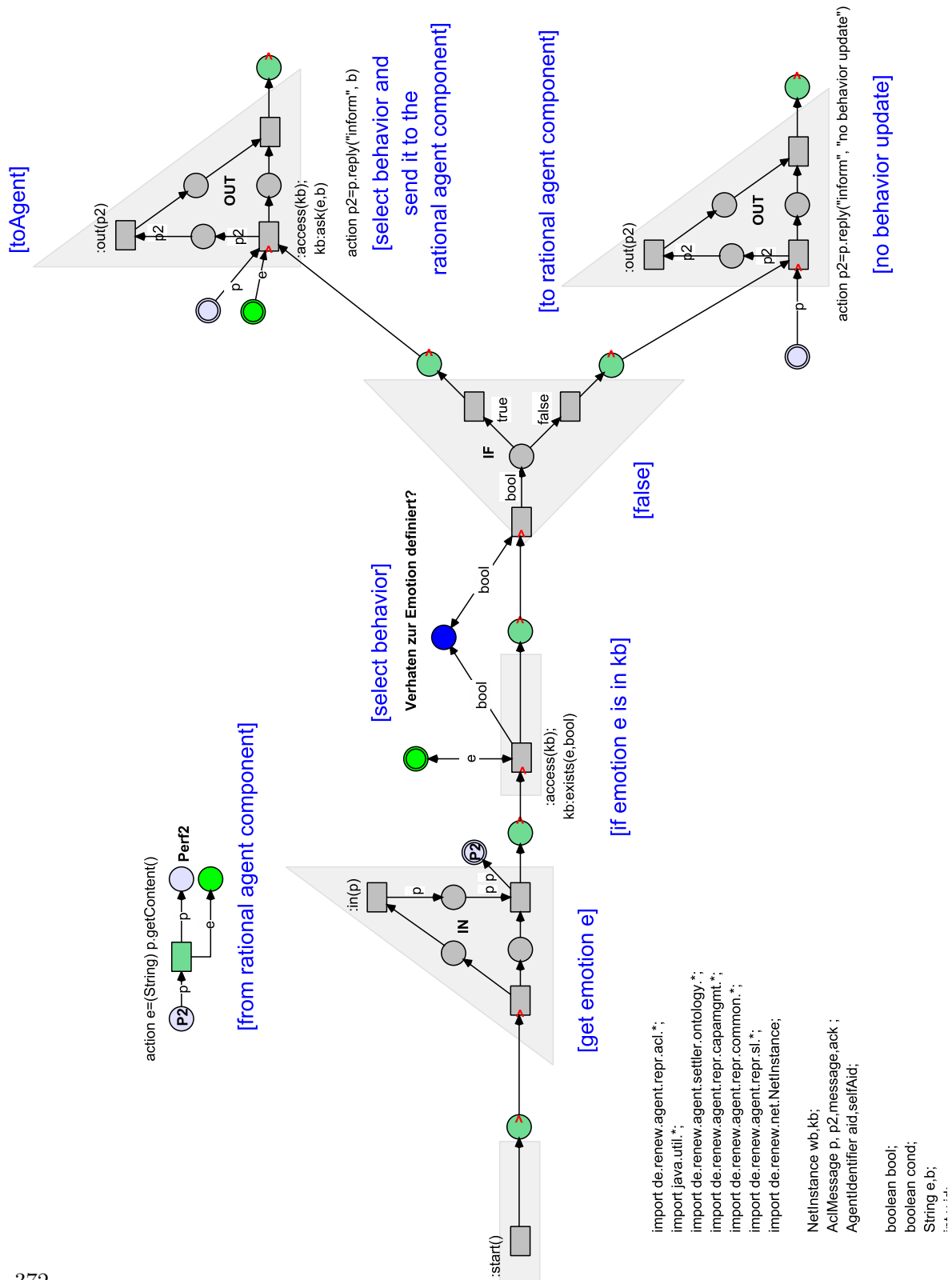


Abbildung 9.16.: Emotionsbasierte Steuerung: Emotionaler Agent

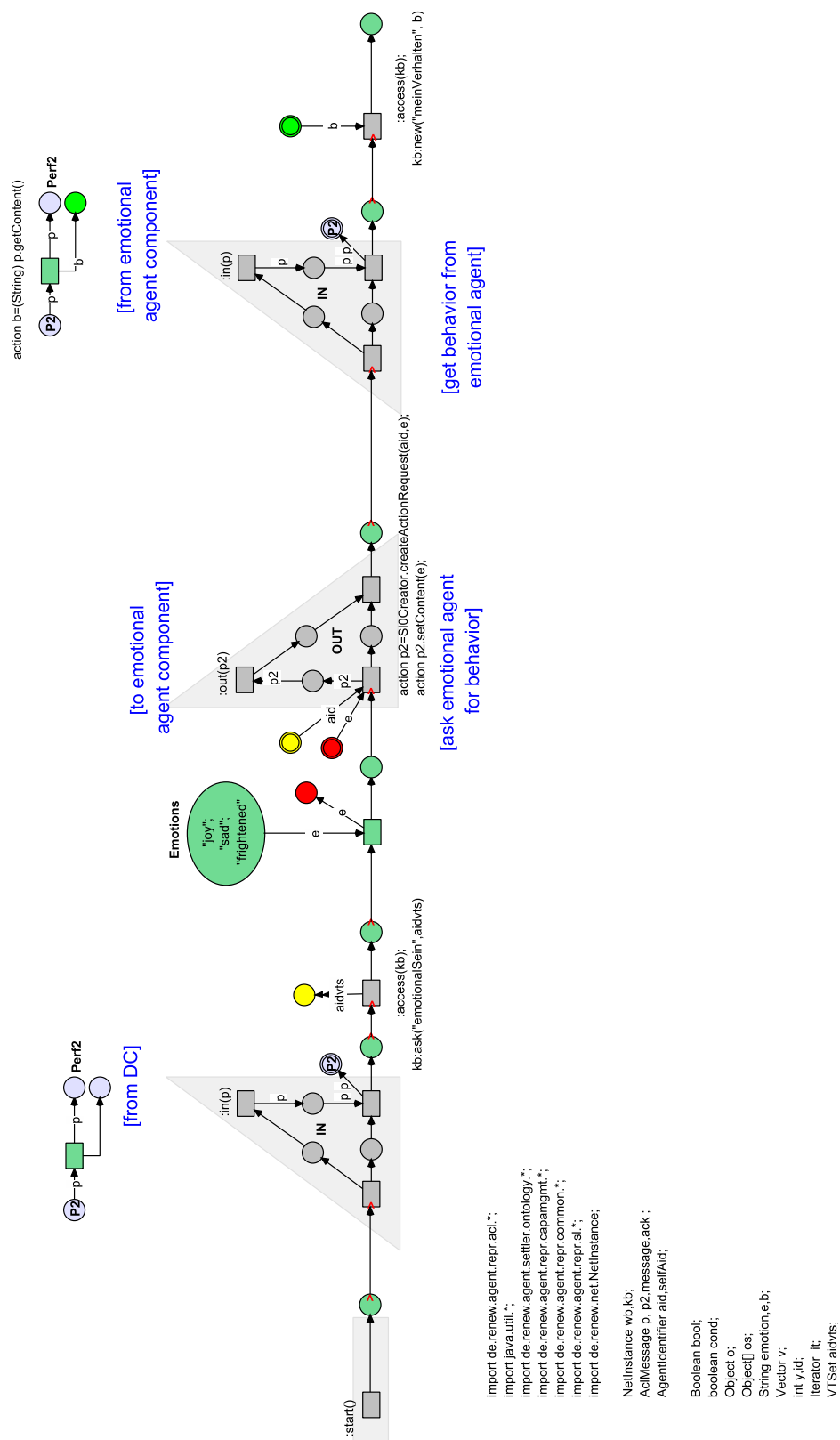
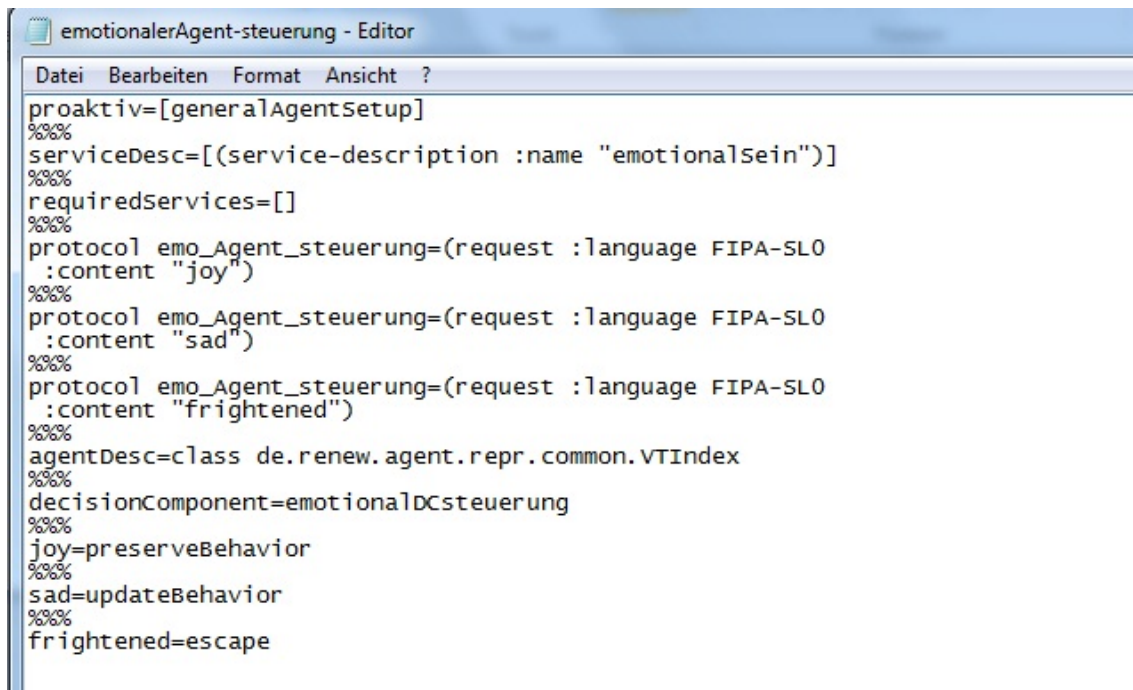


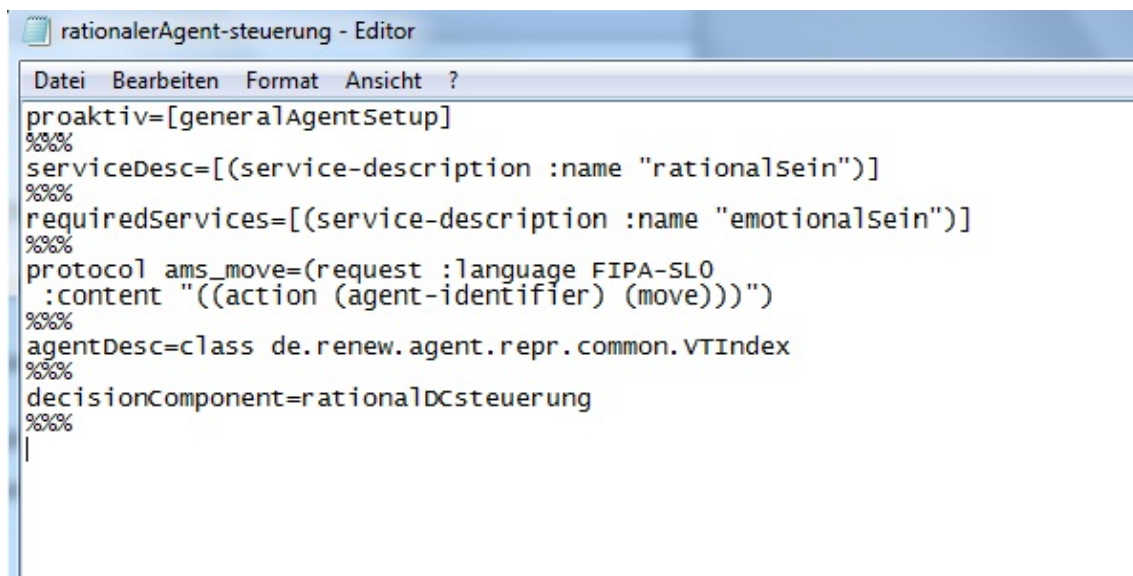
Abbildung 9.17.: Emotionsbasierte Steuerung: Rationaler Agent

9. Anwendungsbeispiele



```
emotionalerAgent-steuerung - Editor
Datei Bearbeiten Format Ansicht ?
proaktiv=[generalAgentSetup]
%%%
serviceDesc=[(service-description :name "emotionalsein")]
%%%
requiredServices=[]
%%%
protocol emo_Agent_steuerung=(request :language FIPA-SLO
:content "joy")
%%%
protocol emo_Agent_steuerung=(request :language FIPA-SLO
:content "sad")
%%%
protocol emo_Agent_steuerung=(request :language FIPA-SLO
:content "frightened")
%%%
agentDesc=class de.renew.agent.repr.common.VTIndex
%%%
decisionComponent=emotionalDCsteuerung
%%%
joy=preserveBehavior
%%%
sad=updateBehavior
%%%
frightened=escape
```

Abbildung 9.18.: Emotionsbasierte Steuerung: Wissensbasis eines emotionalen Agenten in wis-Format.



```
rationalerAgent-steuerung - Editor
Datei Bearbeiten Format Ansicht ?
proaktiv=[generalAgentSetup]
%%%
serviceDesc=[(service-description :name "rationalsein")]
%%%
requiredServices=[(service-description :name "emotionalsein")]
%%%
protocol ams_move=(request :language FIPA-SLO
:content "((action (agent-identifier) (move)))")
%%%
agentDesc=class de.renew.agent.repr.common.VTIndex
%%%
decisionComponent=rationalDCsteuerung
%%%
|
```

Abbildung 9.19.: Emotionsbasierte Steuerung: Wissensbasis eines rationalen Agenten in wis-Format.

9.4. Modell der emotionsbasierten Steuerung in E-MULAN

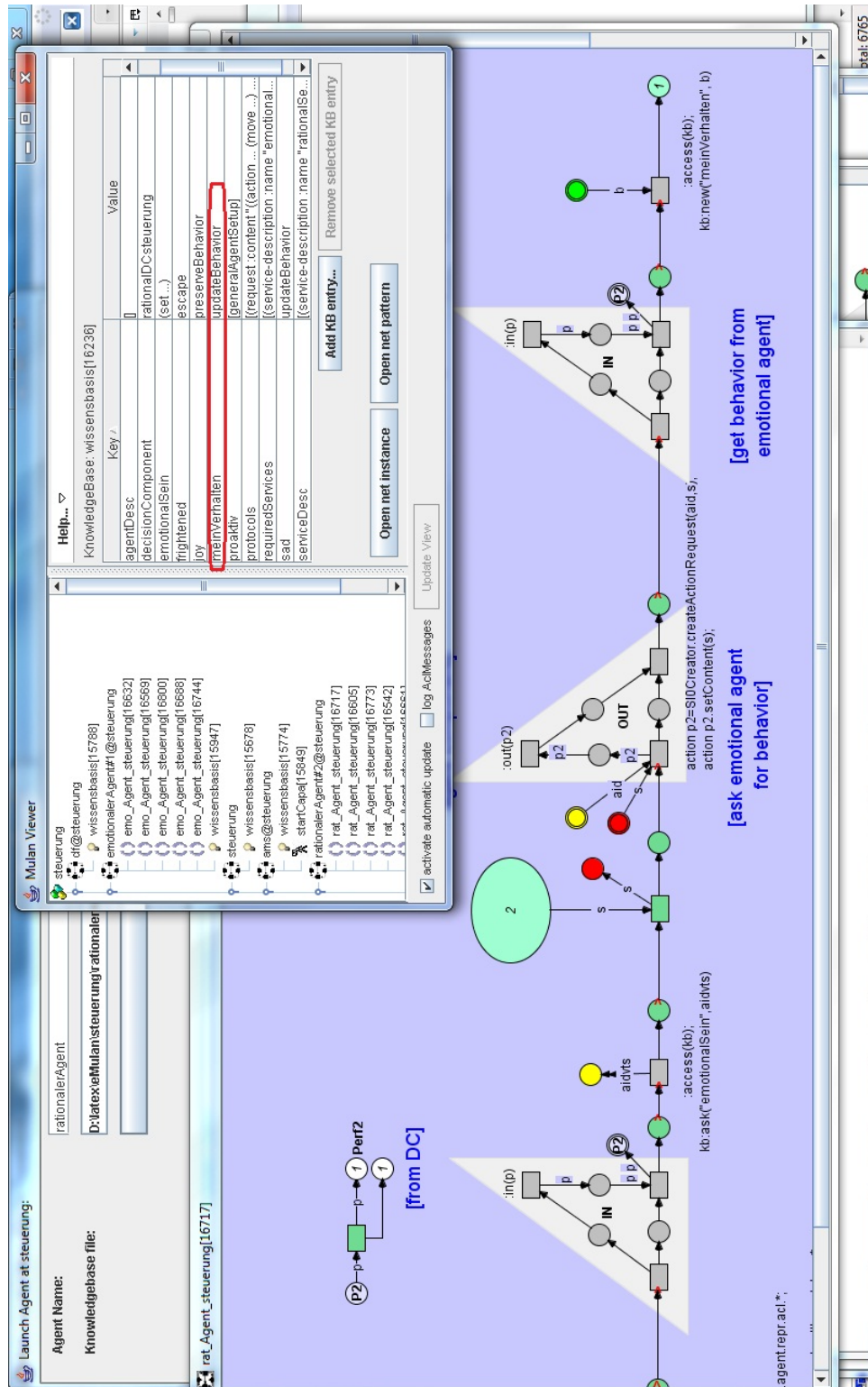


Abbildung 9.20.: Emotionsbasierte Steuerung: Wissensbasis eines rationalen Agenten zur Laufzeit.

9.5. Modell der emotionsbasierten Normetablierung

In Kapitel 6.3.3 wurde das Verhältnis von Emotion und sozialen Normen im Hinblick auf das Strukturationsproblem von Multiagentensystemen untersucht. Es sollte geklärt werden, wie normkonforme bzw. sozial akzeptierte Handlungsmuster bzw. Verhaltensweisen entstehen, warum sie den abweichenden Verhaltensweisen des Akteurs wiederholt vorgezogen werden und welche Rolle der Emotion dabei ggf. zukommt.

Aus der Betrachtung soziologischer und neuropsychologischer Ansätze zu den Wechselwirkungen von Emotion, Norm und normkonformen Verhalten konnte abgeleitet werden, dass bestimmte Emotionen als Instrumente der Normetablierung dienen, indem sie einerseits das Verhalten hin zur Normkonformität steuern und andererseits die Vermeidung der negativen Konsequenzen nicht konformer Handlungen bedingen.

Ausgehend von der Annahme, dass Normen das Handeln der Akteure restringieren, indem sie bestimmte Handlungsoptionen und -alternativen als inadäquat kennzeichnen, wurde mit Hilfe der Erkenntnisse aus [Dam03] gezeigt, dass mit negativen Emotionen assoziierte Handlungen durch die emotionsbasierte Situationskategorisierung unter bestimmten situationalen Gegebenheiten nicht ausgeführt werden, während die mit positiven Emotionen verknüpften Handlungsoptionen von den Akteuren vorgezogen werden. So kommen Regelmäßigkeiten im Handeln der Akteure zustande, die wir als soziale Normen bezeichnet haben.

Mit der Theorie von Elias [Eli76] konnten wir erläutern, wie diese unter Mitwirkung der sozialen Emotionen auf der Mikroebene fortdauernd (re-)internalisiert werden. Ausschlaggebend hierbei ist die Art der Sanktionierung im Falle einer Nichtbeachtung sozialer Normen. Abweichend von der ökonomischen Theorien ([Axe86], [Col91]) wird bei der Sanktionierung abweichenden Verhaltens nicht der Entzug materieller Ressourcen als ausschlaggebend angesehen, sondern vielmehr die emotionalen Reaktionen Verachtung, Geringschätzung, Ekel oder Abscheu, die beim Beobachten des abweichenden Verhaltens entstehen und den Akteur dann zu Sanktionen - materieller oder nicht materieller Art - zwingen, die in der Konsequenz den Verursacher Scham oder Schuld empfinden lassen.

Gerade die Emotionen Scham und Verachtung dienen als Vehikel für die Aufrechterhaltung sozialer Normen, indem sie nach [CC95b] "n-goals (normative goals)" generieren. Normkonformität als Ziel wird also vor allem durch die Furcht vor emotionalen Sanktionen (Verachtung) und der daraus resultierenden Scham gespeist. Zur Veranschaulichung, wie die sozialen Funktionen der Emotion ihre strukturdynamischen Effekte auf der Makroebene hervorbringen, wurde in Kapitel 6.3.3 deren Vermittlung und Wirksamkeit in der sozialen Interaktion betrachtet.

Konformität mit einer sozialen Norm entsteht demnach als Resultat der automatischen Vermeidung von Verhalten, das zu negativen sozialen Emotionen bei den anderen Gruppenmitgliedern führt. Die Demonstration negativer Emotionen (wie Verachtung, Abscheu, Geringschätzung, Ekel, Missbilligung) würde wiederum zur Generierung von negativen Emotionen - wie Scham, Schuld, Verlegenheit - bei dem Normbrecher führen, was es durch normkonformes Verhalten zu vermeiden gilt. Entsprechend führt emotionsgesteuerte Verstärkung (durch Anerkennung und Bewunderung des Beobachters) bei dem Akteur, der das normkonforme Verhalten demonstriert, zur Generierung positiver Emotionen wie Zu-

friedenheit und Stolz. Somit unterstützt die unwillkürliche Antizipierung der negativen / positiven sozialen Emotionen die Internalisierung der sozialen Norm.

Die Mechanismen der emotionsbasierten Sanktionierung des abweichenden bzw. der Verstärkung des normkonformen Verhaltens wurden in Kapitel 6.3.3 mit Referenznetzen veranschaulicht (vgl. Abb. 6.30 bzw. 6.31). Hierbei wurden Interaktionen zwischen einem Akteur, der sich entweder normabweichend oder normkonform verhält, und einem Beobachter, der dieses Verhalten emotional “bewertet“, modelliert. Viel besser und anschaulicher könnte ein solches Szenario natürlich in Form einer Interaktion zwischen zwei Agenten in einem E-MULAN-Referenzmodell dargestellt werden. In diesem Abschnitt werden deshalb die Ergebnisse aus Kapitel 6.3.3 in einem solchen Interaktionsszenario nachgebildet.

Bei der Modellierung nutzen wir die generische Eigenschaft des E-MULAN-Referenzmodells, das uns erlaubt, je nach Modellgegenstand entweder das aufgefaltete Agentenmodell mit der expliziten Trennung in emotionale und rationale Agentenkomponente oder das zusammengefaltete Modell, bei dem das rationale und das emotionale Modell in einem Agenten zusammengefasst sind, zu nutzen. Im Modell der emotionsbasierten Steuerung bietet sich die letztere Möglichkeit an, da eine Interaktion zwischen zwei verschiedenen Agenten eines Multiagenten-Settings dargestellt werden soll. Jeder dieser Agenten ist zur Generierung der Emotionen nach den in den Kapiteln 9.2 und 9.3 beschriebenen Mustern fähig. Das Agentenmodell kann also wiederum auf der E-MULAN-Agentenebene aufgefaltet werden, um die internen Mechanismen der Emotionsgenerierung bzw., falls im Anwendungsszenario erwünscht, auch die emotionsbasierte Steuerung (vgl. Kapitel 9.4), falls im Anwendungsszenario erwünscht, zu spezifizieren.

Szenario: Emotionsbasierte Normetablierung Im Folgenden stellen wir das in Kapitel 6.3.3 vorgestellte Modell der emotionsbasierten Normetablierung als eine Interaktion zwischen zwei Agenten dar. Es wird exemplarisch die Sanktionierung des norm-abweichenden Verhaltens (vgl. Kapitel 6.3.3) und die Verstärkung des norm-konformen Verhaltens (vgl. Kapitel 6.3.3) modelliert. Bei der Modellierung werden die Begriffe der Modelle aus Kapitel 6.3.3 beibehalten: Ein Agent, der durch sein Verhalten eine bestimmte soziale Verhaltensnorm bricht, wird als “Violator“ bezeichnet; den Agenten, der dieses Verhalten beobachtet und sanktioniert, nennen wir “Observer“.

In der Wissensbasis des Observer-Agenten sind Verhaltensweisen gespeichert, die im Widerspruch zur sozialen Norm stehen. Diesen Verhaltensweisen sind emotionale Reaktionen zugeordnet, die darauf abzielen, das abweichende Verhalten des Violator-Agenten zu bestrafen. Dies sind insbesondere die in Kapitel 6.3.3 identifizierten negativen sozialen Emotionen (Verachtung, Geringschätzung, Abscheu etc.), die als emotionale Sanktionen fungieren: Das bloße Kommunizieren dieser Emotionen an den Violator-Agenten stellt bereits eine Bestrafungsaktion dar, die ihn dazu bewegen kann, das nicht konforme Verhalten einzustellen. Das Kommunizieren der Emotion “Verachtung“ erfolgt hier vereinfacht über eine Nachricht an den Violator-Agenten. In einem elaborierteren Beispiel könnte es z.B. durch expressives Verhalten oder bestimmte Aktionen des Observer-Agenten erfolgen, die der Violator-Agent beobachten und entsprechend als Verachtung deuten könnte. Die Kommunikation bzw. Demonstration von Verachtung durch den Observer-Agenten führt zur Generierung von

9. Anwendungsbeispiele

negativen Emotionen - wie Scham, Schuld, Verlegenheit - aufseiten des Violator-Agenten. Dies hat zur Folge, dass der Violator-Agent sein norm-abweichendes Verhalten einstellt - dies kann z.B. durch den Mechanismus der emotionsbasierten Verhaltenssteuerung (siehe Kapitel 9.4) realisiert werden.

Ferner kann ein Lernmechanismus modelliert werden, über den ein Agent lernt, das mit negativen Emotionen verknüpfte abweichende Verhalten dauerhaft zu vermeiden und stattdessen das mit positiven Emotionen verknüpfte normkonforme Verhalten vorzuziehen - die sogenannte Internalisierung der sozialen Norm (vgl. Kapitel 6.3.3). So wäre es möglich, feste Verhaltensstrukturen, die sich auf der Agentenebene unter Mitwirkung der Emotionen Verachtung und Scham dynamisch ausbilden, auf der Ebene der sozialen Struktur als soziale Verhaltensnormen fortwirken zu lassen.

Durch die Konstruktion und eine anschließende Simulation eines solchen Anwendungsbeispiels ließen sich die Behauptungen der soziologischen und neurowissenschaftlichen Ansätze überprüfen, die die Schlüsselrolle der Emotion für die Ausbildung sozialer Normen in einer (Agenten-)Gesellschaft postulieren (z.B. [Dam03], [Els99], [Eli76], [SP00]). Dies wäre sicherlich zur Überprüfung der in Kapitel 6.3.3 aufgestellten Hypothesen interessant, würde aber den Rahmen dieser Arbeit sprengen. Hier soll zunächst einmal nur die prinzipielle Realisierbarkeit der emotionsbasierten Normetablierung im Rahmen des E-MULAN-Referenzmodells exemplarisch demonstriert werden.

Interaktionsdiagramm Das eben beschriebene Szenario der emotionsbasierten Sanktionierung des norm-abweichenden Verhaltens wird in Abb. 9.21 in einem Interaktionsdiagramm dargestellt.

Das Beobachten des Verhaltens eines Violator-Agenten durch den Observer-Agenten wird vereinfacht durch eine Nachricht des Violators an den Observer dargestellt. In dieser Nachricht teilt der Violator dem Observer sein aktuelles Verhalten mit. Falls das Verhalten von der sozialen Norm abweicht, wird der Observer die Emotion **Verachtung** generieren und diese dem Violator in Form einer Nachricht kommunizieren. Falls sich der Violator-Agent normkonform verhält, sendet der Observer-Agent die Emotion **Bewunderung**. Der Violator aktualisiert infolgedessen seinen emotionalen Zustand und setzt ihn im Falle der Geringschätzung durch den Observer auf **Scham** bzw. im Fall der Anerkennung seiner Normkonformität auf **Stolz**.

Protokollnetze Die Protokollnetze der beiden Agenten zum Interaktionsprotokoll aus Abb. 9.21 sind in Abb. 9.22 für den Violator und in Abb. 9.23 für den Observer dargestellt.

Wissensbasen Die Wissensbasis eines Violator-Agenten ist in Abb. 9.24 dargestellt. Darin findet man die Zuordnung der Emotionen des Violator-Agenten zu den Emotionen des Observer-Agenten ("Verachtung" = "Scham", "Bewunderung" = "Stolz"). Der aktuelle emotionale Zustand des Violator-Agenten wird "Scham" sein, weil vom Observer-Agenten die Emotion "Verachtung" als aktuelle Bewertung der Normkonformität des Verhaltens des Violator-Agenten kommuniziert wurde.

9.5. Modell der emotionsbasierten Normetablierung

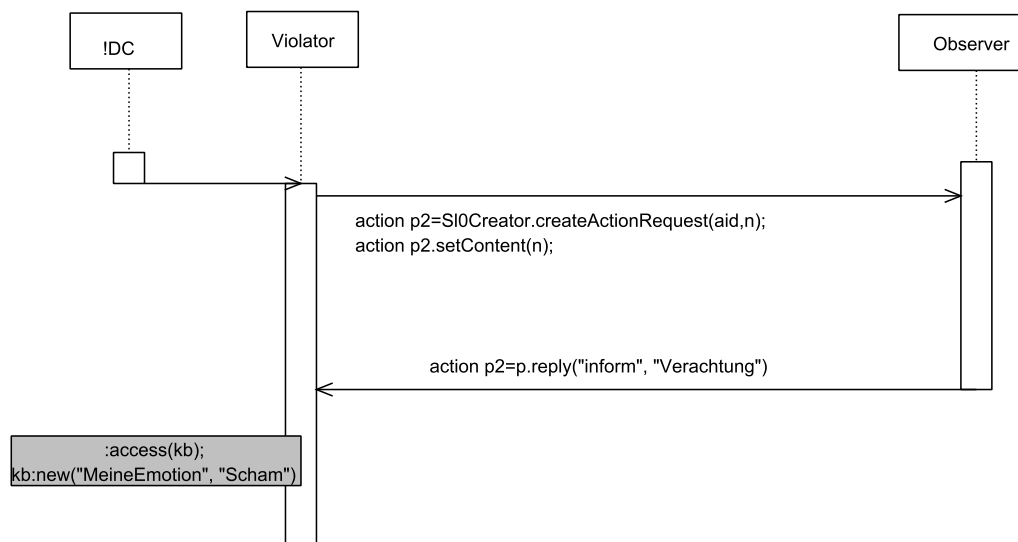
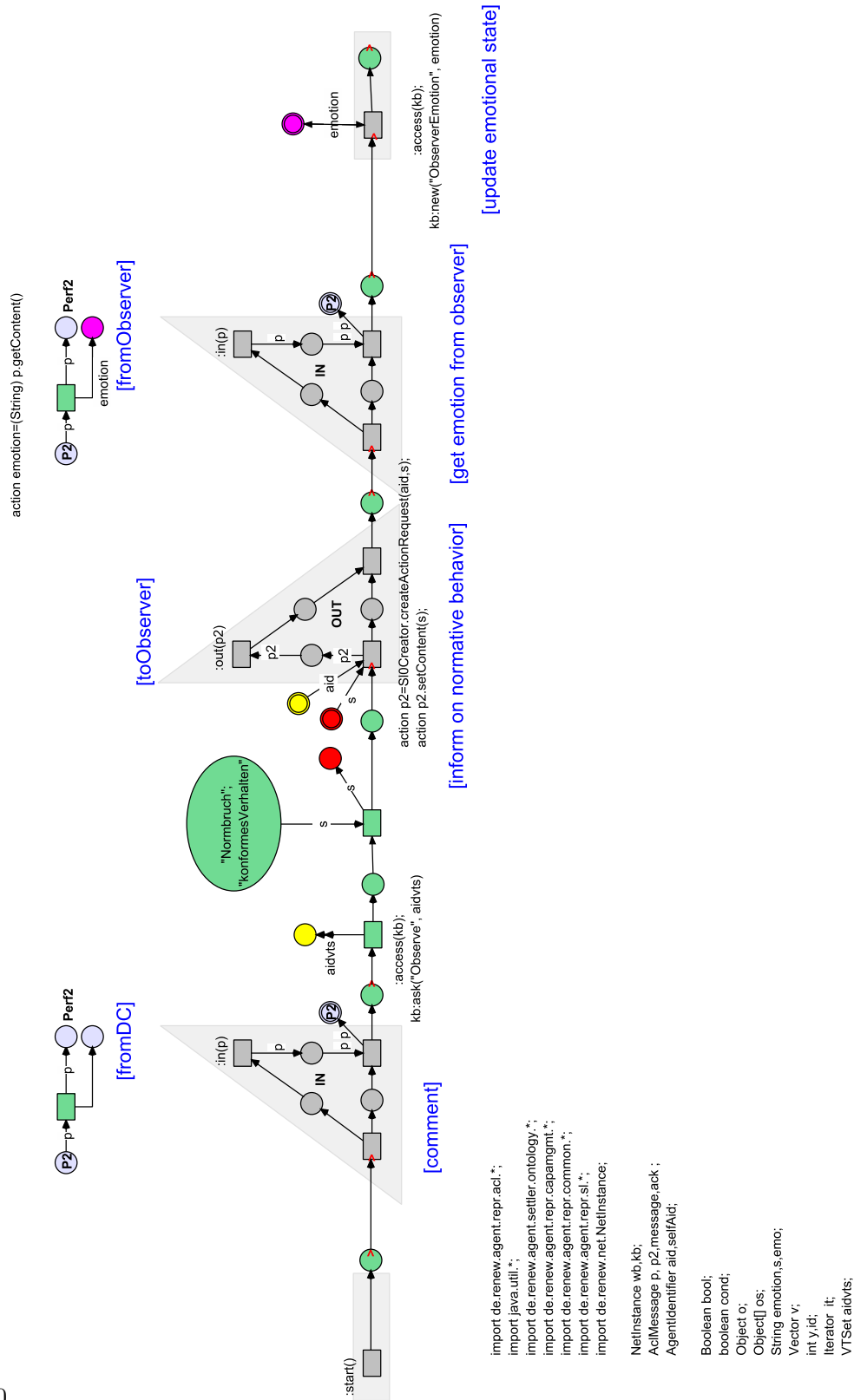


Abbildung 9.21.: Interaktionsdiagramm: Emotionsbasierte Normetablierung



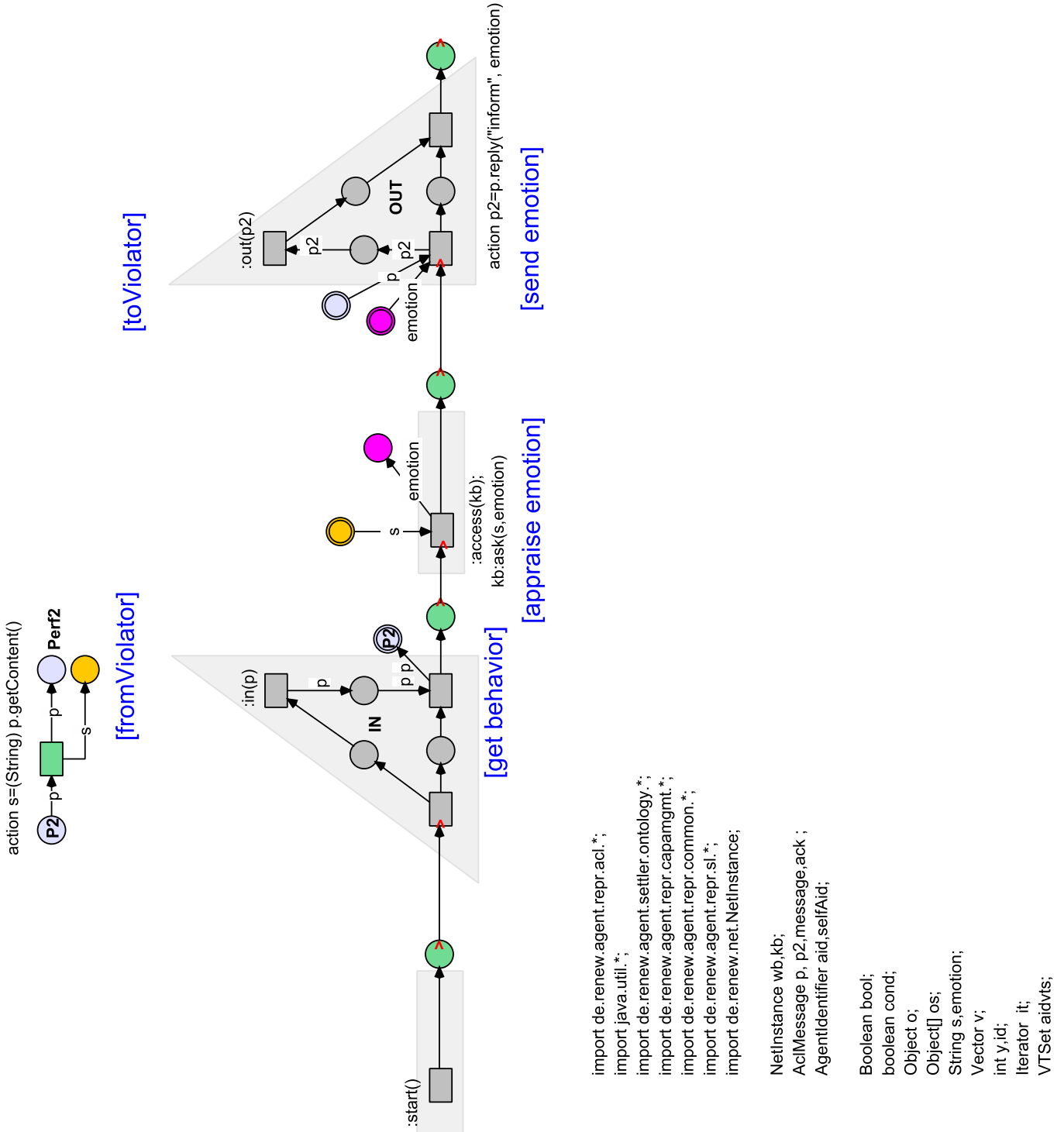


Abbildung 9.23.: Emotionsbasierte Normetablierung: Observer-Agent

9. Anwendungsbeispiele

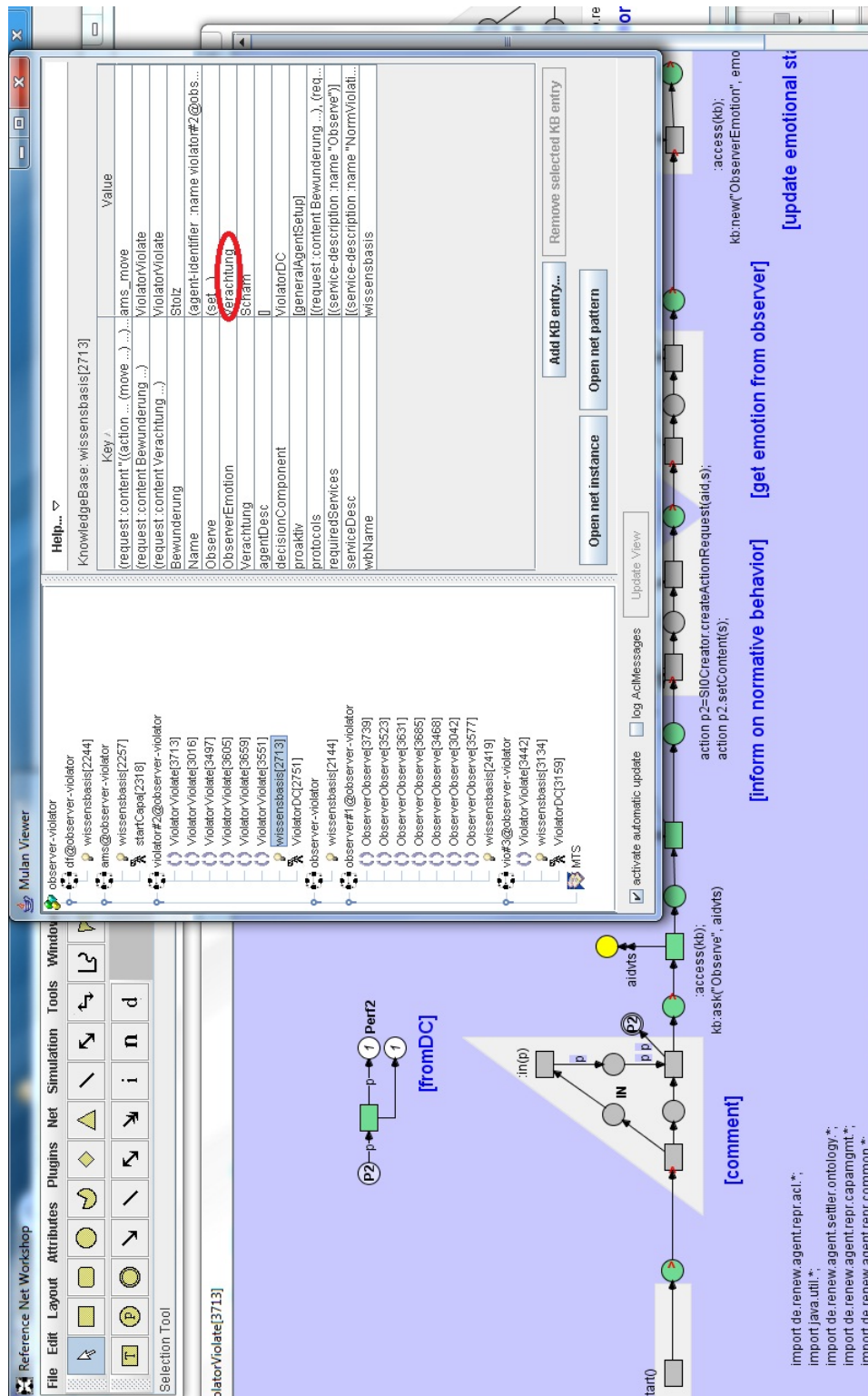


Abbildung 9.24.: Emotionsbasierte Normetablierung: Wissensbasis eines Violator-Agenten

10. Zusammenfassung und Ausblick

Die allgemeine Zielsetzung der vorliegenden Arbeit bestand in der Aufbereitung einer theoretischen Grundlage für Emotionsmodellierung in der Informatik durch Untersuchung und modellhafte Umsetzung von psychologischen, soziologischen und neurowissenschaftlichen Theorien der Emotion. Die Verwendung von Petrinetzen als Beschreibungstechnik für die Emotionstheorien und -modelle sollte im Laufe der Arbeit praktisch erprobt werden. Auf Basis der erstellten Modelle wurde ein gemeinsamer konzeptueller und terminologischer Rahmen zur Beschreibung der informatischen Emotionsarchitekturen erstellt.

Die Entwicklung von Agentenarchitekturen, die sich in ihrem Aufbau an den Mechanismen der menschlichen Kognition orientieren, erfordert als theoretische Basis sozialwissenschaftliche Theorien, die eine funktionale Beschreibung der genannten Systeme liefern und zugleich hinreichend detailliert und formalisierbar sind. Deshalb war eine möglichst umfassende Übersicht über sozialwissenschaftliche Emotionstheorien und aktuelle Ansätze in der informatischen Emotionsforschung ein wichtiges Anliegen dieser Arbeit. Durch die Bestandsaufnahme aktueller emotionstheoretischer Ansätze sollte die theoretische Grundlage für weitere Arbeiten zur Emotionsmodellierung in informatischen Anwendungen geschaffen werden. Die Übersichtsdarstellung der sozialwissenschaftlichen Emotionstheorien und Forschungsergebnisse und deren teilweise Formulierung mit informatischen Darstellungsmitteln sollte die Transferierbarkeit und Integration dieser Ergebnisse in die Informatikdomäne erleichtern.

Die Diskussion der Architekturen emotionaler Agentensysteme im Kapitel 7 hat eine wesentliche Forschungsfragmentierung und mangelnde theoretische Fundierung vieler Emotionsmodelle erkennen lassen. In diesem Zusammenhang stellte sich die Frage nach einem Referenzmodell für Emotionsmodellierung, das für die Analyse, Klassifizierung und den Vergleich verschiedener emotionaler Systemarchitekturen und Prozesse eingesetzt werden kann. Das Referenzmodell E-MULAN wurde auf Basis der referenznetz-basierten Multiagentenarchitektur MULAN entwickelt und so konzipiert, dass es als generisches Modell eine explizite Repräsentation sowohl rationaler / kognitiver als auch emotionaler Prozesse innerhalb desselben Rahmenwerks ermöglicht, sowie insbesondere die Wechselwirkungen zwischen diesen Prozessen darstellbar macht. Damit wurde einerseits eine Modellierungsbasis für die Darstellung emotionaler Prozesse für Agentenanwendungen geschaffen, andererseits aber auch ein Referenzmodell zur Modellierung von konkreten emotionalen Agentenarchitekturen. Das E-MULAN-Modell bildet somit einen gemeinsamen konzeptuellen und terminologischen Rahmen für die Darstellung der emotionstheoretischen Grundlagen und ihre Integration in die informatischen Emotionsmodelle.

Das Interesse der Arbeit wurde im Sinne der Grundlagenforschung möglichst breit angelegt: Es wurden unterschiedliche theoretische Überlegungen und praktische Modellierungsansätze aus verschiedenen Gebieten der Emotionsforschung und der informatischen Emotionsmodel-

10. Zusammenfassung und Ausblick

lierung zusammengetragen mit dem Ziel, diese in den informatischen Kontext systematisch einzuordnen und kritisch zu hinterfragen.

Die gewonnenen Ergebnisse sowie offene Fragen für weiterführende Arbeiten werden in diesem Kapitel zusammengefasst und diskutiert.

10.1. Diskussion der Ergebnisse

Den ersten Teil der Arbeit bildete die Einführung in die emotionstheoretischen Grundlagen und den aktuellen Forschungsstand der Emotionsmodellierung.

In einer Übersicht über aktuelle Forschungsschwerpunkte und Anwendungsgebiete für die Darstellung der Emotionen in der Informatik (Kapitel 2) wurden die informatischen Forschungsgebiete vorgestellt, die sich mit der Emotionsmodellierung befassen. Die wichtigsten Forschungsfragen, Motivationen und Anwendungsbereiche des jeweiligen Gebiets wurden kurz zusammengefasst. Es wurden drei grundsätzliche Motivationen für die Beschäftigung der Informatik mit der Emotionsmodellierung identifiziert: Emotionen werden einerseits als Mittel der technischen Systemoptimierung und Systemleistungssteigerung gesehen, andererseits dienen sie zur Verbesserung der Benutzbarkeit und schließlich werden sie zur Simulation und Validierung von sozialwissenschaftlichen Theorien in informatischen Systeme modelliert.

Die Motivation der funktionalen Systemoptimierung durch Emotion wurde für die vorliegende Arbeit als zentral identifiziert, weil es sich hierbei um die informatische Kernmotivation handelt, die Performanz, Zuverlässigkeit, Kontrollierbarkeit und Effizienz von Informatiksystemen durch die Nachbildung bzw. Einbeziehung natürlicher Emotionsprozesse in die bisher weitgehend rational geprägte Agentenmetapher zu steigern. Dieser Zielsetzung liegt die Hypothese zugrunde, dass soziale menschliche Konstellationen, wie sie in Psychologie, Soziologie oder Neurowissenschaft beschrieben werden, bei ihrer entsprechenden Umsetzung technische Systeme grundlegend verbessern können.

Daraus wurde eine der zentralen Forschungsfragen nach der Auswahl der psychologischen, soziologischen und neurowissenschaftlichen Theorien abgeleitet, die die funktionale Rolle der Emotion innerhalb eines isolierten bzw. sozialen Systems beschreiben. Diese Fragestellung wurde im weiteren Verlauf der Arbeit schwerpunktmäßig behandelt und bestimmte insbesondere die Auswahl der theoretischen Ansätze und informatischen Emotionsarchitekturen in den Kapiteln 6 und 7.

In weiteren Grundlagen-Kapiteln wurde eine kurze Einführung in Agenten- (Kapitel 3) und Multiagentensysteme (Kapitel 3.4) gegeben. Die benutzten Netzformalismen wurden vorgestellt und ihre Verwendung bei der sozionischen Theorienmodellierung erläutert (Kapitel 4). Schließlich wurden die Grundzüge des sozionischen Modellierungsansatzes nach [vLMV03]) dargestellt und die Anwendung dieses Ansatzes für die Modellierung der Emotionstheorien diskutiert. Es wurden zudem die verschiedenen netzbasierten Repräsentationstechniken und Vorgehensweisen vorgestellt. Als ein wichtiger Modellierungsgrundsatz wurde die stufenweise Verfeinerung der einfacheren Modellierungsansätze hin zu komplexeren Konstrukten identifiziert.

Der zweite Teil der Arbeit widmete sich der Aufbereitung der theoretischen Grundlagen der Emotionsforschung für die Informatik. In diesem Teil wurden sozialwissenschaftliche Emotionstheorien und informatische Emotionsmodelle mit Hilfe der Referenznetze nach [Val96] bzw. der vermittelnden Netze nach [vLMV03] konstruiert und analysiert. Das Ziel der Theorieübersicht bestand in erster Linie im Aufbau eines möglichst umfassenden Bildes des emotionstheoretischen Forschungsstandes in der für einen Informatiker leicht zugänglichen Beschreibungssprache der Petrinetz-Modelle.

Eingangs wurde aus Sicht der informatischen Emotionsmodellierung eine Reihe von Bewertungskriterien für die Beurteilung der sozialwissenschaftlichen Emotionstheorien abgeleitet, die bei der Theoriediskussion (Kapiteln 6.1, 6.2 und 6.3) Anwendung fanden. Im nächsten Schritt wurden die sozialwissenschaftlichen Emotionsdefinitionen ausgewählt und diskutiert und eine speziell im Kontext eines technischen Systems tragfähige operationale Definition des Emotionsbegriffs formuliert. Anschließend wurden ausgewählte emotionstheoretische Ansätze der Psychologie, Soziologie und Neurowissenschaften zusammengetragen und unter Verwendung der sozionischen Modellierungsverfahren nachgebildet.

Emotionsdefinition Die zentrale Zielsetzung in Kapitel 5 bestand darin, eine für den hier gewählten Schwerpunkt der funktionalen Systemoptimierung adäquate “funktionale Emotionsdefinition“ zu finden. Darunter wird eine funktionale Beschreibung der Emotion verstanden, welche die Spezifizierung der Emotionsentstehungsprozesse (Emotionsgenerierung) und der Auswirkungen der Emotionen auf das Gesamtverhalten des Systems berücksichtigt. Es wurde also vor allem nach ganzheitlichen Ansätzen gesucht, die Emotion als Reaktion auf Ereignisse im Kontext der Zielverfolgung betrachten und insbesondere die Implikationen, die Emotionen für kognitive Prozesse und Handlungssteuerung haben, zum Forschungsgegenstand machen. Die “funktionale Emotionsdefinition“ wurde durch ein Petrinetz-Modell veranschaulicht, das als Referenzmodell bei der Auswertung der Emotionsdefinitionen im Hinblick auf die Emotionsmodellierung in den informatischen Systemen dienen sollte. Anschließend wurden die verschiedenen Definitionen des Emotionsbegriffs betrachtet und eine Abgrenzung der Emotion zu anderen, in der Literatur oft gleichbedeutend verwendeten, jedoch bezüglich ihrer Funktion im natürlichen System sehr heterogenen Begriffen (Gefühl, Stimmung, Affekt, Reflex, Trieb) gesucht.

Bei der Analyse der Emotionsdefinitionen verschiedener Emotionstheoretiker musste festgestellt werden, dass die Definition des Forschungsgegenstandes ein grundlegendes Problem der zeitgenössischen Emotionsforschung darstellt. Unter dem Begriff *Emotion* werden sehr heterogene Prozesse und Zustände zusammengefasst und es besteht zurzeit kein Konsens bezüglich einer qualitativen Abgrenzung der Emotion zu Phänomenen wie z.B. Stimmungen, Affekten bzw. elementaren Affektzuständen.

Sowohl bei der kategorialen Klassifikation von Emotionen als *Basisemotionen* und *zusammengesetzte* Emotionen als auch bei *dimensionalen Modellen der Emotion* stehen Fragen nach der Emotionsgenerierung bzw. funktionalen Auswirkungen einer Emotion nicht im Zentrum des Forschungsinteresses. Deshalb sind diese beiden zentralen Kategorisierungsmöglichkeiten der Emotionspsychologie für die informatische Modellierung nicht sehr hilfreich, sofern die Entstehung bzw. Funktionalität der emotionalen Mechanismen innerhalb des

10. Zusammenfassung und Ausblick

Systems erfasst werden soll. Auch in den klassischen emotionspsychologischen Arbeitsdefinitionen werden die Prozesse der Emotionsentstehung und die Auswirkung der Emotion auf das Gesamtverhalten nicht explizit definiert. Für die Spezifikation der Emotionalität in einem Computersystem haben sich diese Definitionen als wenig geeignet herausgestellt. Daraufhin wurde auf die einschätzungstheoretischen und neurobiologischen Emotionsdefinitionen zurückgegriffen, die die Auswirkungen der emotionalen Prozesse auf das Verhalten des Gesamtsystems spezifizieren.

Die Emotionsdefinitionen werden im Hinblick auf die Umsetzbarkeit und Verwendbarkeit für die informatische Modellierung bewertet. Ausgewählte Emotionsdefinitionen werden mit Petrinetz-Modellen illustriert, einerseits um die konzeptuellen und terminologischen Zusammenhänge zu veranschaulichen, andererseits um verschiedene Definitionen zu vergleichen bzw. zueinander in Beziehung setzen zu können. Die Modellierung der betrachteten Emotionsdefinitionen mit Petrinetzen ließ gewisse strukturelle Ähnlichkeiten und grundsätzliche Anschlussmöglichkeiten verschiedener Definitionen erkennen.

So konnten die Netzmodelle der Emotionsdefinitionen von Lazarus, Frijda und Scherer zu einem allgemeinen Netz zusammengefasst werden, das den Prozess der Emotionsentstehung generell aus der Sicht der Einschätzungstheoretiker beschreibt.

Des Weiteren wurde bei den Netzmodellen der Basisemotionen-Definition und der einschätzungstheoretischen Emotionsdefinitionen die gleiche Grundstruktur aufgedeckt. Als gemeinsames strukturelles Merkmal der Emotionsgenerierung trat bei beiden Netzmodellen ein interner oder externer Stimulus auf, der als Auslösebedingung für den Emotionsprozess fungiert und entweder zur automatischen Emotionsgenerierung oder zur Auslösung des Einschätzungs- / Bewertungsprozesses führt. Damit wurde deutlich, dass in beiden Definitionen Emotion als eine *Reaktion* auf ein Ereignis dargestellt wird. Emotionen, die durch einen externen Stimulus verursacht werden, wurden deshalb als *reaktive Emotionen* und der zugehörige Generierungsprozess als *reaktive Emotionsgenerierung* bezeichnet. Es wurden zwei strukturell verschiedene Arten der reaktiven Emotionsgenerierung identifiziert: einerseits unmittelbare automatische Reaktionen auf bestimmte (kritische) Ereignisse in der Umgebung, die Basisemotionen (in der Literatur oft auch als primäre oder elementare Emotionen bezeichnet) auslösen; andererseits komplexe deliberative Prozesse der Einschätzung von Umweltereignissen in Bezug auf Ziele, Bedürfnisse und Überzeugungen (vgl. BDIs eines Agenten), die zur Generierung von Emotionen im Sinne der Einschätzungstheorien führen. Mittels (Platz-)Faltung der Netzmodelle der Basisemotionen (Kapitel 5.3) und der einschätzungstheoretischen Emotionsdefinition (Kapitel 5.9) wurde ein generelles Modell der reaktiven Emotionsgenerierung erzeugt, das die beiden Arten der Emotionsgenerierung in einem Netzmodell zusammenfasst. (Kapitel 5.11).

Daraus ergab sich ein generelles Emotionsmodell, das die Bedeutung des Begriffs “Emotion“ im Kontext der Emotionsmodellierung in technischen Systemen, insbesondere in KI-Systemen, spezifiziert (Kapitel 5.12). Der Vergleich dieses Modells mit der Definition des KI-Emotionsmodells nach [TWB07] ergab eine weitgehende Kongruenz der beiden Definitionen. Aus dem generellen Modell der Emotionsgenerierung wurde eine Emotionsdefinition abgeleitet, die zur Spezifikation des Emotionsbegriffs im Kontext der Modellierung der Emotionen in Agenten- und KI-Systemen für die weitere Arbeit zugrunde gelegt wurde.

Modellierung und Auswertung der psychologischen Emotionstheorien Die Analyse der psychologischen Theorien hat gezeigt, dass es bisher in der Psychologie keine *integrative Theorie* der Emotion im Sinne von [ZS00] gibt, die dem Multikomponenten-Charakter des Phänomens Rechnung tragen würde und insbesondere konkrete Vorhersagen über Auslösung und Differenzierung von Emotionsprozessen und die hierbei auftretenden Reaktionsmuster in den verschiedenen (System-)Komponenten machen könnte.

Die meisten psychologischen Emotionstheorien konzentrieren sich auf eine der Konstituenten des Emotionskonstrukts. So sind bereits die klassischen Emotionstheorien wie die James-Lange-Theorie (1894), die Cannon-Bard-Theorie (1929) und die Theorie von Schachter und Singer (1964) lediglich als Teiltheorien anzusehen, da sie sich im Wesentlichen auf subjektive und physiologische Aspekte des Emotionsphänomens fokussieren und nur innerhalb dieses Bereichs ihren Erklärungsanspruch sehen.

Nach einem Überblick über psychologische Emotionstheorien wurden Einschätzungstheorien zur weiteren Spezifizierung ausgewählt. Als Vorteile der Einschätzungstheorien in Bezug auf eine Modellierung der Emotionen zur technischen Systemoptimierung wurden insbesondere ihre leichte Operationalisierbarkeit, ihre funktional orientierte Betrachtung der Emotion sowie ihre gute Übertragbarkeit auf die Agentenmetapher, vor allem auf die grundlegenden Konzepte der BDI-Agentenarchitektur, identifiziert. Bei der Analyse der emotionspsychologischen Theorien wurde keine andere Theorieklasse gefunden, welche die aktuellen empirischen Befunde und die emotionspsychologischen Fakten umfassender zu erklären vermag. In den informatischen Emotionsmodellen werden die psychologischen Einschätzungstheorien deshalb am häufigsten implementiert. Als wichtige Nachteile der Einschätzungstheorien wurden ihre schwierige empirische Überprüfung, die nicht gesicherten Einschätzungsdimensionen sowie die vielerorts mangelhafte Spezifikation der Mechanismen des Einschätzungsprozesses identifiziert.

Die Analyse und Modellierung der Emotionstheorien von Ortony, Clore und Collins [OCC88], Frijda [FS87], Lazarus [Laz91] und Scherer [Sch93] hat gezeigt, dass die meisten Einschätzungstheorien den Emotionsgenerierungsprozess zwar sehr genau spezifizieren, jedoch keine Aussagen über die Auswirkungen der generierten Emotion (z.B. für Entscheiden, Schlussfolgern, Planen, Handlungswahl, Motivation usw.) machen. So werden z.B. in der OCC-Theorie lediglich der Prozess der kognitiven Einschätzung der Ereignisse / Akteure / Objekte in Bezug auf bestimmte Dimensionen spezifiziert. Damit wird zwar die Emotionsgenerierung von der Theorie hinreichend detailliert beschrieben, jedoch nicht die Funktion der generierten Emotionen für den weiteren Verlauf der Systemprozesse, wie etwa die Wahrnehmung, die Informationsverarbeitung oder die Handlungswahl. Will man die OCC-Theorie in einem Agentensystem modellieren, muss von den Modellentwicklern zuerst die Relation der Emotionen und ihrer Auswirkungen im Gesamtsystem entworfen werden. Bei vielen Umsetzungen der OCC-Theorie ([Ell92], [Rei96]) werden an dieser Stelle intuitive Konstrukte der Entwickler eingesetzt, um die fehlende Spezifizierung der Auswirkung von Emotionen auf das Agentenverhalten zu ersetzen.

Auf der allgemeinen Ebene hat die Petrinetz-Modellierung der Emotionstheorie von Ortony et al. [OCC88] eine Reihe von Mehrdeutigkeiten aufgedeckt, die in den Formulierungen der Theorie bezüglich der Differenzierung von drei Emotionsklassen enthalten sind. So liefert

10. Zusammenfassung und Ausblick

die OCC-Theorie keine eindeutige Aussage darüber, ob es sich bei den Emotionsklassen um disjunkte Klassen handelt. Ebenso wird aus der OCC-Theorie nicht klar, wie sich Emotionen verschiedener Klassen (z.B. ereignis- vs. akteursfundierte Emotionen) zueinander verhalten und ob in diesem Fall ein gegenseitiger Ausschluss oder eine nebenläufige Emotionsgenerierung zu modellieren ist. Des Weiteren macht die OCC-Theorie keine Angaben über die Dauer respektive den zeitlichen Verlauf der erzeugten Emotionen.

Die Darstellung der Theorienaussagen mit Netzmodellen ließ anhand der Modellstruktur die in der OCC-Theorie enthaltene Mehrdeutigkeit bezüglich der Differenzierung der Emotionsklassen, des kausalen Ablaufs der Einschätzungsprozesse und der generierten Emotionen erkennen. Dabei kamen die Vorteile der Benutzung von Petrinetzen als Modellierungstechnik für sozialwissenschaftliche Theorien deutlich zum Vorschein.

Von den betrachteten Einschätzungstheorien wurde lediglich in der Theorie von Frijda [FS87], in der speziell die funktionale Rolle der Emotionen bei der Handlungsgenerierung fokussiert wird, der gesamte emotionale Prozess beschrieben. In den Theorien von Ortony et al. [OCC88], Lazarus [Laz91] und Scherer [Sch93] wird an die Einschätzung sich anschließende Anpassungsaktivitäten nicht eingegangen, was offensichtlich als Nachteil dieser Theorien in Bezug auf die ganzheitliche Emotionsmodellierung zu erachten ist. Generell konnte festgestellt werden, dass die Emotionstheorie von Frijda [FS87] aus der Perspektive der funktionalen Optimierung von Agentensystemen ein deutlich größeres Erklärungspotenzial besitzt als die anderen betrachteten Theorien.

Es hat sich gezeigt, dass jede der betrachteten Einschätzungstheorien ihre speziellen Schwerpunkte, aber auch Schwachstellen in Bezug auf die Erklärung und Modellierung des Emotionsphänomens aus informatischer Sicht besitzt. Für die Emotionsmodellierung bedeutet dies, dass die Emotionstheorien ggf. miteinander kombiniert und gegenseitig ausgeglichen werden müssen. Wie in Kapitel 7 gezeigt wurde, ist dies in den meisten informatischen Emotionsmodellen auch der Fall. Die Untersuchung der Validität der Emotionsmodelle, die durch die Kombination mehrerer emotionstheoretischer Ansätze zustande kommen, ist eine offene Forschungsfrage für weiterführende Arbeiten.

Von den betrachteten psychologischen Emotionstheorien berücksichtigen lediglich die von Frijda [FS87] und Scherer [Sch93] explizit die Rolle der sozialen Phänomene auf verschiedenen Stufen des Emotionsprozesses. Insbesondere geht Frijda auf die Rolle der sozialen Normen bei der Generierung und Regulation von Emotionen ein.

Die grundlegende Frage an die psychologische Emotionsforschung, wie die wechselseitige Korrelation zwischen sozialen Phänomenen und der Emotion auf der psychologischen Ebene beschaffen ist, bleibt jedoch offen. Die betrachteten Ansätze haben die These von Collins ([Col84]) bzw. Barbalet ([Bar99]) bestätigt, dass die meisten emotionspsychologischen Forscher "gegenüber soziologischen Befunden weitgehend immun geblieben sind - sie schließen nur wenige und allenfalls eher dekorativ soziologische Beiträge ein" ([Fla03], S. 118).

Modellierung und Auswertung der neurowissenschaftlichen Ansätze Da die experimentellen Beweise einer Beteiligung der Emotionen an kognitiven Denk- und Entscheidungsprozessen in den letzten Jahren vor allem aus der neurowissenschaftlichen Forschung kommen,

wurde in Kapitel 6.2 eine Übersicht zu den wichtigsten Ansätzen der neurowissenschaftlichen Forschung zur Emotion vorgestellt.

Nach der Vorstellung der Somatic-Marker-Hypothese wurde ihr großer Einfluss auf die KI-Emotionsforschung erläutert, die in den Arbeiten von Damasio ein wichtiges Argument dafür gefunden hat, dass der Zugang zum künstlichen Intellekt nicht allein durch Nachbilden der kognitiven Mechanismen unter Vernachlässigung der emotionalen möglich ist.

Die Mechanismen des emotionalen und kognitiven Prozessierens nach LeDoux [LeD96b] wurden ebenso vorgestellt und mit einem Petrinetz modelliert. Die Untersuchungen von LeDoux ([LeD84], [LeD96a], [LeD96b], [LeD98]) liefern die neurophysiologische Begründung der grundlegenden Dichotomie zwischen Emotion und Kognition und geben erste Hinweise darauf, dass emotionale und kognitive Verarbeitungsprozesse nach der Wahrnehmung eines Stimulus parallel ablaufen und sich gegenseitig ergänzen. Diese Erkenntnisse wurden bei der Entwicklung der Referenzarchitektur zur Emotionsmodellierung (Kapitel 8) genutzt, um die konzeptuelle Trennung in emotionale und rationale Architekturkomponenten zu begründen.

Die wichtige Frage nach der Steuerung des sozialen Verhaltens und insbesondere nach der Verankerung der sozialen Gesellschaftsstrukturen der Makroebene in einem Akteur bzw. Agenten, d.h. auf der Mikroebene, wurde eingangs in Hinblick auf die sozionalen Probleme der Strukturierung und -regulation der Multiagentengesellschaften für den theoretischen Teil dieser Arbeit formuliert (vgl. Kapitel 1).

Die Analyse der psychologischen Ansätze in Kapitel 6.1 hat gezeigt, dass insbesondere die Einschätzungstheorien ihren Erklärungsanspruch auf interne psychologische Prozesse beschränken, ohne jedoch soziale Determinanten und Funktionen der Emotionen zu berücksichtigen. Es wurde deutlich, dass mit Hilfe der Einschätzungstheorien lediglich die Generierung der Emotionen unter Einfluss der Umweltfaktoren spezifiziert werden kann, nicht jedoch der Einfluss der Emotion auf Strukturen der Makroebene wie soziale Normen. Da die Frage nach dem Zusammenhang von Emotion und sozialer Norm von der emotionspsychologischen Forschung nicht ausreichend beleuchtet wurde, wurde nun aus der neurowissenschaftlichen Perspektive untersucht, welche neurologischen Mechanismen für die Verfestigung der sozialen Makrostrukturen, insbesondere der sozialen Normen, auf der Mikroebene sorgen.

Auf Basis der Hypothese von Damasio [Dam03] hinsichtlich der Funktionsweise der emotionalen Mechanismen bei der sozialen Gesellschaftsregulation konnte ein generelles Modell der Internalisierung von sozialen Normen durch Emotion abgeleitet werden. Damasio zeigt, wie der Mechanismus der emotionsbasierten Situationskategorisierung dafür sorgt, dass die mit negativen sozialen Emotionen (z.B. Scham, Schuld, Verlegenheit) assoziierten Handlungen unter bestimmten situationalen Gegebenheiten nicht ausgeführt werden, während mit positiven sozialen Emotionen (z.B. Selbstzufriedenheit, Stolz, Anerkennung) verknüpfte Handlungsoptionen von den Akteuren vorgezogen werden. So kommen Regelmäßigkeiten im Handeln der Akteure zustande, die als soziale Normen bezeichnet werden. Die Norm, die als eine wichtige soziologische Handlungsdeterminante der Makroebene zu betrachten ist, wird auf diese Weise unter Mitwirkung der sozialen Emotionen auf der Mikroebene fortdauernd (re-)internalisiert. Die Mechanismen der Sanktionierung normabweichenden Verhaltens und der Internalisierung der sozialen Norm durch Emotion wurden durch Petrinetz-Modelle

10. Zusammenfassung und Ausblick

dargestellt.

Diese Modelle wurden in Kapitel 6.3 wieder aufgegriffen und zusammen mit den soziologischen Ansätzen, die die Etablierung und Aufrechterhaltung der sozialen Normen durch Emotion beleuchten, zu einem generellen Modell der emotionsbasierten Normetablierung (in natürlichen und Multiagenten-Gesellschaften) integriert. Dabei bildeten die neurowissenschaftlichen Ansätze eine wertvolle Ergänzung zu den soziologischen Theorien: Das Modell der Norminternalisierung nach Damasio [Dam03] lieferte bei der Modellierung des generellen Prozesses der Normetablierung durch Emotion einen fehlenden Baustein.

Auswertung der soziologischen Emotionstheorien Bei der soziologischen Theorienübersicht wurde speziell nach solchen Ansätzen gesucht, die den Zusammenhang der emotionalen Handlungsdeterminanten der Mikro- und Makroebene beleuchten und explizit die Frage nach den Wechselwirkungen von Emotion und sozialer Norm betrachten.

In einer Reihe von soziologischen Theorien wird der Emotion Scham eine wichtige Rolle bei der Ausbildung der sozialen Normen in einer Gesellschaft zugeschrieben (so bei Simmel [Sim01], Parsons [Par94] und Durkheim [Dur95]).

Zusammengefasst konnte festgestellt werden, dass mehrere der betrachteten Ansätze die gleiche Hauptthese postulieren: Negative soziale Emotionen (wie Scham, Verachtung, Abscheu) spielen bei der Sanktionierung von Verhalten, das von gesellschaftlichen Verhaltensnormen und -konventionen abweicht, eine entscheidende Rolle. Die Internalisierung der gemeinsamen Verhaltensnormen durch jedes einzelnes Gesellschaftsmitglied wird mit Hilfe der Emotion erzwungen und so letztendlich die Ausbildung einer gemeinsamen sozialen Norm vorbereitet.

Konstruktivistische Theorien betonen die soziale Konstruktion der Emotionen und vertreten die These, dass Emotionen das Produkt sozialkultureller Deutungs- und Interpretationsmuster sind (Shott 1979, Hochschild 1983, Thoits 1989, Tritt 1991).

Symbolisch-interaktionistische Ansätze (Hochschild [Hoc79], Shott [Sho79], Averill [Ave80]) sind eine Teilmenge der konstruktivistischen Ansätze, welche zur Identifizierung und Wahrnehmung einer Emotion die Notwendigkeit einer kognitiven Interpretation der physiologischen Stimuli durch den Akteur annehmen. Die Ausdifferenzierung der Emotionen in "physiologische Stimuli" und aufgrund der kognitiven Auswertung "akut gefühlter Emotionen" ([Hoc79]) passt gut zu den einschätzungstheoretischen Sichtweisen.

Die Integrierbarkeit der symbolisch-interaktionistischen Ansätze mit den psychologischen Emotionstheorien wurde als offene Frage für weiterführende Arbeiten identifiziert.

Die Betrachtung von Kempers positivistischem Ansatz ([Kem78], [Kem84]) hat dessen besonderen Stellenwert für die Emotionsforschung offenbart. Kemper untersucht soziale Faktoren der Emotionsgenese, die weder in psychologischen Emotionstheorien noch in symbolisch-interaktionistischen Ansätzen thematisiert werden. Aufgrund des logisch-kausalen Regelmusters, mit dem Kemper die Emotionsgenese beschreibt, und seiner Vergleichbarkeit mit psychologischen Einschätzungstheorien wurde vermutet, dass sich diese durch Kempers Modell ergänzen ließen. Die detaillierte Untersuchung der Integrierbarkeit von Kempers Modell der Emotionsgenese in die Einschätzungstheorien bleibt jedoch weiterführenden Arbeiten vorbehalten.

Die Betrachtung der positivistischen Ansätze hat ferner ergeben, dass sie im Wesentlichen auf Interaktionssituationen fokussieren und deshalb für die Verbesserung von Benutzerinteraktion und Benutzer-Modellierung durch Emotionseinbettung viele interessante Möglichkeiten bieten. Es wurde die Möglichkeit der Ergänzung von Multiagentenarchitekturen um sozialwissenschaftlich fundierte Mechanismen zur Verhaltensmodellierung anderer Agenten in MAS durch Kempers "Social interactional Theory of Emotions" angedeutet. Eine genaue Untersuchung der Anwendbarkeit von Kempers Theorie zur Optimierung der Benutzerinteraktion wurde jedoch nicht durchgeführt und bleibt somit ebenfalls zukünftigen Arbeiten vorbehalten.

Ein Modell der Etablierung und Aufrechterhaltung sozialer Normen durch Emotion

Die Übersicht über neurowissenschaftliche Ansätze in Kapitel 6.2 und über sozialwissenschaftliche Theorien in Kapitel 6.3 hat den gemeinsamen Leitgedanken vieler Ansätze erkennen lassen, dass Emotion eine wesentliche Rolle bei der Etablierung und Aufrechterhaltung sozialer Strukturen (insbesondere Normen) spielt, indem sie eine Grundlage für die Internalisierung von sozialen Normen durch einzelne Akteure bildet. Auf diese Erkenntnis und eine theoretische Ausarbeitung von Staller und Petta [SP00] zurückgreifend, wurde die Möglichkeit der Emotionsimplementierung für eine flexible Kontrolle und Koordination eines Multiagentensystems diskutiert. Staller und Petta [SP00] schlagen die Implementierung von Emotion in Multiagenten-Settings zur Konstituierung von sozialen Normen vor (vgl. Kapitel 2.2). Zur Unterstützung ihrer Thesen wurde eine Reihe von emotionstheoretischen Grundlagen verschiedener Disziplinen zusammengestellt, welche allesamt die Existenz einer wechselseitigen Beziehung zwischen Emotion und sozialen Normen bestätigen: Demnach führt die emotionsbasierte Verankerung sozialer Normen in einem Agenten eines Multiagentensystems und die dadurch bedingte Erzwingung normkonformen Verhaltens zur Ausbildung dynamischer und flexibler Kontrollstrukturen auf der MAS-Ebene.

Die grundlegenden theoretischen Aussagen wurden schrittweise in Petrinetz-Modelle übersetzt, so dass am Ende die Kernaussagen der Ansätze von Elias ([Eli76], [ES90]), Elster ([Els89], [Els96], [Els99]) und Damasio ([Dam94], [Dam03]) in einer Reihe von Petrinetzen anschaulich dargestellt werden konnten. Als nächstes wurde anhand der Ansätze gezeigt, welche Funktionen Emotion bei der Etablierung der sozialen Strukturen und sozialen Normen zukommen. Schließlich wurde ein allgemeines Modell der Normetablierung durch Emotion unter Einbeziehung der in Kapitel 6.2 erstellten Modelle hergeleitet, das auf der Relation zwischen sozialen Normen und Emotion basiert. Dieses wurde in Kapitel 9 als ein Anwendungsbeispiel für das E-MULAN-Referenzmodell nachgebildet.

Auswertung der informatischen Emotionsarchitekturen Der Übersicht der Emotionsmodellierungsansätze hat in erster Linie die Heterogenität der informatischen Emotionsmodelle gezeigt. Nach der Illustration verschiedener Klassifizierungsmöglichkeiten emotionaler Agentenarchitekturen wurden die ausgewählten KI- und VKI-Modelle der Emotion vorgestellt und mit Petrinetzen modelliert.

Bei den VKI-Ansätzen wurden zwei Multiagentenarchitekturen - von Elliot [Ell92] und von Bates und Reilly [BLR92b] - untersucht. Es hat sich gezeigt, dass die beiden Ansätze

10. Zusammenfassung und Ausblick

die funktionalen Einflüsse der Emotion auf das Systemverhalten nicht implementieren. Nur der Ansatz von Aubé und Senteni hat den funktionalen Einsatz der Emotionen zur Ressourcenkontrolle und Agentenkommunikation in einem MAS illustriert.

Bei der Betrachtung der KI-Ansätze wurden die emotionalen Architekturen von Staller und Petta [SP98], Velásquez [Vel98], Dörner [DSD01], Bach [Bac03] und Sloman [Slo01] vorgestellt. Aus der Übersicht der Modellierungsansätze wurde ersichtlich, dass die Emotionskonzepte der Forscher teilweise sehr weit auseinanderliegen.

Wegen des Fehlens einer gemeinsamen emotionstheoretischen Grundlage sind informatische Modelle, die auf unterschiedlichen Emotionsdefinitionen gründen, nur bedingt miteinander vergleichbar; die informatische Diskussion um die Emotionsmodellierung wird durch die stark divergierenden Emotionskonzepte der Forscher sehr erschwert.

Des Weiteren ist für die informatische Emotionsmodellierung eine intuitive Vorgehensweise charakteristisch; viele Modelle weisen einseitige oder nur unzureichende sozialwissenschaftliche Fundierung auf. Häufig werden nur bestimmte Elemente einer Emotionstheorie herausgegriffen, die dann um eigene Konzepte erweitert werden. Insbesondere wird die soziale Komponente der Emotionalität oftmals vernachlässigt: Nach wie vor dominieren bei der Entwicklung emotionaler Systeme kognitionspsychologische Emotionstheorien, auch dann, wenn es sich um künstliche Sozialsysteme, beispielsweise auf Basis von Multiagentensystemen handelt (z.B. [Ell92], [Bat94], [AS96], [Vel97], [SP98]).

Um die Unvollständigkeit der "Teiltheorien" auszugleichen, sind die Forscher aufseiten der Informatik meist gezwungen, mehrere Emotionstheorien gleichzeitig in ihren Modellen umzusetzen oder ausgewählte Aspekte verschiedener Theorien zu kombinieren (z.B. [Bat94], [Can97], [EO92], [GM04]). Diese Vorgehensweise kann aber wegen der uneinheitlichen Terminologie Schwierigkeiten bereiten und macht eine genaue Untersuchung der grundlegenden theoretischen Konzepte unbedingt erforderlich, um sicherzustellen, dass die Theorien von gleichen Grundbegriffen und Vorannahmen bezüglich der Emotion ausgehen. Die Validität von solchen kombinierten Emotionstheorien wurde deshalb in Frage gestellt.

Die Diskussion der Ansätze hat somit diverse Probleme der Emotionsmodellierung erkennen lassen, die auf das Fehlen einer gemeinsamen Beschreibungssprache für emotionale Prozesse / Architekturen zurückzuführen sind. Um informatische Emotionsmodelle / emotionale Agentenarchitekturen miteinander vergleichen und zueinander in Beziehung setzen zu können, wird eine einheitliche Modellierungstechnik als architekturunabhängiges Rahmenwerk benötigt, das eine Repräsentation emotionaler Agentenarchitekturen, Konzepte und Prozesse ermöglicht.

Referenzmodell für Emotionsmodellierung Um eine adäquate theoretische Fundierung der Emotionsmodellierung in (verteilten) künstlichen Systemen zu erschaffen, wurde im darauffolgenden Kapitel ein Vorschlag für ein kohärentes konzeptuelles Rahmenwerk unterbreitet, das die Analyse der Emotion aus verschiedenen Perspektiven erleichtern soll. Das Ziel bestand u.a. darin, ein Modellierungsrahmenwerk zur Verfügung zu stellen, das einerseits als konzeptionelle Modellierungstechnik zur Modellierung und zum Entwurf emotionaler Agentenarchitekturen eingesetzt werden kann, andererseits als Darstellungstechnik zur Repräsentation emotionaler Prozesse auf verschiedenen Ebenen dienen kann.

Emotion wurde im Kontext eines Mikro-Makro-Rahmenwerks betrachtet und ihre Wirkungsweise auf vier verschiedenen Ebenen untersucht. Die Ebenen korrespondieren unmittelbar mit den entsprechenden Ebenen der MULAN-Architektur [Röl04], die, mit den emotionstheoretischen Grundlagen kombiniert, eine Basis für das Modellierungsrahmenwerk bildet. Die MULAN-Architektur wurde insbesondere benutzt, um die kognitiven Komponenten des Rahmenwerks abzubilden.

Durch eine *Shadow*-Architektur wurde zu jeder rationalen / kognitiven Rahmenwerkkomponente eine mit dieser korrespondierende Komponente zur Modellierung der Emotion in Beziehung gesetzt. Bei den emotionalen Komponenten wurde auf eine einfache Anbindung an die existierende Architektur geachtet. Der wechselseitige oder unidirektionale Einfluss zwischen den Rahmenwerkkomponenten wurde durch synchrone Kanäle realisiert.

Die Möglichkeit expliziter Modellierung der kognitiven und emotionalen Prozesse wurde als wichtiges Feature des Modellierungsrahmenwerks gefordert. Die “rationalen“ und “emotionalen“ Rahmenwerk-Komponenten können bei der Modellierung durch flexible Verbindung mittels synchroner Kanäle sowohl separat als auch gemeinsam eingesetzt werden. Die rein rationalen Modelle können mit rationalen Rahmenwerkkomponenten beschrieben werden; die zuvor nicht berücksichtigten emotionalen Aspekte können durch nachträgliche Modellierung mit emotionalen Rahmenwerkkomponenten ergänzt werden.

Damit stellt die E-MULAN-Architektur eine spezielle Sicht auf das System dar, in die insbesondere emotionale Prozess bzw. Verhaltensaspekte ausgegliedert werden, währendin die rationale / kognitive Verhaltenskomponenten weiterhin in der MULAN-Architektur verbleiben.

Die Kopplung der beiden unterschiedlichen Modellierungsebenen (Rationalität und Emotion) erlaubt zum einen die theoretische Untersuchung wechselseitiger Zusammenhänge und Abhängigkeiten und zum anderen die technische Zusammenführung und somit eine direkte Umsetzung emotionaler Modelle im Kontext von Multiagentensystemen.

Anschließend wurde die Implementierung einzelner Architekturkomponenten dargestellt und diskutiert. Eine wichtige Grundlage für diese Darstellung wurde bereits durch die Modellierung ausgewählter Emotionstheorien und Agentenarchitekturen mit Petrinetz-Modellen gelegt. Auf diese Modelle wurde nun bei der Ausgestaltung einzelner Architekturkomponenten zurückgegriffen. Es wurde gezeigt, wie sich diese Komponenten aus den in Kapiteln 5 bis 6.2 dargestellten psychologischen, soziologischen bzw. neurowissenschaftlichen Emotionstheorien herleiten lassen und wie die abgeleiteten Architekturkomponenten zur Modellierung der Emotionstheorien angewendet werden können.

Für die Synchronisation der aufgefalteten E-MULAN-Modellebenen wurde die MULAN-Architektur um einen Synchronisationsmechanismus erweitert. Die bisherige MULAN-Architektur sah in der konkreten Umsetzung mittels CAPA lediglich eine asynchrone Kommunikation vor. Durch die Erweiterung von MULAN um eine synchrone Kommunikation wurde eine direkte Synchronisation auf Agentenebene (insbesondere der emotionalen und rationalen Agentenkomponenten) möglich. Dadurch lassen sich viele Anwendungsszenarien im E-MULAN-Referenzmodell einfacher umsetzen und simulationstechnisch nachverfolgen. Um das exemplarisch zu zeigen, wurde in Kapitel 8.4.4 mit dem erweiterten Modell eine synchro-

10. Zusammenfassung und Ausblick

ne Kommunikation zwischen einer rationalen und einer emotionalen Agentenkomponente realisiert und mit einem Kausalnetz veranschaulicht.

Der hier für E-MULAN präsentierte Synchronisationsmechanismus erweitert auch MULAN um eine synchrone Schnittstelle, er ist nicht auf die Anwendungsdomäne der Emotionstheorien beschränkt. Inwieweit andere Agentenmodelle von dem hier vorgestellten Verfahren Anregungen für eine synchrone Architekturweiterung erhalten, bleibt in weiterführenden Arbeiten zu klären. Die Umsetzung der Möglichkeit zur synchronen Modellierung in konkreten Simulationen bzw. der Parallelisierung von Einschätzungen im Emotionsgenerierungsprozess ist nur im Rahmen einer differenzierten Betrachtung der betreffenden Theorien möglich und kann nur auf der Seite der Psychologie erfolgen. Im Rahmen dieser Arbeit wird diese Fragestellung explizit ausgeschlossen, da nur die Modellierung der bestehenden etablierten theoretischen Emotionsgrundlagen angestrebt wird und keine Umschreibung / Weiterentwicklung von Emotionstheorien auf Basis der verwendeten Modellierungstechnik.

Die Implementierung des Rahmenwerks und seine Verwendung als Denkraum und Repräsentationssprache für emotionstheoretische Konzepte wurde im Anschluss anhand von ausgewählten Anwendungsbeispielen illustriert. Als Anwendungsbeispiele wurden vorzugsweise solche Szenarien gewählt, die bereits im theoretischen Teil der Arbeit ausführlich diskutiert wurden: Insbesondere wurde die Anwendung des Referenzmodells E-MULAN zur Modellierung der reaktiven (Kapitel 9.2) und proaktiven (Kapitel 9.3) Emotionsgenerierung, zur Modellierung der emotionsbasierten Verhaltenssteuerung (Kapitel 9.4), sowie zur Etablierung vom normkonformen Verhalten (Kapitel 9.5) herangezogen.

10.2. Offene Fragen und Ausblick

Die Ergebnisse dieser Arbeit fügen sich in ein interdisziplinäres Feld der Emotionserforschung und -modellierung ein, das zurzeit vor allem auf der Seite der Informatik rasante Entwicklungen aufweist. Angesichts dieses schnellen Wachstums der letzten Jahrzehnte ist eine Konsolidierung der Forschungsergebnisse wünschenswert, da ansonsten eine starke Forschungsfragmentierung und mangelnde Vernetzung einem weiteren Erkenntniswachstum auf diesem Gebiet entgegenwirkt. Zahlreiche Emotionsforscher betonen die Notwendigkeit einer gemeinsamen Modellierungssprache zur Beschreibung und Darstellung von Emotionstheorien ([Sch02], [Slo04]). Von großem Interesse ist die Entwicklung einer Modellierungstechnik, mit der sowohl sozialwissenschaftliche als auch informatische Theorien und Modelle der Emotion in einer allgemein zugänglichen und verständlichen Weise dargestellt werden könnten.

Im Rahmen dieser Arbeit wurde der sozionische Modellierungsansatz benutzt, um psychologische, soziologische und neurowissenschaftliche Emotionstheorien und -modelle zu repräsentieren. Das Ziel lag hierbei einerseits in der Erprobung eines petrinetzbasierten Modellierungsansatzes für die Abbildung der emotionalen Theorien, andererseits in der Abbildung der theoretischen Grundlagen der Emotion mit einer (semi-)formalen informatischen Modellierungstechnik. Es wurde ein Referenzmodell für die Emotionsmodellierung vorgeschlagen, das als Rahmenwerk für die Modellierung von Emotionstheorien bzw. zur Implementierung von Emotion in informatischen (Multi-)Agentensystemen dienen kann.

Bei der Betrachtung der emotionstheoretischen Grundlagen bzw. der informatischen Emotionsmodelle hat sich eine Reihe von weiterführenden Forschungsfragen ergeben, die im Folgenden zusammengefasst werden.

Die Untersuchung der emotionstheoretischen Grundlagen von Emotion machte deutlich, dass eine interdisziplinäre Betrachtung der Emotion, die für die informatische Modellierung von essenzieller Bedeutung ist, bisher nicht im ausreichenden Maße stattgefunden hat. Die zeitgenössische Emotionsforschung kann der Informatik lediglich eine Reihe von “Teiltheorien“ zur Verfügung stellen, die jeweils Aspekte des Emotionsphänomens beschreiben, die Integration der verschiedenen Ansätze in einer interdisziplinären Theorie der Emotion ist jedoch bisher nicht erfolgt (vgl. [ZS00]).

Bislang wurden in Bezug auf eine umfassende Konzeptualisierung der Emotion kaum Querbezüge zwischen den wegweisenden Ergebnissen verschiedener Disziplinen (Psychologie, Soziologie, Neuro- und Kognitionswissenschaften) hergestellt. Die Emotionstheorien verschiedener Disziplinen stehen derzeit nebeneinander, ohne dass Versuche unternommen werden, die verschiedenen Ansätze in einer interdisziplinären Emotionstheorie zu vereinen. Eine solche Theorie wird insbesondere in typischen Anwendungsdomänen für emotionale Agenten benötigt, wie z.B. in der MCI oder der VKI.

In diesem Zusammenhang lassen sich von der Zusammenarbeit der Informatik mit der psychologischen, soziologischen und neurowissenschaftlichen Emotionsforschung wichtige Impulse für die Erarbeitung, Modellierung und simulationstechnische Validierung des übergreifenden Theorieansatzes erwarten ([vLMV03]; [Mal98], [vSvL05]). Die Tatsache, dass die Informatik im Anwendungskontext der Agenten- und Multiagentensysteme mit dem Problem der Entstehung und Reproduktion sozialer Ordnung befasst ist, lässt insbesondere eine Kooperation mit den Sozialwissenschaften zu diesem Themenkomplex überaus erfolgreich erscheinen (vgl. [vLMV03]).

U.a. wäre eine Evaluierung des in [Dam03] bzw. [SP00] vorgeschlagenen Mechanismus der emotionsbasierten Etablierung der Normen auf der Makroebene mit Hilfe des vorgeschlagenen Referenzmodells denkbar. Dazu wäre eine umfangreiche Implementation erforderlich, bei der eine Gesellschaft aus mehreren emotionalen Agenten nachgebildet und ihre Strukturodynamik in einer Reihe von Simulationen untersucht wird.

Die MULAN-basierte emotionale Agentenarchitektur auf Basis von E-MULAN liefert eine um Emotion verbesserte Architektur für Agentensysteme, ohne jedoch einen Bezug zu speziellen Anwendungsbereichen herzustellen, d.h. es handelt sich um eine allgemeine und weitgehend universell einsetzbare Architektur. Weiterführende Arbeiten müssen klären, für welche Anwendungsfelder innerhalb des breiten Forschungsgebiets “Affective Computing“ die E-MULAN-Architektur bzw. die auf Basis von E-MULAN entwickelten Agentenarchitekturen besonders geeignet sind.

Die Möglichkeiten und Grenzen des hier vorgeschlagenen Modellierungsansatzes für die Repräsentation der Emotionstheorien und emotionaler Agentenarchitekturen können ebenso erst in den weiterführenden Arbeiten abschließend geklärt werden: Durch die Modellierung von mehreren verschiedener Emotionstheorien und Emotionsmodelle unter Verwendung von E-MULAN-Modellierungstechnik kann diese erprobt und evaluiert werden und ihre Verwendbarkeit zur konzeptuellen Beschreibung sozialwissenschaftlicher Emotionstheorien

10. Zusammenfassung und Ausblick

bzw. zur Repräsentation informatischer Emotionsmodelle und -architekturen praktisch nachgewiesen werden.

Die Umsetzung der Möglichkeit zur synchronen Modellierung in konkreten Simulationen bzw. Parallelisierung von Einschätzungen im Emotionsgenerierungsprozess ist nur im Rahmen der differenzierten Betrachtung der betreffenden Theorien möglich und kann nur auf der Seite der Psychologie erfolgen. Im Rahmen dieser Arbeit wird diese Fragestellung explizit ausgeschlossen, da nur die Modellierung der bestehenden etablierten theoretischen Emotionsgrundlagen angestrebt wird und keine Umschreibung / Weiterentwicklung von Emotionstheorien auf Basis der verwendeten Modellierungstechnik betrieben werden soll.

Das E-MULAN-Referenzmodell ist sowohl als Modellierungsrahmen zum Entwurf einer konkreten ausführbaren emotionalen (Multi-)Agentenarchitektur als auch als konzeptuelles und terminologisches Referenzmodell zur Klassifizierung und zum Vergleich sozialwissenschaftlicher Emotionstheorien und informatischer Systemarchitekturen einsetzbar. Aufgrund der operationalen Semantik durch die Petrinetze liegt eine Simulation nahe bzw. in Verbindung mit MULAN, CAPA und dem PAOSE-Ansatz die direkte Implementierung. Die Erprobung des Referenzmodells für die Modellierung und die anschließende Simulation einer oder mehrerer Emotionstheorien bzw. Theorieteile auf verschiedenen Architekturebenen wäre eine weitere Forschungsaufgabe für weiterführende Arbeiten.

Die Implementierung einer konkreten emotionalen Agentenarchitektur auf Basis von E-MULAN würde eine simulationstechnische Erprobung vieler Anwendungsszenarien ermöglichen, die sowohl innerhalb als auch außerhalb der Informatik von Interesse sind (z.B. für psychologische und soziologische Ansätze zum Phänomen der Emotion). Die bereits existierenden MULAN-Anwendungen können bei Bedarf mit Hilfe des Referenzmodells um Emotionsrepräsentation und -verarbeitung ergänzt werden. Das Vorgehen und ggf. die Entwicklung einer Werkzeugunterstützung für eine solche Integration bleiben jedoch ebenfalls weiterführenden Arbeiten vorbehalten.

Die Tragfähigkeit des hier vorgeschlagenen Modellierungsansatzes für die Repräsentation der Emotionstheorien und der emotionalen Agentenarchitekturen könnte in einem solchen Rahmen abschließend geklärt werden: Durch die Modellierung von mehreren Emotionstheorien und Emotionsmodellen unter Verwendung der hier vorgestellten E-MULAN-Modellierungstechnik kann diese erprobt und evaluiert sowie ihre Verwendbarkeit zur konzeptuellen Beschreibung sozialwissenschaftlicher Emotionstheorien bzw. zur Repräsentation informatischer Emotionsmodelle und -architekturen praktisch nachgewiesen werden.

Die analytische Trennung von Emotion und Kognition bzw. die Separierung der Emotion von anderen Anteilen sozialer Modelle mit Hilfe einer (semi-)formalen Modellierungstechnik ist ein wichtiger Schritt für das grundsätzliche Verständnis sozialer Theorien. Der Beitrag für die Informatik ergibt sich unmittelbar aus der Möglichkeit, noch genauere Abbildungen der Theorien vornehmen zu können. Dabei ist aber das tiefgehende Verständnis für die inhaltlichen Bestandteile der jeweiligen Emotionstheorien von großer Bedeutung. Die Komplexität der Modelle und die methodischen Unterschiede stellen dabei eine besondere Herausforderung dar. Ohne die Beteiligung sozialwissenschaftlicher Partner können innerhalb der Informatik somit keine neuen Modelltheorien erschlossen werden. Die Umsetzung der emotionstheoretischen Erkenntnisse zur Gestaltung konkreter Anwendungen, z.B. im MCI-

Bereich, stellt eine weitere Herausforderung dar. Hier ergeben sich neue Möglichkeiten der Modellierung. Durch die Verknüpfung über die verschiedenen Ebenen hinweg können systematisch konkrete Querbezüge zwischen den bisher vereinzelt Theorien hergestellt werden. Dabei ist die Beteiligung der Sozialwissenschaften unabdingbar, da die Verortung und die Absicherung über die jeweiligen Verfahren und Methoden zwingend notwendig sind.

Gleichzeitig kann mittels E-MULAN eine Simulationsumgebung erstellt werden, die eine ebenenübergreifende Repräsentation ermöglicht und so Simulations- und Prognoseverfahren für gezielte Untersuchungen bereitstellt. Werden die Ideen aus E-MULAN für die Entwicklung von Multiagentensystemen aufgegriffen, dann ergeben sich ebenfalls neue Möglichkeiten. Gezielt können in diese softwaretechnischen Systeme Mechanismen eingebaut werden, die als Steuerung fungieren. Dabei verfolgt man die Idee der Emotion als einer basalen Steuerung menschlichen Verhaltens, die durch die Umsetzung des Synchronisationsmechanismus von Agenten und anderen Einheiten der (E)-MULAN-Architektur über Ebenen hinweg möglich wird. Wird diese Erweiterung in Verbindung mit E-MULAN genutzt, dann ergibt sich für die emotionalen Bestandteile von komplexen Gesellschaften eine bisher nicht möglich gewesene separate Darstellung mittels expliziter Modelle. In den Anwendungsbeispielen im Rahmen dieser Arbeit wurde bisher lediglich die Realisierbarkeit der Steuerungsfunktion der Emotion in E-MULAN exemplarisch in einem einfachen Anwendungsszenario demonstriert (Kapitel 9.4).

Die Konstruktion und Implementierung eines elaborierteren Anwendungsbeispiels, das den kompletten Emotionsprozess von der Stimulusbewertung bis zur Auswahl einer optimalen Verhaltensweise aufgrund der akkumulierten emotionalen Erfahrungen modelliert, wäre eine weitere interessante Fragestellung für weiterführende Arbeiten. Dazu könnten die beschriebenen Modelle der reaktiven und proaktiven Emotionsgenerierung genutzt werden, um Emotionen dynamisch zur Laufzeit zu generieren. Ebenso könnten die Verhaltensstrategien in Abhängigkeit von dem Modellgegenstand, der Anwendungsdomäne, dem Modellierungszweck und ggf. unter Berücksichtigung der emotionstheoretischen Fundierung der Emotion-Verhalten-Relation gewählt werden. Zudem könnten sowohl konkrete Verhaltensweisen als auch generelle Tendenzen - im Sinne der Handlungstendenzen nach [FS87] - angegeben werden. Um das Agentenverhalten für eine konkrete Anwendung abzubilden, wäre die Repräsentation der Verhaltensweisen durch die MULAN-Protokolle denkbar.

Die Erweiterung und Übertragung der Ergebnisse auf andere Agentenarchitekturen und Agentenentwicklungsansätze wäre ebenfalls in zukünftigen Arbeiten zu behandeln. Für eine engere Kopplung von Modellen, insbesondere von petrinetzbasierten Agentensystemen, wurde hier ein Synchronisationsmechanismus über synchrone Kanäle dargestellt. Damit lassen sich für die E-MULAN-Architektur die zuvor analytisch getrennten rationalen Bestandteile mit den emotionalen Bestandteilen der kognitiven Theorien wieder vollständig und effizient integrieren. Diese Modellierung kann im Prinzip für jede Art der (MULAN-)Modellierung verwendet werden. Sie ist nicht auf die Anwendungsdomäne der Emotionstheorien beschränkt. Jegliche zu synchronisierenden Agenten können mittels dieses Verfahrens modelliert werden. Insbesondere wurde hier eine technische Umsetzung für die CAPA-Umgebung vorgestellt. Andere Agentenmodelle können von dem hier vorgestellten Verfahren wichtige Anregungen für eine synchrone Architekturweiterung übernehmen.

10. Zusammenfassung und Ausblick

Eine Adaption der konzeptuellen Trennung von Emotion und Kognition für BDI-Modelle (E-BDI) oder Simulationsumgebungen von sozialen Gruppen oder menschlichen Akteuren liegt nahe. Dabei sollten die semiformalen Darstellungen der Emotionstheorien aus dem ersten Teil der Arbeit von Nutzen sein, wobei der vermittelnde Charakter der semiformalen Petrinetz-Modelle sowohl für die Informatiker als auch für die Sozialwissenschaftler von Bedeutung ist. Beide können unmittelbar auf eine vereinfachte Version der Theorien zurückgreifen. Zwingend ist jedoch für eine umfassende und nachhaltige Verwendung der Modelle, dass die jeweiligen Beteiligten sich nach einer ersten Orientierung verstärkt mit den Grundlagen der Theorien und ihren Einbettungsmöglichkeiten hinsichtlich des gewünschten Zwecks auseinandersetzen müssen.

Zumindest für die Informatik werden durch die Petrinetz-basierten Darstellungen und die RENEW / MULAN / CAPA-Umgebungen verschiedene Umgebungen zur Verfügung gestellt. RENEW erlaubt die prinzipielle Einbettung in eine Petrinetz-Simulation. Weiterhin stehen Nebenläufigkeit, Konfliktlösung und Synchronisation als basale Konzepte für eine Java-Implementierung bereit. Mit MULAN werden konzeptuelle Brücken für eine zielgerichtete Strukturierung von großen Modellen möglich. Die Eigenschaften der jeweiligen Teilnetze werden spezifischer und erlauben somit eine anwendungsnähere Modellierung. Mit CAPA steht eine Infrastruktur zur Verfügung, die die Simulation auf einem Rechner noch stärker unterstützt, aber insbesondere werden Mittel für die Entwicklung verteilter Implementationen und Simulationen bereitgestellt. Diese vielfältige Umgebung ermöglicht viele anwendungsspezifische Nutzungen, die in weiterführenden Arbeiten geklärt und erörtert werden sollten. Offen bleibt, inwiefern die sehr speziellen Petrinetz-Modelle außerhalb der Petrinetz-Community akzeptiert und aufgegriffen werden. Oft werden Vereinfachungen angestrebt, z.B. durch vermeintlich simplere Modellierungstechniken. Diese haben aber meist nicht mehr die Möglichkeit einer semantisch umfangreichen Repräsentation der Inhalte.

Neben diesen bisher eher technischen oder abstrakten Betrachtungen von Emotionstheorien stellt sich aber auch die Frage nach der Verwendung der Ergebnisse in der Praxis. Durch das immer tiefere Verständnis des Emotionsphänomens werden immer bessere Möglichkeiten der Benutzermodellierung geschaffen. Im Bereich der sozialen (webbasierten) und damit technisch auswertbaren Netzwerke kann die Benutzermodellierung in Bezug auf die Erfassung der emotionalen Zustände einzelner Netzwerk-Nutzer gezielt betrieben werden. Dies wird durch immer besser auf das einzelne Individuum ausgerichtete Untersuchungen unterstützt. Die immer schärferen Benutzerprofile, die sehr vorteilhaft sein können, um gezielt bessere Dienste bereitzustellen, bringen jedoch auch die Möglichkeit des Missbrauchs bzw. der Manipulation mit sich. So können etwa die Netzwerk-Nutzer unterschwellig in eine bestimmte Richtung emotional beeinflusst werden, ohne dass sie dies selbst bemerken, kontrollieren oder verhindern können. Dass mit der Automatisierung der Erfassung und Gestaltung von emotionssensiblen Modellen in den persönlichen Bereich jedes Einzelnen eingegriffen wird, ist offensichtlich. Diese Arbeit liefert u.a. die Möglichkeit einer tieferen Einsicht in diese Zusammenhänge und ermöglicht somit die Diskussion dieser Thematik und deren zielgerichtete Gestaltung, z.B. in Form einer emotionalen Firewall zum Schutz der Persönlichkeit, wie dies im technischen Bereich schon länger existiert.

Literaturverzeichnis

- [Ado99] ADOLPHS, R.: *Social Cognition and the Human Brain*. Trends in Cognitive Sciences, 3(12):469–479, 1999.
- [Ado03] ADOLPHS, R.: *Cognitive Neuroscience of Human Social Behaviour*. Nature Reviews Neuroscience, 4:165–178, 2003.
- [ADP06] AYLETT, R. S., J. DIAS und A. PAIVA: *An affectively-driven planner for synthetic characters*. In: *ICAPS, AAAI Press*, 2006.
- [AI99] ABE, J. A. und C. E. IZARD: *The Developmental Functions of Emotions: An Analysis in Terms of Differential Emotions Theory*. Cognition and Emotion, 13:523–549, 1999.
- [AKGR00] ANDRE, E., M. KLESEN, S. GEBHARD, P. ALLEN und T. RIST: *Integrating Models of Personality and Emotions into Lifelike Characters*. In: PAIVA, A. (Herausgeber): *Affective Interactions: Towards a New Generation of Computer Interfaces*, LNAI 1814, Seiten 150–165. Springer, Berlin, Heidelberg, 2000.
- [ARM98] ANDRE, E., T. RIST und J. MULLER: *WebPersona: A life-like presentation agent for the world-wide web*. Knowledge-Based Systems, 11:25–36, 1998.
- [ARMK00] ANDRE, E., T. RIST, S. V. MULKEN und M. KLESEN: *The automated design of believable dialogues for animated presentation teams*. In: CASSELL, J., J. SULLIVAN, S. PREVOST und CHURCHILL E. (Herausgeber): *Embodied Conversational Agents*, Seiten 220–255. MIT Press: Cambridge, MA, 2000.
- [AS96] AUBE, M. und A. SENTENI: *Emotions as Commitments Operators: A Foundation for Control Structure in Multi-Agents Systems*. In: VELDE, W. VAN DE und J. W. PERRAM (Herausgeber): *Agents Breaking Away. Proceedings of the 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'96)*, Eindhoven/NL, Band 1038, Seiten 13–25. Berlin, Springer, 1996.
- [ASSCL95] ARMONY, J. L., D. SERVAN-SCHREIBER, J. D. COHEN und J. E LEDOUX: *An anatomically constrained neural network model of fear conditioning*. Behavioral Neuroscience, 109:246–257, 1995.
- [ASSCL97] ARMONY, J. L., D. SERVAN-SCHREIBER, J. D. COHEN und J. E LEDOUX: *Computational modeling of emotion: Explorations through the anatomy and physiology of fear conditioning*. Trends in Cognitive Science, 1:28–34, 1997.

Literaturverzeichnis

- [Aub98] AUBE, M.: *A commitment theory of emotions*. In: *Proceedings of the 1998 AAAI Fall Symposium, Emotional and Intelligent: The Tangled Knot of Cognition (AAAI Technical Report FS-98-03)*. Orlando, FL: AAAI Press, 1998.
- [Aub05] AUBÉ, M.: *Beyond needs: Emotions and the commitments requirement*. In: DAVIS, D. N. (Herausgeber): *Visions of mind: Architectures for cognition and affect*, Seiten 21–44. Hershey, PA : Idea Group. Inc., 2005.
- [Aub09] AUBÉ, M.: *Unfolding commitments management: A systemic view of emotions*. In: *Handbook of research on synthetic emotions and sociable robotics: New applications in affective computing and artificial intelligence*, Seiten 198–227. Hershey, PA: Information Science Reference., 2009.
- [Aus72] AUSTIN, J.L.: *Zur Theorie der Sprechakte (How to do things with words)*. Reclam, Stuttgart, 1972.
- [Ave80] AVERILL, J. R.: *A Constructivist View of Emotion*. In: PLUTCHIK, R. und H. KELLERMAN (Herausgeber): *Emotion: Theory, Research, and Experience. Vol. 1., Theories of Emotion*, Seiten 305–339. New York, NY: Academic Press, 1980.
- [Axe86] AXELROD, R.: *An evolutionary approach to norms*. American Political Science Review, 80(4):1095–1111, 1986.
- [BA08] BECKER-ASANO, C.: *WASABI: Affect simulation for agents with believable interactivity*. Doktorarbeit, Bielefeld, University of Bielefeld, 2008. letzter Zugriff: 11.01.2012.
- [BAB02] BAZZAN, A. L., D. F. ADAMATTI und R. BORDINI: *Extending the Computational Study of Social Norms with a Systematic Model of Emotions*. In: *Proceedings of the 16th Brazilian Symposium on Artificial Intelligence*, Band 2507 der Reihe *LNCS*, Seiten 108–117. Berlin: Springer, 2002.
- [BABD02] BERTHOZ, S., J. L. ARMONY, R. J. BLAIR und R. J. DOLAN: *An fMRI study of intentional and unintentional (embarrassing) violations of social norms*. Brain, 125(8):1696–1708, 2002.
- [Bac02] BACH, J.: *Enhancing Perception and Planning of Software Agents with Emotion and Acquired Hierarchical Categories*. In: *Proceedings of MASHO 02, German Conference on AI KI 2002*. Karlsruhe, Germany, 2002.
- [Bac03] BACH, J.: *The MicroPsi Agent Architecture*. In: *Proceedings of ICCM-5*. Universitäts-Verlag Bamberg, 2003.
- [Bac05] BACH, J.: *Representations for a Complex World: Combining Distributed and Localist Representations for Learning and Planning*. In: SPRINGER

- (Herausgeber): *Biomimetic Neural Learning for Intelligent Robot*, Band 3575 der Reihe *LNCS*, 2005.
- [Bac08] BACH, J.: *Seven Principles of Synthetic Intelligence*. In: *Seven Principles of Synthetic Intelligence. Proceedings of 1st Conference on Artificial General Intelligence*, Seiten 63–74, 2008.
- [Bar99] BARBALET, J. M.: *William James' Theory of Emotions: Filling in the Picture*. *Journal for the Theory of Social Behaviour*, 29(3):251–266, 1999.
- [Bar06] BARRETT, L.F.: *Solving the emotion paradox: Categorization and the experience of emotion*. *Personality and Social Psychology Review*, 10:20–46, 2006.
- [Bat94] BATES, J.: *The Role of Emotions in Believable Agents*. In: *Communications of the ACM*, Band 37 (7), Seiten 122–125, 1994.
- [BB05] BREAZEAL, C. und R. BROOKS: *A Functional Perspective*. In: FELLOUS, J.-M. und A. M. ARBIB (Herausgeber): *Who needs emotions?: the brain meets the robot*, Oxford University Press, Seiten 271–311, 2005.
- [BC06] BLANCHARD, A. und L. CAÑAMERO: *Developing Affect-modulated behaviors: stability, exploration, exploitation, or imitation?* In: *6th International Workshop on Epigenetic Robotics, Paris*, 2006.
- [BD07] BUSEMEYER, J. R. und E. DIMPERIO: *Integrating emotional processing into decision-making models*. In: GREY, W. D. (Herausgeber): *Integrated models of cognitive systems*. Oxford, Oxford University Press, 2007.
- [BDD00] BECHARA, A., H. DAMASIO und A. R. DAMASIO: *Emotion, Decision Making and the Orbitofrontal Cortex*. *Cerebral Cortex*, 10(3):295–307, 2000.
- [BDDA94] BECHARA, A., A. R. DAMASIO, H. DAMASIO und S. W. ANDERSON: *Insensitivity to future consequences following damage to human prefrontal cortex*. *Cognition*, 50:7–15, 1994.
- [BDTD97] BECHARA, A., H. DAMASIO, D. TRANEL und A. R. DAMASIO: *Deciding Advantageously Before Knowing the Advantageous Strategy*. *Science*, 275(5304):1293–1295, 1997.
- [BDTD05] BECHARA, A., H. C. DAMASIO, D. TRANEL und A. DAMASIO: *The Iowa gambling task (IGT) and the somatic marker hypothesis (SMH): Some questions and answers*. *Cognitive Sciences*, 9:159–162, 2005.
- [BFG⁺97] BONASSO, R. P., J. R. FIRBY, E. GAT, D. KORTENKAMP, D. P. MILLER und M. G. SLACK: *Experiences with an Architecture for Intelligent, Reactive Agents*. *Journal of Experimental and Theoretical Artificial Intelligence*, 9(2):237–256, 1997.

Literaturverzeichnis

- [BHRH00] Busetta, P., N. Howden, R. Rönquist und A. Hodgson: *Structuring BDI Agents in Functional Clusters*. In: Jennings, Nicholas und Yves Lépérance (Herausgeber): *Intelligent Agents VI. Agent Theories, Architectures, and Languages*, Band 1757 der Reihe *Lecture Notes in Computer Science*, Seiten 277–289. Springer Berlin / Heidelberg, 2000.
- [BLR92a] Bates, J., A. B. Loyall und W. S. Reilly: *An Architecture for Action, Emotion, and Social Behaviour*. In: Castelfranchi, C. und E. Werner (Herausgeber): *Lecture Notes In Computer Science, Selected papers from the 4th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, Artificial Social Systems*, Seiten 55–68, 1992.
- [BLR92b] Bates, J., A. B. Loyall und W. S. Reilly: *Integrating Reactivity, Goals, and Emotion in a Broad Agent*. In: *Proceedings of the 14th Annual Conference of the Cognitive Science Society*. Bloomington, Indiana, 1992.
- [BMO01] Bauer, B., J. P. Müller und J. Odeh: *Agent UML : A Formalism for Specifying Multiagent Interaction*. In: Ciancarini, P. und M. Wooldridge (Herausgeber): *Agent-Oriented Software Engineering*, Seiten 91–103. Springer, 2001.
- [BodAHN03] Burghouts, R., D. Op den Akker, M. Poel Heylen und A. Nijholt: *An Action Selection Architecture for an Emotional Agent*. In: *Recent Advances in Artificial Intelligence, Proceedings of FLAIRS 16*, Seiten 293–297, 2003.
- [Bou70] Bourdieu, P.: *Zur Soziologie der symbolischen Formen*. Suhrkamp, Frankfurt a. M., 1974 (franz. 1970).
- [BOvD00] Bauer, B., J. Odeh und P. H. van Dyke: *Extending UML for Agents*. In: *Proceedings of Agent-Oriented Information Systems Workshop*, Seiten 3–17, 2000.
- [BP02] Blocher, K. und R. W. Picard: *Affective Social Quest: Emotion Recognition Therapy for Autistic Children*. In: K. Dautenhahn, A. Bond, L. Canamero und B. Edmonds (Herausgeber): *Socially Intelligent Agents - Creating Relationships with Computers and Robots*, Cambridge: Cambridge University Press, Kluwer Academic Publishers, The Netherlands, 2002.
- [BPR01] Bellifemine, F., A. Poggi und G. Rimassa: *JADE: A FIPA-2000 Compliant Agent Development Environment*. In: *In Proceedings of the Fifth International Conference on Autonomous Agents*. ACM Press, 2001.
- [Bra90] Bratman, M. E.: *Intentions, Plans and Practical Reason*. Harvard University Press, Cambridge, MA, 1990.
- [Bra97] Bradshaw, J. M.: *Software Agents*. Menlo Park: AAAI Press / MIT Press, 1997.

- [BRHL99] BUSETTA, P, R RÖNNQUIST, A HODGSON und A LUCAS: *JACK Intelligent Agents - Components for Intelligent Agents in Java*. AgentLink News Letter, Seiten 2–6, 1999.
- [BTS02] BUSEMEYER, J. R., J. R. TOWNSEND und J. C. STOUT: *Motivational Underpinnings of Utility in Decision Making: Decision Field Theory Analysis of Deprivation and Satiation*. In: MOORE, S. (Herausgeber): *Emotional Cognition*, John Benjamins: Amsterdam, 2002.
- [Buc00] BUCK, R.: *The Epistemology of Reason and Affect*. In: BOROD, J.C. (Herausgeber): *Neuropsychology of Emotion*, Seiten 31–55, New York, NY:Oxford University Press, 2000.
- [Bur98] BURT, A.: *Research into Emotionally Intelligent Agents: The Outline of a Resource-Oriented Approach*. In: PETTA, P. (Herausgeber): *Grounding Emotions in Adaptive Systems at 5th International Conference of the Society for Adaptive Behavior (SAB'98)*, Seiten 43–48. Zurich, Switzerland: AAAI Press, 1998.
- [BZO96] BEN-ZE'EV, A. und K. OATLEY: *The Intentional and Social Nature of Human Emotions: Reconsideration of the Distinctions between Basic and Non-Basic Emotions*. Journal for the Theory of Social Behaviour, 26:81–94, 1996.
- [Cab10] CABAC, L.: *Modeling Petri Net-Based Multi-Agent Applications*, Band 5 der Reihe *Agent Technology – Theory and Applications*. Logos Verlag, Berlin, 2010.
- [Can27] CANNON, W. B.: *The James-Lange theory of emotion: A critical examination and an alternative theory*. American Journal of Psychology, 39:106–124, 1927.
- [Can97] CANAMERO, D.: *Modeling motivations and emotions as a basis for intelligent behavior*. In: *International Conference on Autonomous Agents*, Marina del Rey, CA, 1997.
- [Can03] CANAMERO, D.: *Designing Emotions for Activity Selection in Autonomous Agents*. In: TRAPPL, R., P. PETTA und S. PAYR (Herausgeber): *Emotions in Humans and Artifacts*, Seiten 115–149. Cambridge/MA: The MIT Press, 2003.
- [Cas01] CASTELFRANCHI, C.: *The Theory of Social Functions: Challenges for Computational Social Science and Multi-Agent Learning*. Cognitive Systems Research, 2(1):5–38, 2001.

Literaturverzeichnis

- [CBA⁺02] CACIOPPO, J., G. BERNTSON, R. ADOLPHS, C. CARTER, R. DAVIDSON, M. MCCLINTOCK, B. MCEWEN, M. MEANEY, D. SCHACTER, E. STERNBERG, S. SUOMI und S. TAYLOR: *Foundations in Social Neuroscience*. Cambridge/MA: The MIT Press, 2002.
- [CBC⁺00] CASSELL, J., T. BICKMORE, L. CAMPBELL, H. VILHJALMSSON und H. YAN: *Human conversation as a system framework: Designing embodied conversational agents*. In: CASSELL, J., J. SULLIVAN, S. PREVOST und E. CHURCHILL (Herausgeber): *Embodied Conversational Agents*, Seiten 29–63, MIT Press: Boston, 2000.
- [CC95a] CONTE, R. und C. CASTELFRANCHI: *Norms as mental objects*. In: *Proceedings of the 5th European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW 1993)*, LNAI Vol. 957, Seiten 186–196, Berlin, Springer, 1995.
- [CC95b] CONTE, R. und C. CASTELFRANCHI: *Understanding the functions of norms in social groups through simulation*. In: GILBERT, N. und R. CONTE (Herausgeber): *Artificial Societies: The Computer Simulation of Social Life*, Seiten 252–267, UCL Press, London, 1995.
- [CCP98] CASTELFRANCHI, C., R. CONTE und M. PAOLUCCI: *Normative reputation and the costs of compliance*. *Journal of Artificial Societies and Social Simulation*, 1(3):5–38, 1998.
- [CdSCS11] CAMPANO, S., E. DE SEVIN, V. CORRUBLE und N. SABOURET: *Simulating Affective Behaviours: An Approach Based on the COR Theory*. In: *ACII 2011, Part I, LNCS 6974*, Seiten 457–466, Berlin Heidelberg, 2011. Springer. letzter Zugriff: 10.01.2012.
- [CG95] CONTE, R. und N. GILBERT: *Introduction: Computer simulation for social theory*. In: GILBERT, N. und R. CONTE (Herausgeber): *Artificial Societies: The Computer Simulation of Social Life*, Seiten 1–18, UCL Press, London, 1995.
- [CH94] CHRISTENSEN, S. und N.D. HANSEN: *Coloured Petri Nets Extended with Channels for Synchronous Communication*. In: VALETTE, R. (Herausgeber): *Application and Theory of Petri Nets*, LNCS, Seiten 159–178. Springer, 1994.
- [CK02] CONATI, C. und M. KLAWE: *Socially Intelligent Agents in Educational Games*. In: DAUTENHAHN, K., A. BOND, D. CANAMERO und B. EDMONDS (Herausgeber): *Socially Intelligent Agents - Creating Relationships with Computers and Robots*, Kluwer Academic Publishers, The Netherlands, 2002.
- [CM81] CLOCKSIN, W. F. und C. S. MELLISH: *Programming in Prolog*. Springer, New York, 1981.

- [CM04] CONATI, C. und H. MACLAREN: *Evaluating a Probabilistic Model of Student Affect*. In: *Proceedings of ITS 2004, 7th International Conference on Intelligent Tutoring Systems*, Seiten 385–396, Maceio, Brazil, 2004.
- [CMO72] COHEN, M. D., J. G. MARCH und J. P. OLSEN: *A garbage can model of organizational choice*. *Administrative Science Quarterly*, 17(1):1–25, 1972.
- [CN89] CARLEY, K. und A. NEWELL: *The Value of Cognitive Foundations for Dynamic Social Theory*. *Journal of Mathematical Sociology*, 14(2/3):171–208, 1989.
- [Col75a] COLBY, K.: *Artificial Paranoia: A Computer Simulation of Paranoid Process*. Pergamon Press, 1975.
- [Col75b] COLLINS, R.: *Conflict Sociology*. Academic Press, New York, 1975.
- [Col84] COLLINS, R.: *The Role of Emotion in Social Structure*. In: SCHERER, K. und P. EKMAN (Herausgeber): *Approaches to Emotion*, Seiten 385–396, Hillsdale/NJ, Lawrence Erlbaum, 1984.
- [Col91] COLEMAN, J. S.: *Grundlagen der Sozialtheorie. Bd. 1, Handlungen und Handlungssysteme*. Oldenbourg, München, 1991.
- [Con02] CONATI, C.: *Probabilistic assessment of user's emotions in educational games*. *Appl. Artif. Intelligence*, Special issue on Merging Cognition and Affect in HCI, 16:555–575, 2002.
- [CP02] COSENTINO, M. und C. POTTS: *A CASE Tool Supported Methodology for the Design of Multi-Agent Systems*. In: *Proceedings of the International Conference on Software Engineering Research and Practice*. CSREA Press, 2002.
- [CP09] CLORE, G. und J. PLAMER: *Affective guidance of intelligent agents: How emotion controls cognition*. *Cognitive Systems Research*, 10(1):21–30, 2009.
- [CSC94] CLORE, G. L., N. SCHWARZ und M. CONWAY: *Cognitive Causes and Consequences of Emotion*. In: WYER, R. S. und T. K. SRULL (Herausgeber): *Handbook of Social Cognition. 2. Aufl.*, Seiten 323–417. Hillsdale: Lawrence Erlbaum, 1994.
- [CZ02] CONATI, C. und X. ZHOU: *Modeling students' emotions from cognitive appraisal in education games*. In: *Proceedings of ITS 2002, 6th International Conference on Intelligent Tutoring Systems*, Biarritz, France, 2002.
- [Dam94] DAMASIO, A. R.: *Descartes' Error: Emotion, Reason, and the Human Brain*. Avon Books, New York, 1994.

Literaturverzeichnis

- [Dam96] DAMASIO, A. R.: *The somatic marker hypothesis and the possible functions of the prefrontal cortex*. In: *Proceedings of the Royal Society*, Band 351, Seiten 1413–1420, 1996.
- [Dam03] DAMASIO, A. R.: *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Harvest Books, 2003.
- [Dau98] DAUTENHAHN, K.: *The art of designing socially intelligent agents: science, fiction and the human in the loop*. *Applied Artificial Intelligence Journal*, 12(7-8):573–617, 1998.
- [Dau00] DAUTENHAHN, K.: *Socially Intelligent Agents and the Primate Social Brain - Towards a Science of Social Minds*. In: *Proceedings of the AAAI Fall Symposium Socially Intelligent Agents - The Human in the Loop*, Technical Report FS-00-04, Seiten 35–51. Menlo Park: AAAI Press, 2000.
- [Dau02] DAUTENHAHN, K.: *Socially Intelligent Agents in Educational Games*. In: DAUTENHAHN, K., BOND A., D. CANAMERO und B. EDMONDS (Herausgeber): *Socially Intelligent Agents - Creating Relationships with Computers and Robots*. Kluwer Academic Publishers, 2002.
- [DB07] DUNCAN, S. und L.F. BARRETT: *Affect is a form of cognition: A neurobiological analysis*. *Cognition and Emotion*, 21:1184–1211, 2007.
- [Den71] DENNETT, D. C.: *Intentional systems*. *The Journal of Philosophy*, 1971.
- [dFGQ05] FREITAS, J. S. DE, R. GUDWIN und J. QUEIROZ: *Emotion in artificial intelligence and artificial life research: facing problems*. *Lecture Notes in Computer Science*, Seiten 501–501, 2005.
- [DLD⁺01] DÖRNER, D., P. LEVI, F. DETJE, M. BECHT und D. LIPPOLD: *Der agentenorientierte, sozionische Ansatz mit PSI*. *Sozionik aktuell*, 1(2), 2001.
- [DM85] DYER, M. G. und E.T. MUELLER: *Towards a computational theory of human daydreaming*. In: *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, Seiten 120–129, Irvine, CA, 1985.
- [DMR03] DUVIGNEAU, M., D. MOLDT und H. RÖLKE: *Concurrent Architecture for a Multi-agent Platform*. In: GIUNCHIGLIA, FAUSTO, JAMES ODELL und GERHARD WEISS (Herausgeber): *Agent-Oriented Software Engineering III. Third International Workshop, Agent-oriented Software Engineering (AOSE) 2002, Bologna, Italy, July 2002. Revised Papers and Invited Contributions*, Band 2585 der Reihe *LNCIS*. Springer, 2003.
- [Dor97] DORAN, J. E.: *Distributed Artificial Intelligence and Emergent Social Complexity*. In: LEEUW, S. E. VAN DE und J. MCGLADE (Herausgeber): *Time, Process, and Structured Transformation in Archaeology*, Band 26 der Reihe *One World Archaeology Series*, Seiten 283–297. London: Routledge, 1997.

- [DP05] DIAS, J. und A. PAIVA: *Feeling and Reasoning: a Computational Model for Emotional Agents*. In: *Proceedings of 12th Portuguese Conference on Artificial Intelligence, EPIA*. Springer, 2005.
- [DP07] DALGLEISH, T. und M. POWER: *Cognition and Emotion*. Psychology Press, 2007.
- [dQFT⁺04] QUERVAIN, D.J.-F. DE, U. FISCHBACHER, V. TREYER, M. SCHELLHAMMER, U. SCHNYDER, A. BUCK und FEHR A.: *The Neural Basis of Altruistic Punishment*. *Science*, 305:1254–1258, 2004.
- [Dör94] DÖRNER, D.: *Über die Mechanisierbarkeit der Gefühle*. In: KRÄMER, S. (Herausgeber): *Geist, Gehirn, Künstliche Intelligenz*, Seiten 130–161. Berlin: deGruyter, 1994.
- [DSD01] DÖRNER, D., H. SCHAUB und F. DETJE: *Das Leben von PSI: Über das Zusammenspiel von Kognition, Emotion und Motivation - oder: Eine einfache Theorie komplexer Verhaltensweisen*. *Sozionik aktuell*, 1(2), 2001.
- [DSG03] DAVIDSON, R. J., K. R. SCHERER und H. GOLDSMITH: *Handbook of Affective Sciences*. Oxford University Press, New York, 2003.
- [Dur95] DURKHEIM, E.: *Die Regeln der soziologischen Methode*. René König, Frankfurt am Main, Suhrkamp, (1999)[1895].
- [Duv02] DUVIGNEAU, M.: *Bereitstellung einer Agentenplattform für petrinetzbasierte Agenten*. Diplomarbeit, Universität Hamburg, FB Informatik, Hamburg, 2002.
- [Duv03] DUVIGNEAU, M.: *Petrinetzbasierte Agentenprogrammierung*. In: JUH’AS, GABRIEL und ROBERT LORENZ (Herausgeber): *Algorithms and Tools for Petri Nets. 10th Workshop Proceedings*, Seiten 98–101, Eichstätt, Germany, sep 2003. GI Special Interest Group on Petri Nets and Related System Models, Katholische Universität Eichstätt-Ingolstadt.
- [Dye87] DYER, M. G.: *Emotions and their computations: Three computer models*. *Cognition and Emotion*, 1:323–347, 1987.
- [EB98] ELLIOTT, C. D. und J. BRZEZINSKI: *Autonomous Agents as Synthetic Characters*. *AI Magazine*, 19 (2):13–30, 1998.
- [EF75] EKMAN, P. und W. V. FRIESEN: *Unmasking the Face*. Englewood Cliffs, NJ: Prentice Hall, 1975.
- [EG93] ELLIOTT, C. D. und SIEGLE G.: *Variables Influencing the Intensity of Simulated Affective States*. Technischer Bericht, Stanford University, Menlo Park, CA, 1993.

Literaturverzeichnis

- [Ekm82] EKMAN, P.: *Emotions in the Human Face*. Cambridge University Press, Cambridge, 1982.
- [Ekm89] EKMAN, P.: *The argument and evidence about universals in facial expressions of emotion*. In: WAGNER, H. und A. S. MANSTEAD (Herausgeber): *Handbook of Social Psychophysiology*, Seiten 143–164, Chichester: Wiley, 1989.
- [Ekm92] EKMAN, P.: *An Argument for Basic Emotions*. Cognition and Emotion, 6(3/4):169–200, 1992.
- [EL04] EISENBERGER, N. I. und M. D. LIEBERMAN: *Why rejection hurts: a common neural alarm system for physical and social pain*. Trends in Cognitive Sciences, 8(7):294–300, 2004.
- [Eli76] ELIAS, N.: *Über den Prozess der Zivilisation. Soziogenetische und psychogenetische Untersuchungen. Band I: Wandlungen des Verhaltens in den weltlichen Oberschichten des Abendlandes*. Suhrkamp, Frankfurt am Main, 1976.
- [Ell92] ELLIOTT, C. D.: *The Affective Reasoner. A Process Model of Emotions in a Multi-Agent System*. PhD Thesis, Technical Report 32, Intitute for the Learning Sciences, Northwestern University, 1992.
- [Ell93] ELLIOTT, C. D.: *Using the Affective Reasoner to Support Social Simulations*. In: *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, Band 1, Seiten 194–200. Los Altos: Morgan Kaufmann, 1993.
- [Ell96] ELLIOTT, C. D.: *I picked up Catapia and other stories: A multimodal approach to expressivity for “emotionally intelligent“ agents*. In: *Proceedings of the First International Conference on Autonomous Agents*, Seiten 451–457, Marina del Rey, CA, February 1996.
- [Ell97] ELLIOTT, C. D.: *Affective reasoner personality models for automated tutoring systems*. In: *Proceedings of the Workshop on Pedagogical Agents*, Seiten 33–39, Kobe, Japan, 1997. Eighth World Conference on Artificial Intelligence in Education.
- [Els89] ELSTER, J.: *Social Norms and Economic Theory*. Journal of Economic Perspectives, 3(4):99–117, 1989.
- [Els96] ELSTER, J.: *Rationality and the Emotions*. The Economic Journal, 106(483):1386–1397, 1996.
- [Els99] ELSTER, J.: *Alchemies of the Mind: Rationality and the Emotions*. Cambridge University Press, Cambridge, 1999.

- [ENIY98] EL-NASR, M.S., T.R. IOERGER und J. YEN: *PETEEI: A PET with Evolving Emotional Intelligence*. Technical Report, Computer Science Departement, Texas University, 1998.
- [ENJT00] EL-NASR, M.S., YEN J. und IOERGER T.R.: *FLAME: A Fuzzy Logic Adaptive Model of Emotions*. Autonomous Agents and Mult-agent Systems, 3(3):219–257, 2000.
- [EO92] ELLIOTT, C. D. und A. ORTONY: *Point of View: Modeling the Emotions of Others*. In: *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, Seiten 809–814, Hillsdale, NJ, 1992. Lawrence Erlbaum Associates.
- [ES90] ELIAS, N. und J. L. SCOTSON: *Etablierte und Außenseiter*. Suhrkamp, Frankfurt am Main, 1990.
- [ES94] ELSAESSER, C. und M.G. SLACK: *Integrating deliberative planning in a robot architecture*. In: *Proceedings of the AIAA NASA Conference on Intelligent Robots in Field Factory Service and Space*, 1994.
- [FADT11] FEINSTEIN, J.S., R. ADOLPHS, A. DAMASIO und D. TRANEL: *The human amygdala and the induction and experience of fear*. Current Biology, 21(1):34–38, 2011.
- [FDM11] FIX, JULIA, MICHAEL DUVIGNEAU und DANIEL MOLDT: *Bereitstellung eines Synchronisationsmechanismus für Mulan-basierte Agenten*. In: BERGENTHUM, ROBIN und JOERG DESEL (Herausgeber): *Algorithmen und Werkzeuge für Petrinetze. 18. Workshop AWPN 2011, Hagen, September 2011. Tagungsband*, Seiten 8–14, 2011.
- [Fer92] FERGUSON, I. A.: *Touring Machines: Autonomous Agents with Attitudes*. Computer, 25(5):51–55, 1992.
- [Fer95] FERBER, J.: *Multiagentensysteme: Eine Einführung in die Verteilte Künstliche Intelligenz*. Addison-Wesley, München, 2001 (1995).
- [FG97] FRANKLIN, S. und A. GRAESSER: *Is it an agent, or just a program?: A taxonomy for autonomous agents*. In: MÜLLER, J. P., M. J. WOOLDRIDGE, und N. R. JENNINGS (Herausgeber): *ECAI'96 Workshop on Agent Theories, Architectures, and Languages: Intelligent Agents III*, Band 1193 of LNAI, Seiten 21–36. Berlin: Springer, 1997.
- [FG02] FEHR, E. und S. GAECHTER: *Altruistic punishment in humans*. Nature, 415:137–140, 2002.
- [fIPA02a] INTELLIGENT PHYSICAL AGENTS, FOUNDATION FOR: *FIPA Agent Management Specification*. Technischer Bericht, Document no. SC00023J, 2002.

Literaturverzeichnis

- [fIPA02b] INTELLIGENT PHYSICAL AGENTS, FOUNDATION FOR: *FIPA Communicative Act Library Specification*. Technischer Bericht, Document no. SC00037J, 2002.
- [Fir89] FIRBY, R. J.: *Adaptive execution in complex dynamic worlds*. Technischer Bericht, PhD thesis, Yale University, 1989.
- [Fix04] FIX, J.: *Emotionale Agenten*. Diplomarbeit, Universität Hamburg, FB Informatik, Hamburg, 2004.
- [FKT89] FRIJDA, N. H., P. KUIPERS und E. TERSCHURE: *Relations among emotion, appraisal, and emotional action readiness*. Journal of Personality and Social Psychology, 57:212–228, 1989.
- [Fla03] FLAM, H.: *Soziologie der Emotion*. Konstanz, UVK, 2003.
- [FM93] FRIJDA, N. H. und D. MOFFAT: *A model of emotions and emotion communication*. In: *Proceedings of RO-MAN 1993: 2nd IEEE International Workshop on Robot and Human Communication*, 1993.
- [FM94] FRIJDA, N. H. und D. MOFFAT: *Modeling emotion*. Cognitive Studies, 1(2):5–15, 1994.
- [FM06] FIX, J. und D. MOLDT: *Modelling Emotions in Multi-Agent Systems based on Petri-Net Modelling Technique*. In: REICHARDT, D., P. LEVI und J.-J. CHR MEYER (Herausgeber): *Proceedings of the 1st Workshop on Emotion and Computing - Current Research and Future Impact. Workshop at KI 2006, Bremen, Germany*, Seiten 19–22, University of Bremen, 2006.
- [FM09] FIX, J. und D. MOLDT: *A Reference Architecture for Modelling of Emotional Agent Systems*. In: BRAUBACH, LARS, WIEBKE VAN DER HOEK, PAOLO PETTA und ALEXANDER POKAHR (Herausgeber): *Multiagent System Technologies. 7th German Conference, MATES 2009, Hamburg, Germany, September 9-11, 2009. Proceedings*, Band 5774 der Reihe LNAI, Seiten 189–194. Berlin: Springer Verlag, 2009.
- [Fra88] FRANK, R.: *Passions within Reason: The Strategic Role of the Emotions*. W. W. Norton, New York, 1988.
- [Fri86] FRIJDA, N. H.: *The Emotions*. Cambridge: Cambridge University Press, 1986.
- [Fri94] FRIJDA, N. H.: *Emotions Require Cognitions, Even If Simple Ones*. In: EKMAN, P. und R.J. DAVIDSON (Herausgeber): *The Nature of Emotion*, Seiten 197–202. New York: Oxford University Press, 1994.
- [Fri07] FRIJDA, N.H.: *The laws of emotion*. Mahwah: Erlbaum., 2007.

- [Fri09] FRIJDA, N.H.: *Emotion experience and its varieties*. Emotion Review, 1:261–271, 2009.
- [FS87] FRIJDA, N. H. und J. SWAGERMAN: *Can computers feel? Theory and design of an emotional system*. Cognition and Emotion, 1:235–258, 1987.
- [FU04] FEHR, E. und FISCHBACHER U.: *Social Norms and Human Cooperation*. Trends in Cognitive Science, 8(4):185–190, 2004.
- [Gas93] GASSER, L.: *Social conceptions of knowledge and action: Dai foundations and open system semantics*. Artificial Intelligence, 47, Foundations of Artificial Intelligence:107–138, 1993.
- [Gen87] GENESERETH, M. R. UND NILSSON, N.: *Logical foundations of artificial intelligence*. Morgan Kaufmann, San Mateo, CA, 1987.
- [GI89] GEORGEFF, M. P. und F. F. INGRAND: *Decision-making in an embedded reasoning system*. In: *Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, Seiten 972–978, Detroit, MI, 1989.
- [GL87] GEORGEFF, M. P. und A. L. LANSKY: *Reactive reasoning and planning*. In: *Proceedings of the 6th National Conference on Artificial Intelligence (AAAI-87)*, Seiten 677–682, Seattle, WA, 1987.
- [GL00] GMYTRASIEWICZ, P. und C. LISETTI: *Using Decision Theory to Formalize Emotions for Multi-Agent Systems*. In: *In Proceedings of the 2nd ICMA’00 Workshop on Game Theoretic and Decision Theoretic Agents*, Boston, 2000.
- [GM01] GRATCH, J. und S. MARSELLA: *Tears and Fears: Modeling emotions and emotional behaviours in synthetic agents*. In: *Proceedings of the 5th International conference on Autonomous Agents*, Seiten 278–285, 2001.
- [GM03] GRATCH, J. und S. MARSELLA: *Fight the Way you Train: The role and limits of Emotions in training for combat*. The Brown Journal of World Affairs, 10(1), 2003.
- [GM04] GRATCH, J. und S. MARSELLA: *A Domain-independent Framework for Modeling Emotion*. Cognitive Systems Research, 5(4):269–306, 2004.
- [GM05] GRATCH, J. und S. MARSELLA: *Lessons from Emotion Psychology for the Design of Lifelike Characters*. Applied Artif. Intelligence, 19 (3/4):215–233, 2005.
- [GM09a] GRATCH, J. und S. MARSELLA: *Assessing the validity of appraisal-based models of emotion*. In: *International Conference on Affective Computing and Intelligent Interaction*, Amsterdam, IEEE, 2009.
- [GM09b] GRATCH, J. und S. MARSELLA: *Modeling the Antecedents and Consequences of Emotion*. Journal of Cognitive Systems Research, 10(1):1–5, 2009.

Literaturverzeichnis

- [GMP02] GIUNCHIGLIA, F., J. MYLOPOULOS und A. PERINI: *The Tropos Software Development Methodology: Processes, Models and Diagrams*. In: *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems*. ACM Press, 2002.
- [Gof56] GOFFMAN, E.: *Embarrassment and Social Organization*. American Journal of Sociology, 62:264–271, 1956.
- [Gol95] GOLEMAN, D.: *Emotional Intelligence*. Bantam Books, New York, 1995.
- [Gra00] GRATCH, J.: *Émile: Marshalling Passions in Training and Education*. In: *Proceedings of the 4th International Conference on Intelligent Agents*, Barcelona, Spain, 2000.
- [GV03] GIRAULT, C. und R. VALK: *Petri Nets for Systems Engineering - A Guide to Modeling, Verification, and Applications*. Springer, 2003.
- [Had96] HADDADI, A.: *Communication and Cooperation in Agent Systems*. LNAI, 1056, 1996.
- [Hew86] HEWITT, C.: *Offices are open systems*. ACM Transactions on Office Information Systems, 4(3):271–287, 1986.
- [Hew10] HEWELT, M.: *UnitEd - Einheitentheoretisches Rahmenwerk zur Modellierung von diskreten, adaptiven Systemen*. Diplomarbeit, Universität Hamburg, FB Informatik, 2010.
- [HJW07] HARMON-JONES, E. und P. WINKIELMAN: *A brief overview of social neuroscience*. In: *Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior*, Seiten 3–11. Guilford Publications: New York, 2007.
- [Hoc79] HOCHSCHILD, A.: *Emotion Work, Feeling Rules, and Social Structure*. American Journal of Sociology, 85(3):551–575, 1979.
- [HP98] HEALEY, J. und R. W. PICARD: *StartleCam: A Cybernetic Wearable Camera*, 1998.
- [HS99] HUHNS, M. N. und L. M. STEPHENS: *Multiagent Systems and Societies of Agents*. In: WEISS, G. (Herausgeber): *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, Seiten 79–120. Cambridge, MA: MIT Press, 1999.
- [Hub99] HUBER, M.: *JAM: A BDI-Theoretic Mobile Agent Architecture*. In: ETZIONI, O., J.P. MÜLLER und J. BRADSHAW (Herausgeber): *Proceedings of the 3rd International Conference on Autonomous Agents (Agents 99)*, Seiten 236–243. Seattle, WA: ACM Press, 1999.

- [IGG99] IGLESIAS, C. A., M. GARIJO und J. C. GONZALES: *A Survey of Agent-oriented Methodologies*. In: *Proceedings of the 5th International Workshop on Agent Theories, Architectures, and Languages (ATAL'98)*. Berlin: Springer, 1999.
- [Iza91] IZARD, C. E.: *The Psychology of Emotions*. Plenum Press, New York, 1991.
- [Jen92] JENSEN, K.: *Colour Petri nets, Basic Methods, Analysis Methods and Practical Use*. EATCS monographs on theoretical computer science. Springer, Berlin, 1992.
- [JR98] JOHNSON, W. L. und J. RICKEL: *Steve, An animated pedagogical agent for procedural training in virtual environments*. SIGART Bulletin, 8:16–21, 1998.
- [JRJR01] J., SCHEIRER, FERNANDEZ R., KLEIN J. und PICARD R.W.: *Frustrating the User on Purpose: A Step Toward Building an Affective Computer*. Interaction with Computers, 2001.
- [JSW99] JENNINGS, N. R., K. SYCARA und M. WOOLDRIDGE: *A Roadmap of Agent Research and Development*. Autonomous Agents and Multi-Agent Systems, 1(1):7–38, 1999.
- [JV80] JANTZEN, M. und R. VALK: *Formal Properties of Place/Transition Nets*. In: BRAUER, W. (Herausgeber): *Lecture Notes in Computer Science: Net Theory and Applications, Proc. of the Advanced Course on General Net Theory of Processes and Systems, Hamburg, 1979*, Band 84, Seiten 165–212, Berlin, Heidelberg, New York, 1980. Springer.
- [JW98] JENNING, N.R und M. J. WOOLDRIDGE: *Applications of intelligent agents*. Springer, 1998.
- [Kem78] KEMPER, T.D.: *A Social Interactional Theory of Emotions*. New York: Wiley, 1978.
- [Kem84] KEMPER, T.D.: *Power, status and emotions: a sociological contribution to a psycho-physiological domain*. In: SCHERER K., EKMAN P. (Herausgeber): *Approaches to Emotion*, Seiten 369–384. Erlbaum, Hillsdale, NJ, 1984.
- [KGR96] KINNY, D., M. GEORGEFF und A. RAO: *A methodology and modelling technique for systems of BDI agents*. In: VELDE, W. VAN DE und J.W. PERRAM (Herausgeber): *Agents Breaking Away: Proceedings of the Seventh European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, Band 1038 der Reihe LNAI, Seiten 151 – 164, Berlin, 1996. Springer.
- [KH99] KELTNER, D. und J. HAIDT: *Social functions of emotions at four levels of analysis*. Cognition and Emotion, 13(5):505–521, 1999.

Literaturverzeichnis

- [Kir97] KIRSCH, D.: *The Sentic Mouse: Developing a tool for Measuring Emotional Valence*. Technischer Bericht, MIT Media Laboratory Perceptual Computing Section, 1997.
- [KK81] KLEINGINNA, P. R. und A. M. KLEINGINNA: *A categorized list of emotion definitions, with suggestions for a consensual definition*. Motivation and Emotion, 5:345–379, 1981.
- [KLL⁺00] KUMMER, O., A. LAUE, M. LIEDTKE, D. MOLDT und HEIKO RÖLKE: *Höhere Petrinetze zur kompakten Modellierung und Implementierung von Verhalten*. In: GIESE, HOLGER und STEPHAN PHILIPPI (Herausgeber): *Visuelle Verhaltensmodellierung verteilter und nebenläufiger Software-Systeme, 8. Workshop des Arbeitskreises GROOM*, Techreport 24/00-I, Seiten 27–32. Universität Münster, 2000.
- [KLR03] KÖHLER, M., R. LANGER und H. RÖLKE: *Sozionik: Modellierung soziologischer Theorie*, Kapitel Sozionisches Modellieren mit Petrinetzen. Ein kommentiertes Protokoll. Wirtschaft – Arbeit – Technik. Lit-Verlag, Münster, 2003.
- [KMP01] KAPOOR, A., S. MOTA und R.W. PICARD: *Toward a Learning Companion that Recognizes Affect*. MIT Media Lab Tech Report No 543, Proceedings from Emotional and Intelligent II, AAAI FallSymposium, 2001.
- [KMR01a] KÖHLER, M., D. MOLDT und H. RÖLKE: *Einheitliche Modellierung von Agenten und Agentensystemen mit Referenznetzen*. In: JABLONSKI, S., S. KIRN, M. PLAHA, E. SINZ, A. ULBRICH-VOM ENDE und G. WEISS (Herausgeber): *Tagungsunterlagen: Verteilte Informationssysteme auf der Grundlage von Objekten, Komponenten und Agenten (vertIS 2001)*, Universität Bamberg, 4-5. Oktober 2001, Seiten 3–20, oct 2001.
- [KMR01b] KÖHLER, M., D. MOLDT und H. RÖLKE: *Modelling the Structure and Behaviour of Petri Net Agents*. In: COLOM, J.M. und M. KOUTNY (Herausgeber): *Proceedings of the 22nd Conference on Application and Theory of Petri Nets*, LNCS, Seiten 224–241. Springer, 2001.
- [KMRV05] KÖHLER, M., D. MOLDT, H. RÖLKE und R. VALK: *Linking Micro and Macro Description of Scalable Social Systems using Reference Nets*. In: FISCHER, K., M. FLORIAN und T. MALSCH (Herausgeber): *Socionics: Sociability of Complex Social Systems*, Band 3413 der Reihe LNAI, Seiten 51–67. Springer, 2005.
- [KNB⁺03] KREMPELS, K.-H., J. NIMIS, L. BRAUBACH, A. POKAHR, R. HERRLER und T. SCHOLZ: *Entwicklung intelligenter Multi-Multiagentensysteme - Werkzeugunterstützung, Lösungen und offene Fragen*. In: *INFORMATIK 2003 Innovative Informatikanwendungen. GI-Edition Lecture Notes in Informatics*, Seiten 31–46. Köllen Verlag, 2003.

- [KR01] KÖHLER, M. und H. RÖLKE: *Towards a Unified Approach for Modeling and Verification of Multi Agent Systems*. In: MOLDT, DANIEL (Herausgeber): *Proceedings of the Workshop on Modelling, object, components, and agents (MOCA'01)*, Seiten 85–104. Universität Aarhus, 2001.
- [KR02] KÖHLER, M. und H. RÖLKE: *Modelling of social theories*. In: LINDEMANN, G., DANIEL MOLDT, M. PAOLUCCI und B. YU (Herausgeber): *International Workshop on Regulated Agent-Based Social Systems: Theories and Applications (RASTA'02) as part of AAMAS 2002*, 2002.
- [KTB03] KRAMER, N. C., B. TIETZ und G. BENTE: *Effects of embodied interface agents and their gestural activity*. In: *Intelligent Virtual Agents, Kloster Irsee, Germany*, 2003.
- [Kum98] KUMMER, O.: *Simulating Synchronous Channels and Net Instances*. In: DESEL, JÖRG, PETER KEMPER, EKKART KINDLER und ANDREAS OBERWEIS (Herausgeber): *Forschungsbericht Nr. 694: 5. Workshop Algorithmen und Werkzeuge für Petrinetze*, Forschungsbericht Nr. 694, Seiten 73–78. Fachbereich Informatik, Universität Dortmund, 1998.
- [Kum00] KUMMER, O.: *Undecidability in Object-Oriented Petri Nets*. Petri Net Newsletter, 59:18–23, 2000.
- [Kum02] KUMMER, O.: *Referenznetze*. Logos Verlag, Berlin, 2002.
- [KWD⁺04] KUMMER, O., F. WIENBERG, M. DUVIGNEAU, J. SCHUMACHER, M. KÖHLER, D. MOLDT, H. RÖLKE und R. VALK: *An Extensible Editor and Simulation Engine for Petri Nets: Renew*. In: CORTADELLA, JORDI und WOLFGANG REISIG (Herausgeber): *Applications and Theory of Petri Nets 2004. 25th International Conference, ICATPN 2004, Bologna, Italy, June 2004. Proceedings*, Band 3099 der Reihe *Lecture Notes in Computer Science*, Seiten 484–493, Heidelberg, jun 2004. Springer.
- [KWDC09] KUMMER, OLAF, FRANK WIENBERG, MICHAEL DUVIGNEAU und LAWRENCE CABAC: *Renew – User Guide (Release 2.2)*. University of Hamburg, Faculty of Informatics, Theoretical Foundations Group, Hamburg, aug 2009. letzter Zugriff: 07.01.2012.
- [LA08] LEICHTENSTERN, K. und E. ANDRÉ: *User-Centred Development of Mobile Interfaces to a Pervasive Computing Environment*. In: *Proceedings of the 1st International Conference on Advances in Computer-Human Interaction (ACHI 2008)*, IEEE, 2008.
- [Lab96] LABROU, Y.: *Semantics for an Agent Communication Language*. PhD Theses, University of Maryland, Baltimore Country, 1996.

Literaturverzeichnis

- [Lab01] LABROU, Y.: *Standardizing Agent Communication*. In: LUCK, M. (Herausgeber): *Multi Agent Systems and Applications*, Band 2086, Seiten 74–97. Springer, 2001.
- [Lan05] LANGER, R.: *Anerkennung und Vermögen - Eine Analyse von Selbstorganisationsprozessen in Bildungsinstitutionen.*, Band 1 der Reihe *Schriftenreihe Pädagogik im Widerspruch*. Monsenstein und Vannerdat, 2005.
- [LAO70] LAZARUS, R. S., J. R. AVERILL und E. M. OPTON: *Toward a cognitive theory of emotion*. In: ARNOLD, M.B. (Herausgeber): *Feelings and emotions*, Seiten 207–232, New York, 1970. Academic Press.
- [Laz91] LAZARUS, R.S.: *Emotion and adaptation*. Oxford University Press, New York, 1991.
- [LB02] LISETTI, C. und N. BIANCHI: *Modeling Multimodal Expression of User's Affective Subjective Experience*. *User Modeling and User-Adapted Interaction*, 12(1):49–84, 2002.
- [LBS99] LESTER, J. C., STONE B.A. und G.D. STELLING: *Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments, User Model*. *User-Adapted Instruction*, 9:1–44, 1999.
- [LeD84] LEDOUX, J.E.: *Cognition and Emotion: Processing Functions and Brain Systems*. In: GAZZANIGA, M. S. (Herausgeber): *Handbook of Cognitive Neuroscience*, Seiten 357–368, New York, 1984. Plenum Press.
- [LeD96a] LEDOUX, J.E.: *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Simon and Schuster, New York, 1996.
- [LeD96b] LEDOUX, J.E.: *The Nature of Emotion*. In: EKMAN, P. und R.J. DAVIDSON (Herausgeber): *Handbook of Cognitive Neuroscience*, Seiten 216–223, New York, 1996. Oxford University Press.
- [LeD98] LEDOUX, J.E.: *Das Netz der Gefühle. Wie Emotionen entstehen*. Carl Hanser Verlag, Wien, 1998.
- [LeD00] LEDOUX, J.E.: *Emotion Circuits in the Brain*. *Annual Review of Neuroscience*, 23:155–184, 2000.
- [Lev99] LEVENSON, R. W.: *The intrapersonal functions of emotion*. *Cognition and Emotion*, 13:481–504, 1999.
- [Lev03] LEVENSON, R. W.: *Autonomic specificity and emotion*. In: *Handbook of Affective Sciences*, Seiten 212–224, New York, 2003. Oxford University Press.
- [Lim01] LIMITED, EMORPHIA: *FIPA-OS V2.1.0 Distribution Notes*. Technischer Bericht, FIPA, 2001.

- [Lin01] LIND, J.: *Iterative Software Engineering for Multiagent Systems: the MASSIVE Method*. Lecture Notes in Computer Science and Lecture Notes in Artificial Intelligence, 1994:Berlin: Springer, 2001.
- [LS87] LEVENTHAL, H. und K. R. SCHERER: *The relationship of emotion to cognition: A functional approach to a semantic controversy*. Cognition and Emotion, 1, Seiten 3–28, 1987.
- [Mac49] MACLEAN, P. D.: *Psychosomatic disease and the “visceral brain“*. Recent developments bearing on the Papez theory of emotion. Psychosom Med, 11:338–353, 1949.
- [Mac96] MACLEAN, P. D.: *Limbic system*. In: BEAUMONT, J. G., P. A. KENEALY und J. C. ROGERS (Herausgeber): *The Blackwell dictionary of neuropsychology*, Oxford, 1996. Basil Blackwell.
- [Mae91] MAES, P.: *The agent network architecture (ANA)*. SIGART Bulletin, 2(4):115–120, 1991.
- [Mal98] MALSCH, T.: *Die Provokation der “artificial societies“*. In: MALSCH, T. (Herausgeber): *Sozionik. Soziologische Ansichten über künstliche Sozialität*, Berlin, 1998. Edition Sigma.
- [Man83] MANDL, H. UND EULER, H. A.: *Begriffsbestimmungen*. In: H. MANDL, H. A. EULER UND (Herausgeber): *Emotionspsychologie. Ein Handbuch in Schlüsselbegriffen*, Seiten 5–11, München, 1983. Urban und Schwarzenberg.
- [MAS02] MATEAS, M. und A. ANDREW STERN: *A Behavior Language for Story-based Believable Agents*. IEEE Intelligent Systems, 17 (4):39–47, 2002.
- [MB93] MCFARLAND, D. und T. BÖSSER: *Intelligent Behavior in Animals and Robots*. MIT Press, Cambridge, MA, 1993.
- [McD69] MCDUGALL, W.: *An outline of psychology (4th ed.)*. Methuen, London, 1928/1969.
- [MF00] MOFFAT, D. und N. H. FRIJDA: *Functional models of emotion*. In: HATANO, G., N. OKADA und H. TANABE (Herausgeber): *Affective Minds: Proceedings of the 13th Toyota Conference*, Shizuoka, Japan, 2000. Elsevier Health Sciences.
- [MFP93] MOFFAT, D., N. H. FRIJDA und R.H. PHAF: *Analysis of a computer model of emotions*. Prospects for artificial intelligence, Seiten 219–228, 1993.
- [MFvLvS06] MOLDT, D., J. FIX, R. VON LÜDE und C. VON SCHEVE: *Emotion als theorieleitende Kategorie in Soziologie und Informatik: Zur emotionsbasierten Modellierung von Struktur dynamiken in künstlichen und natürlichen Gesellschaften*. In: *Post Proceedings: Ordnungsbildung und Erkenntnisprozesse (OrdEr) 2005*, Seiten 103–116, 2006.

Literaturverzeichnis

- [MG01] MARSELLA, S. und J. GRATCH: *Modeling the interplay of plans and emotions in multi-agent simulations*. In: *Cognitive Science Society*, Edinburgh, Scotland, 2001.
- [MG02] MARSELLA, S. und J. GRATCH: *A Step toward Irrationality: Using emotion to change belief*. In: *First International Joint Conference on Autonomous Agents and Multi-agent Systems*, Bologna, Italy, 2002.
- [MG03] MARSELLA, S. und J. GRATCH: *Modeling coping behaviors in virtual humans: Don't worry, be happy*. In: *Second International Joint Conference on Autonomous Agents and Multi-agent Systems*, Melbourne, Australia, 2003.
- [MG04] MAO, W. und J. GRATCH: *Social judgment in multiagent interactions*. In: *Third International Joint Conference on Autonomous Agents and Multiagent Systems*, 2004.
- [MG09] MARSELLA, S. und J. GRATCH: *EMA: A Model of Emotional Dynamics*. *Journal of Cognitive Systems Research*, 10(1):70–90, 2009.
- [MGP12] MARSELLA, S., J. GRATCH und P. PETTA: *Computational Models of Emotion*. In: SCHERER, K.R., T. BÄNZIGER und E. ROESCH (Herausgeber): *A blueprint for an affectively competent agent: Cross-fertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing*, Band in press der Reihe *Lecture Notes in Computer Science*. Oxford: Oxford University Press, 2012.
- [Min86] MINSKY, M.: *The Society of Mind*. Simon and Schuster, New York, Cambridge, England, 1986.
- [MJL00] MARSELLA, S., W. L. JOHNSON und C. LABORE: *Interactive pedagogical drama*. In: *Fourth International Conference on Autonomous Agents*, Montreal, Canada, 2000.
- [ML09] MARINIER, R. P. und J. E. LAIRD: *A computational unification of cognitive behavior and emotion*. *Cognitive Systems Research*, 10(1):48–69, 2009.
- [MMSL01] MORENO, R., R. E. MAYER, H. SPIRES und J. LESTER: *The case for social agency incomputer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents?* *Cognition and Instruction*, 19:177–214, 2001.
- [Moh05] MOHAMAD, Y.: *Integration von emotionaler Intelligenz in Interface-Agenten am Beispiel einer Trainingssoftware für lernbehinderte Kinder*. Doktorarbeit, Fraunhofer series in information and communication technology, 2005, 4, 2005.

- [Mol05] MOLDT, D.: *Petrinetze als Denkzeug*. In: FARWER, BERNDT und DANIEL MOLDT (Herausgeber): *Object Petri Nets, Processes, and Object Calculi*, Nummer FBI-HH-B-265/05, Seiten 51–70, 2005.
- [Mol06] MOLDT, D.: *Paose: A way to develop distributed software systems based on Petri nets and agents*. Proceedings of The Fourth International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems (MSVVEIS'06), 2006.
- [MOO04] MOLDT, D., S. OFFERMANN und J. ORTMANN: *A Proposal for Petri Net Based Web Service Application Modeling*. In: KOCH, N., P. FRATERNALI und M. WIRSING (Herausgeber): *Engineering - 4th International Conference, ICWE 2004*, Band 3140 der Reihe *Lecture Notes in Computer Science*, Seiten 93–97. Springer, 2004.
- [Mor89] MORRIS, W. N.: *Mood. The frame of mind*. Springer, New York, 1989.
- [MPS04] MARSELLA, S., D. PYNADATH und READ S.: *PsychSim: Agent-based modeling of social interactions and influence*. In: *International Conference on Cognitive Modeling*, 2004.
- [MSR01] MEYER, W.-U., A. SCHÜTZWOHL und R. REISENZEIN: *Einführung in die Emotionspsychologie (Band 1, 2. Auflage: Die Emotionstheorien von Watson, James und Schachter)*. Bern: Huber, 2001.
- [MV93] MACKENTHUN, R. und R. VALK: *Nachweisbar sichere Software durch Modellieren und Verifizieren mit Hilfe von Petrinetzen*. Technischer Bericht FBI-HH-M-222, Universität Hamburg, Fachbereich Informatik, 1993.
- [MVDT04] MOHAMAD, Y., C. VELASCO, S. DAMM und H. TEBARTH: *Cognitive Training with Animated Pedagogical Agents (TAPA) in Children with Learning Disabilities*. In: *ICCHP 2004 9th International Conference on Computers Helping People with Special Needs*, Universite Pierre et MarieCurie, Paris, France, 2004.
- [MvS00] MOLDT, D. und C. VON SCHEVE: *Soziologisch adäquate Modellierung emotionaler Agenten*. In: *Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen (ABIS) 2000*, Seiten 117–131, nstitut für Semantische Informationsverarbeitung, Universität Osnabrück, 2000.
- [MvS01a] MOLDT, D. und C. VON SCHEVE: *Emotional Actions for Emotional Agents*. In: *Proceedings of the AISB'01 Symposium on Emotion, Cognition, and Affective Computing. 21.-24. März 2001, York/UK*, 2001.
- [MvS01b] MOLDT, D. und C. VON SCHEVE: *Emotions and Multimodal Interface-Agents: A Sociological View*. In: OBERQUELLE, HORST (Herausgeber):

Literaturverzeichnis

- Mensch und Computer 2001*, 05.–08. März in Bad Honnef (Bonn), Nummer 55 in *Berichte des German Chapter of the ACM*, Stuttgart, 2001. Teubner.
- [MvS02] MOLDT, D. und C. VON SCHEVE: *Emotions in Hybrid Social Aggregates*. In: HERCZEG, MICHAEL, WOLFGANG PRINZ und HORST OBERQUELLE (Herausgeber): *Mensch & Computer 2002 – Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten*, Band 56 der Reihe *Berichte der German Chapter of the ACM*, Seiten 343–352, Stuttgart, Leipzig, Wiesbaden, aug 2002. 2. fachübergreifende Konferenz, Hamburg, 2. bis 5. September 2002, Teubner.
- [MW97] MOLDT, D. und F. WIENBERG: *Multi-Agent-Systems based on Coloured Petri Nets*. In: *Application and Theory of Petri Nets 1997*, Lecture Notes in Computer Science, Seiten 82–101, Berlin Heidelberg New York, 1997. Springer Verlag.
- [New80] NEWELL, A.: *Physical symbol systems*. *Cognition*, 4:135–143, 1980.
- [NL95] NORMAN, T. J. und D. LONG: *Goal Creation in Motivated Agents*. In: *Intelligent Agents: Theories, Architectures, and Languages*, LNAI 890, Seiten 277–290. Springer, 1995.
- [NS72] NEWELL, A. und H. A. SIMON: *Human Problem Solving*. Englewood Cliffs, Prentice Hall, 1972.
- [OCC88] ORTONY, A., G. L. CLORE und A. COLLINS: *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, 1988.
- [OEM00] OTTO, J., H. A. EULER und H. MANDL: *Begriffsbestimmungen*. In: OTTO J., EULER H. A. UND MANDL H. (Herausgeber): *Handbuch Emotionspsychologie*, Seiten 11 – 18. Weinheim: Beltz, PsychologieVerlagsUnion., 2000.
- [OJ96] OATLEY, K. und J. M. JENKINS: *Understanding Emotions*. Blackwell, 1996.
- [OJL87] OATLEY, K. und P. N. JOHNSON-LAIRD: *Towards a Cognitive Theory of Emotions*. *Cognition and Emotion*, 1 (1):29–50, 1987.
- [OPB01] ODELL, J., H. V. D. PARUNAK und B. BAUER: *Representing Agent Interaction Protocols in UML*. In: *Proceedings of the First International Workshop on Agent-oriented SoftwareEngineering (AOSE2000)*. Berlin: Springer, 2001.
- [OS03] OLIVEIRA, E. und L. SARMENTO: *Emotional advantage for adaptability and autonomy*. In: *AAMAS*, Seiten 305–312. ACM, 2003.
- [OT90] ORTONY, A. und T. J. TURNER: *What's basic about basic emotions?* *Psychological Review*, 97:315–331, 1990.

- [Pan98] PANKSEPP, J.: *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press, New York, 1998.
- [Pan06] PANKSEPP, J.: *Emotional endophenotypes in evolutionary psychiatry*. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 30:774–784, 2006.
- [Pap37] PAPEZ, J. W.: *A proposed mechanism of emotion*. Neuropsychiatry: Clin. Neuroscience, 7(1):103–112, 1995 (1937).
- [Par51] PARSONS, T.: *The Social System*. Free Press, New York, 1951.
- [Par94] PARSONS, T.: *Aktor, Situation und normative Muster: Ein Essay zur Theorie sozialen Handelns*. Suhrkamp, Frankfurt a. M., 1994.
- [PBL05] POKAHR, A., L. BRAUBACH und W. LAMERSDORF: *A Goal Deliberation Strategy for BDI Agent Systems*. In: EYMANN, TORSTEN, FRANZISKA KLÜGL, WINFRIED LAMERSDORF, MATTHIAS KLUSCH und MICHAEL HUHNS (Herausgeber): *Multiagent System Technologies*, Band 3550 der Reihe *Lecture Notes in Computer Science*, Seiten 82–93. Springer Berlin / Heidelberg, 2005.
- [PCLC07] PIZZI, D., F. CHARLES, J.-L. LUGRIN und M. CAVAZZA: *Interactive Storytelling with Literary Feelings*. In: PAIVA, A., R. PRADA und R. W. PICARD (Herausgeber): *Affective Computing and Intelligent Interaction*, Band 4738, Seiten 630–641. Berlin: Springer, 2007.
- [Pes08] PESSOA, L.: *On the relationship between emotion and cognition*. Nature Reviews. Neuroscience, 9:148–158, 2008.
- [Pet62] PETRI, C. A.: *Kommunikation mit Automaten*. Dissertation, Schriften des IIM 2, Rheinisch-Westfälisches Institut für Instrumentelle Mathematik an der Universität Bonn, Bonn, 1962.
- [Pet03] PETTA, P.: *The Role of Emotions in a Tractable Architecture for Situated Cognizers*. In: R., TRAPPL, PETTA P. und PAYR S. (Herausgeber): *Emotions in Humans and Artifacts*, Seiten 250–287, Cambridge, Massachusetts, 2003. The MIT Press.
- [Pfe82] PFEIFER, R.: *Cognition and emotion: an information processing approach*. CIP Working Paper 436, Carnegie-Mellon University, 1982.
- [Pfe88] PFEIFER, R.: *Artificial intelligence models of emotion*. In: HAMILTON, V., G. BOWER und N. FRIJDA (Herausgeber): *Cognitive Perspectives on Emotion and Motivation. Proceedings of the NATO Advanced Research Workshop on Cognitive Science Perspectives on Emotion, Motivation, and Cognition*, volume 44 of NATO ASI series, Seiten 287–320, Dordrecht. Kluwer Academic, 1988.

Literaturverzeichnis

- [PG08] PICARD, R. W. und M. GOODWIN: *Developing Innovative Technology for Future Personalized Autism Research and Treatment*. Autism Advocate, 50(1):32–39, 2008.
- [PH97] PICARD, R. W. und J. HEALEY: *Affective Wearables*, 1997.
- [Pic97] PICARD, R. W.: *Affective Computing*. MIT Press, Cambridge, 1997.
- [Pic03] PICARD, R. W.: *What Does it Mean for a Computer to Have Emotions?* In: R. TRAPPL, P. PETTA und S. PAYR (Herausgeber): *Emotions in Humans and Artifacts*, Cambridge/Massachusetts, 2003. MIT Press.
- [Plu94] PLUTCHIK, R.: *The psychology and biology of emotion*. Harper Collins College Publishers, New York, 1994.
- [PMMI03] PRENDINGER, H., S. MAYER, J. MORI und M. ISHIZUKA: *Persona effect revisited*. In: *Intelligent Virtual Agents*, Kloster Irsee, 2003.
- [PN85] PFEIFER, R. und D. W. NICHOLAS: *Toward computational models of emotion*. In: STEELS, L. und J.A. CAMPBELL (Herausgeber): *Progress in Artificial Intelligence*, Chichester, U.K., Ellis Horwood, 1985.
- [POM08] PEREIRA, D., E. OLIVEIRA und N. MOREIRA: *Formal Modelling of Emotions in BDI Agents*. In: SADRI, F. und K. SATOH (Herausgeber): *Computational Logic in Multi-Agent Systems*, Band 5056/2008 der Reihe *Lecture Notes in Computer Science*, Seiten 62–81, Berlin, Heidelberg, 2008. LNAI, Springer.
- [PPB⁺04] PICARD, R. W., S. PAPERT, W. BENDER, B. BLUMBERG, C. BREAZEAL, D. CAVALLO, T. MACHOVER, M. RESNICK, D. ROY und C. STROHECKER: *Affective Learning—A Manifesto*. BT Technical Journal, 22(4):253–269, 2004.
- [PPP07] PAIVA, A., R. PRADA und R. W. PICARD (Herausgeber): *Affective Computing and Intelligent Interaction*, Band 4738. Berlin: Springer, 2007.
- [PS96] PEKRUN, R. und U. SCHIEFELE: *Emotions- und motivationspsychologische Bedingungen der Lernleistung*. In: WEINERT, F.E. (Herausgeber): *Psychologie des Lernens und der Instruktion*, Enzyklopädie der Psychologie, Serie Pädagogische Psychologie, Bd. 2, Seiten 153 – 180, Göttingen: Hogrefe., 1996.
- [PT06] PETTA, P. und R. TRAPPL: *Emotions and Agents*. In: LUCK, M., V. MARIK, O. STEPANKOVA und R. TRAPPL (Herausgeber): *Multi-Agent Systems and Applications*, Band 2086 der Reihe *Lecture Notes in Computer Science*, Seiten 301–316. Springer Berlin / Heidelberg, 2006. letzter Zugriff: 10.01.2012.
- [PVH01] PICARD, R. W., E. VYZAS und J. HEALEY: *Toward Machine Emotional Intelligence: Analysis of Affective Physiological State*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(10):1175–1191, 2001.

- [R.10] R., ADOLPHS: *Emotion*. Current Biology, 20(13):549–552, 2010.
- [Rao91] RAO, A. UND GEORGEFF, M.: *Modeling rational agents within a BDI-architecture*. In: R. FIKES, E. SANDEWALL (Herausgeber): *Proceeding of Knowledge Representation and Reasoning*, Seiten 473–484, San Mateo, CA, 1991. Morgan Kaufmann.
- [Rao92] RAO, A. UND GEORGEFF, M.: *An abstract architecture for rational agents*. In: C. RICH, W. SWARTOUT, B. NEBEL (Herausgeber): *Proceeding of Knowledge Representation and Reasoning*, Seiten 439–449. Morgan Kaufmann, 1992.
- [Rao93] RAO, A. UND GEORGEFF, M.: *A model-theoretic approach to the verification of situated reasoning systems*. In: C. RICH, W. SWARTOUT, B. NEBEL (Herausgeber): *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI-93)*, Seiten 318–324, Chambery, France, 1993.
- [Rao95] RAO, A. UND GEORGEFF, M.: *BDI-agents: from theory to practice*. In: *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, San Francisco, 1995.
- [RB92] REILLY, W. S. und J. BATES: *Building Emotional Agents*. Technical Report CMU-CS-92-143, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1992.
- [RE09] ROTHERMUND, K. und A. EDER: *Emotion und Handeln*. In: V., BRANDSTÄTTER und J.H. OTTO (Herausgeber): *Handbuch der Allgemeinen Psychologie: Motivation und Emotion*, Seiten 675–685, Göttingen: Hogrefe, 2009.
- [Ree91] REEVES, J. F.: *Computational morality: A process model of belief conflict and resolution for story understanding*. Technischer Bericht, Technical Report UCLA-AI-91-05, UCLA Artificial Intelligence Laboratory, 1991.
- [Ree10] REESE, C.: *Prozess-Infrastruktur für Agentenanwendungen*, Band 3 der Reihe *Agent Technology – Theory and Applications*. Logos Verlag, Berlin, 2010.
- [Rei96] REILLY, W. S.: *Believable Social and Emotional Agents*. PhD Thesis CMU-CS-96-138, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, May 1996.
- [Rei00a] REISENZEIN, R.: *Einschätzungstheoretische Ansätze in der Emotionspsychologie*. In: OTTO, J.H., H.A. EULER und H. MANDL (Herausgeber): *Handbuch der Emotionspsychologie*, Seiten 117–138, Weinheim: Psychologie-VerlagsUnion, 2000.

Literaturverzeichnis

- [Rei00b] REISENZEIN, R.: *Worum geht es in der Debatte um die Basisemotionen.* In: F., FÖRSTERLING und SILNY L.-M. (Herausgeber): *Kognitive und motivationale Aspekte der Motivation*, Seiten 205–237, Göttingen, Hogrefe, 2000.
- [RHA⁺99] ROTHBAUM, B. O., L. F. HODGES, R. ALARCON, D. READY, F. SHAHAR, K. GRAAP, J. PAIR, P. HEBERT, B. WILLS und D. BALTZELL: *Virtual environment exposure therapy for PTSD vietnam veterans: A case study.* Trauma Stress, 12:263–272, 1999.
- [Röl99] RÖLKE, H.: *Modellierung und Implementation eines Multi-Agenten-Systems auf der Basis von Referenznetzen.* Diplomarbeit, FBIUniHH, 1999.
- [Röl04] RÖLKE, H.: *Modellierung von Agenten und Multiagentensystemen - Grundlagen und Anwendungen*, Band 2. Logos Verlag, Berlin, 2004.
- [RN95] RUSSEL, S. und P. NORVIG: *Artificial Intelligence: A modern approach.* Englewood Cliffs, NJ:Prentice Hall, 1995.
- [RN96] REEVES, B. und C. NASS: *The media equation: how people treat computers, television, and new media like real people and places.* Cambridge University Press, New York, NY, USA, 1996.
- [Rol05] ROLLS, E.: *Emotion Explained.* Oxford University Press, 2005.
- [Ros96] ROSEMAN, I. J.: *Appraisal determinants of discrete emotions.* Cognition and Emotion, 3:161–200, 1996.
- [Rot99] ROTH, G.: *Fühlen, Denken, Handeln: Wie das Gehirn unser Verhalten steuert.* Suhrkamp, Frankfurt am Main, 1999.
- [RS04] RETICULAR SYSTEMS, INC.: *AgentBuilder – User Guide.* Reticular Systems, Inc, 2004. letzter Zugriff: 07.01.2012.
- [Rus94] RUSSELL, J. A.: *Is there Universal Recognition of Emotion from Facial Expression ? A Review of the Cross-Cultural Studies.* Psychological Bulletin, 115:102–141, 1994.
- [Rus03] RUSSELL, J. A.: *Core affect and the psychological construction of emotion.* Psychological Review, 110(1):145–172, 2003.
- [RVC03] RYOKAI, K., C. VAUCELLE und J. CASSELL: *Virtual Peers as Partners in Storytelling and Literacy Learning.* Journal of Computer Assisted Learning, 19(2):195–208, 2003.
- [RWUS96] REISENZEIN, R., MEYER W.-U. und A. SCHÜTZWOHL: *Reactions to surprising events: A paradigm to emotion research.* In: FRIJDA, N. (Herausgeber): *Proceedings of the 9th conference of the International Society for Research on Emotions*, Seiten 292–296, Toronto, Canada:ISRE, 1996.

- [SA81] SCHMIDT-ATZERT, L.: *Emotionspsychologie*. Kohlhammer, Stuttgart, 1981.
- [SA96] SCHMIDT-ATZERT, L.: *Lehrbuch der Emotionspsychologie*. Kohlhammer, Stuttgart, 1996.
- [SC81] SLOMAN, A. und M. CROUCHER: *Why robots will have emotions*. In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI'81)*, Seiten 197–202, Vancouver. Canada, 1981.
- [Sch84] SCHERER, K. R.: *On the nature and function of emotion: A component process approach*. In: SCHERER, K.R. und P. EKMAN (Herausgeber): *Approaches to Emotion*, Seiten 293–318. Erlbaum, Hillsdale, NJ, 1984.
- [Sch88] SCHERER, K. R.: *Criteria for Emotion-Antecedent Appraisal: A Review*. In: HAMILTON, V., BOWER G. H. und N. H. FRIJDA (Herausgeber): *Cognitive Perspectives on Emotion and Motivation*, Seiten 89–126. Dordrecht: Kluwer, 1988.
- [Sch90] SCHWARZ, N.: *Feelings as information: Informational and motivational functions of affective states*. In: HIGGINS, E.T. und SORRENTINO (Herausgeber): *Handbook of motivation and cognition*, Seiten 527–561. NY: Guilford, 1990.
- [Sch93] SCHERER, K. R.: *Studying the Emotion-Antecedent Appraisal Process: An Expert System Approach*. *Cognition and Emotion*, 7 (3/4):325–355, 1993.
- [Sch99a] SCHERER, K. R.: *Appraisal theory*. *Handbook of Cognition and Emotion*. Wiley, Chichester, 1999.
- [Sch99b] SCHERER, K. R.: *On the sequential nature of appraisal processes: Indirect evidence from a recognition task*. *Cognition and Emotion*, 13 (6):763–793, 1999.
- [Sch01] SCHERER, K. R.: *Appraisal considered as a process of multi-level sequential checking*. *Appraisal processes in emotion: Theory, Methods, Research*. Oxford University Press, New York and Oxford, 2001.
- [Sch02] SCHEUTZ, M.: *Agents with or without Emotions?* In: *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference*, Seiten 89–93. AAAI Press, 2002.
- [Sch03] SCHEFF, T. J.: *Shame in Self and Society*. *Symbolic Interaction*, 26(2):239–262, 2003.
- [SE85] SMITH, C. A. und P. C. ELLSWORTH: *Patterns of Cognitive Appraisal in Emotion*. *Journal of Personality and Social Psychology*, 48:813–838, 1985.
- [SH99] SAAM, N. J. und A. HARRER: *Simulating Norms, Social Inequality, and Functional Change in Artificial Societies*. *Journal of Artificial Societies and Social Simulation*, 2(1), 1999.

Literaturverzeichnis

- [Sho79] SHOTT, S.: *Emotion and Social Life: A Symbolic Interactionist Analysis*. American Journal of Sociology, 84(4):1317–1334, 1979.
- [Sim01] SIMMEL, G.: *Zur Psychologie der Scham*. In: DAHME, H.-J. und O. RAMMSTEDT (Herausgeber): *Georg Simmel. Schriften zur Soziologie*, Seiten 151–158, Frankfurt am Main, 1986[1901]. Suhrkamp.
- [Sim67] SIMON, H. A.: *Motivational and emotional controls of cognition*. Psychological Review, 74:29–39, 1967.
- [SJO⁺02] SILVERMAN, B. G., M. JOHNS, K. O'BRIEN, R. WEAVER und J. CORNWELL: *Constructing Virtual Asymmetric Opponents from Data and Models in the Literature: Case of Crowd Rioting*. In: *11th Conference on Computer-Generated Forces and Behavior Representation*, Seiten 97–105. Orlando FL, 2002.
- [SL90] SMITH, C. A. und R. LAZARUS: *Emotion and Adaptation*. In: PERVIN, L. A. (Herausgeber): *Handbook of Personality: Theory and research*, Seiten 609–637. Guilford Press: NY, 1990.
- [Slo92] SLOMAN, A.: *Prolegomena to a Theory of Communication and Affect*. In: ORTONY, A., J. SLACK und O. STOCK (Herausgeber): *Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues*, Heidelberg, Springer, 1992.
- [Slo93] SLOMAN, A.: *The mind as a control system*. In: C., HOOKWAY und D. PETERSON (Herausgeber): *Philosophy and the Cognitive Science*, Seiten 69–110. Cambridge University Press, 1993.
- [Slo01] SLOMAN, A.: *Varieties of affect and the CogAff architecture schema*. In: *Agents and Cognition: Proceedings of the AISB'01 Symposium on Emotion, Cognition, and Affective Computing*, Seiten 39–48, York, 2001. SSAISB.
- [Slo03] SLOMAN, A.: *How many separately evolved emotionale beasts live within us?* In: TRAPPL, R., P. PETTA und PAYR S. (Herausgeber): *Emotions in Humans and Artifacts*, Seiten 35–114, Cambridge/MA, 2003. The MIT Press.
- [Slo04] SLOMAN, A.: *What are emotion theories about?* In: *Proceedings at the AAAI 2004 Spring Symposium*, Stanford University, 2004.
- [SMBB08] SLATER, S., R. MORETON, K. BUCKLEY und A. BRIDGES: *A Review of Agent Emotion Architectures*. Eludamos. Journal for Computer Game Culture, 2:203 – 214, 2008.
- [SP81] SMITH, C. A. und L. K. POPE: *Appraisal and Emotion. The Interactional Contributions of Dispositional and Situational Factors*. In: CLARK, M.S.

- (Herausgeber): *Emotion and Social Behavior*, Seiten 32–62, Newbury Park, CA, 1981. Sage.
- [SP98] STALLER, A. und P. PETTA: *Towards a Tractable Appraisal-Based Architecture for Situated Cognizers*. In: CANAMERO D., NUMAOKA C., PETTA P. (Herausgeber): *Grounding Emotions in Adaptive Systems*, Workshop Notes, Fifth International Conference of the Society for Adaptive Behaviour (SAB98), Zurich, Switzerland, 1998.
- [SP00] STALLER, A. und P. PETTA: *Introducing Emotions into the Computational Study of Social Norms*. Journal of Artificial Societies and Social Simulation, 4, 2000.
- [SRA⁺03] SANFEY, A.G., J. K. RILLING, J. A. ARONSON, L. E. NYSTROM und J. D. COHEN: *The Neural Basis of Economic Decision-Making in the Ultimatum Game*. Science, 300:1755–1758, 2003.
- [SS02] SLOMAN, A. und M. SCHEUTZ: *A framework for comparing agent architectures*. In: *Proceedings of UK Workshop on Computational Intelligence (UKCI'02)*, Birmingham, UK, 2002.
- [ST97] SHOHAM, Y. und M. TENNENHOLTZ: *On the emergence of social conventions: modeling, analysis, and simulations*. Artificial Intelligence, 94(1-2):139–166, 1997.
- [Ste02] STEHR, M.-O.: *Programming, Specification, and Interactive Theorem Proving — Towards a Unified Language based on Equational Logic, Rewriting Logic, and Type Theory*. Doktorarbeit, Computer Science Department, University of Hamburg, Germany, 2002.
- [Tho89] THOITS, P.A.: *The Sociology of Emotions*. Annual Review of Sociology, 15:317–342, 1989.
- [TM03] TELL, V. und D. MOLDT: *Ein Petrinetzsystem zur Modellierung selbstmodifizierender Petrinetze*. In: SCHMIDT und STAHL (Herausgeber): *Proceedings of the 12th Workshop on Algorithms and Tools for Petri Nets (AWPN 05)*, Seiten 36–41, 2003.
- [TPP03] TRAPPL, R., P. PETTA und S. PAYR: *Emotions in Humans and Artifacts*. MIT Press, Cambridge/Massachusetts, 2003.
- [TT02] THEBERGE-TURMEL, MICHAUD F.: *Mobile Robotic Toys and Autism, Creating Relationships with Computers and Robots*. In: DAUTENHAHN, K., BOND A., D. CANAMERO und B. EDMONDS (Herausgeber): *Socially Intelligent Agents - Creating Relationships with Computers and Robots*, The Netherlands, 2002. Kluwer Academic Publishers.

Literaturverzeichnis

- [Tur99] TURNER, J. H.: *Toward a General Sociological Theory of Emotions*. Journal for the Theory of Social Behaviour, 29(2):133–161, 1999.
- [Tur02] TURNER, S. P.: *Brains / Practices / Relativism. Social Theory after Cognitive Science*. Chicago: University of Chicago Press, 2002.
- [TWB07] TANGUY, E., P. WILLIS und J. J. BRYSON: *Emotions as durative dynamic state for action selection*. In: *In Proceedings of the 20 th International Joint Conference on Artificial Intelligence*, Seiten 1537–1542. Morgan Kaufmann, 2007.
- [Vaa03] VAAS, R.: *Emotionen*. Lexikon der Neurowissenschaft, 2003.
- [Val93] VALK, R.: *Bridging the Gap Between Place- and Floyd-Invariants with Applications to Preemptive Scheduling*. In: MARSAN, M., AJMONE (Herausgeber): *Lecture Notes in Computer Science; Application and Theory of Petri Nets 1993, Proceedings 14th International Conference, Chicago, Illinois, USA*, Band 691, Seiten 433–452. Springer, 1993.
- [Val96] VALK, R.: *On Processes of Object Petri Nets*. Fachbereichsbericht FBI-HH-B-185/96, Universität Hamburg, 1996.
- [Val98] VALK, R.: *Petri Nets as Token Objects - An Introduction to Elementary Object Nets*. In: DESEL, J. und M. SILVA (Herausgeber): *19th International Conference on Application and Theory of Petri nets, Lisbon, Portugal*, Nummer 1420 in LNCS, Seiten 1–25, Berlin, 1998. Springer.
- [Val00a] VALK, R.: *Concurrency in Communicating Object Petri Nets*. In: AGHA, G., F. DE CINDIO und G. ROZENBERG (Herausgeber): *Concurrent Object-Oriented Programming and Petri Nets.*, Berlin, 2000. Springer Verlag.
- [Val00b] VALK, R.: *Mobile and Distributed Object versus Central Referencing*. In: GRABOWSKI, JENS und HEYMER, STEFAN (Herausgeber): *Proceedings of 10. GI-ITG-Fachgespräch FBT 2000: Formale Beschreibungstechniken für verteilte Systeme, Lübeck, June 2000*, Seiten 7–27, Aachen, 2000. Shaker Verlag.
- [Val00c] VALK, R.: *Relating Different Semantics for Object Petri Nets, Formal Proofs and Examples*. Technical Report FBI-HH-B-226, University of Hamburg, Department for Computer Science Report/00, 2000.
- [vdAMVW99] AALST, W. VAN DER, D. MOLDT, R. VALK und F. WIENBERG: *Enacting Interorganizational Workflows Using Nets in Nets*. In: BECKER, JÖRG, MICHAEL ZUR MÜHLEN und MICHAEL ROSEMAN (Herausgeber): *Proceedings of the 1999 Workflow Management Conference Workflow-based Applications*, Seiten 117–136, 1999.

- [Vel97] VELÁSQUEZ, J.: *Modeling Emotions and Other Motivations in Synthetic Agents*. In: *Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI-97)*. Providence, RI: MIT/AAAI Press: 10-15, 1997.
- [Vel98] VELÁSQUEZ, J.: *Modeling Emotion-Based Decision-Making*. In: *Proceedings of the 1998 AAAI Fall Symposium, Emotional and Intelligent: The Tangled Knot of Cognition (AAAI Technical Report FS-98-03)*. Orlando, FL: AAAI Press, 1998.
- [Vel99] VELÁSQUEZ, J.: *From Affect Programs to Higher Cognitive Emotions: An Emotion-Based Control Approach*. In: *Online Proceedings of the Workshop on Emotion-Based Agent Architectures (EBAA'99)*. Workshop of the 3 International Conference on Autonomous Agents (Agents'99), 1999.
- [vLMV03] LÜDE, R. VON, D. MOLDT und R. VALK: *Sozionik: Modellierung soziologischer Theorie*, Band 2 der Reihe *Reihe: Wirtschaft – Arbeit – Technik*. Lit-Verlag, Münster - Hamburg - London, 2003.
- [VM97] VELÁSQUEZ, J. und P. MAES: *Catheris: a computational model of emotions*. In: *AGENTS '97: Proceedings of the first international conference on Autonomous agents*, Seiten 518–519, New York, NY, USA, 1997. ACM.
- [VP99] VYZAS, E. und R.W. PICARD: *Offline and Online Recognition of Emotion Expression from Physiological Data*. Technischer Bericht, MIT Media Lab Tech Report No 488, 1999.
- [vRS84] REEKUM, C.M. VAN und K.R. SCHERER: *Cognitive determinants of emotions: A structural theory*. In: SHAVER, P. (Herausgeber): *Review of personality and social psychology*, Band 5, Beverly Hills, CA, 1984. Sage.
- [vRS97] REEKUM, C.M. VAN und K.R. SCHERER: *Levels of processing in emotion-antecedent appraisal*. In: G., MATTHEWS (Herausgeber): *Cognitive science perspectives on personality and emotion*, Seiten 259–300, Elsevier, Amsterdam, 1997.
- [vS04] SCHEVE, C. VON: *Emotionale Agenten - Eine explorative Annäherung aus soziologischer Perspektive*. Diplomarbeit, Universität Hamburg, Hamburg, 2004.
- [vSM04] SCHEVE, C. VON und D. MOLDT: *Emotion: Theoretical Investigations and Implications for Artificial Social Aggregates*. In: G. LINDEMAN, D. MOLDT und P. PAOLUCCI (Herausgeber): *Regulated Agent-Based Social Systems*, Band LNAI Vol. 2934, Seiten 189–209. Berlin: Springer, 2004.
- [vSMFvL06] SCHEVE, C. VON, D. MOLDT, J. FIX und R. VON LÜDE: *My Agents Love to Conform: Norms and Emotion in the Micro-Macro Link*. Computational and Mathematical Organization Theory (CMOT), 2006.

Literaturverzeichnis

- [vST02] STEEN, M. VAN und A. S. TANENBAUM: *Distributed Systems - Principles and Paradigms*. Prentice Hall, 2002.
- [vSvL05] SCHEVE, C. VON und R. VON LÜDE: *Emotion and Social Structures. Towards an Interdisciplinary Approach*. Journal for the Theory of Social Behaviour, 35(3):303–328, 2005.
- [WD01] WOOD, M.F. und S.A. DELOACH: *An Overview of the Multiagent System Engineering Methodology*. In: CIANCARINI, P. und M. WOOLDRIDGE (Herausgeber): *Proceedings of the First International Workshop on Agent-oriented Software Engineering*. Springer, 2001.
- [Web22] WEBER, M.: *Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie*. Tübingen: Mohr, 1976 [1922].
- [Weh98] WEHRLE, T.: *Motivations behind modeling emotional agents: Whose emotions does your robot have?* In: *Proceedings of the Workshop on Grounding Emotions in Adaptive Systems, 5th International Conference of the Society for Adaptive Behaviour (SAB'98)*, Zürich, 1998.
- [Wei66] WEIZENBAUM, J.: *ELIZA - A computer program for the study of natural languagecommunication between man and machine*. Communications of the ACM, 9(1):36–45, 1966.
- [Wei99] WEISS, G.: *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press, Cambridge, MA, 1999.
- [WEKBM08] WESTER-EBBINGHAUS, M., M. KÖHLER-BUSSMEIER und D. MOLDT: *From Multi-Agent to Multi-Organization Systems: Utilizing Middleware Approaches*. In: ARTIKIS, ALEXANDER, GAUTHIER PICARD und LAURENT VERCOUTER (Herausgeber): *International Workshop Engineering Societies in the Agents World (ESAW 08)*, 2008.
- [Wie01] WIENBERG, F.: *Informations- und prozesorientierte Modellierung verteilter Systeme auf der Basis von Feature-Structure-Netzen*. Dissertation, Universität Hamburg, Fachbereich Informatik, 2001.
- [Win06] WINTHER, E.: *Motivation in Lernprozessen*. DUV Deutscher Universitäts-Verlag, 2006.
- [WJ95] WOOLDRIDGE, M. und N. R. JENNINGS: *Intelligent agents: Theory and practice*. Knowledge Engineering Review, 10:115–152, 1995.
- [WJK00] WOOLDRIDGE, M., N. R. JENNINGS und D. KINNY: *The Gaia Methodology for Agent-Oriented Analysis and Design*. Journal of Autonomous Agents and Multi-Agent Systems, 3(3):285–312, 2000.

- [Woo97] WOOLDRIDGE, M.: *Agent-based software engineering*. IEEE Proceedings on SoftwareEngineering, 144:23–37, 1997.
- [Woo02] WOOLDRIDGE, M.: *An Introduction to Multiagent Systems*. John Wiley and Sons Ltd, 2002.
- [WR20] WATSON, J. B. und R. RAYNER: *Conditioned emotional reactions*. American Psychologist, 55:313–317, 1920.
- [Wri94] WRIGHT, I. P.: *An emotional agent: the detection and control of emergent states in autonomous resource-bounded agents*. Technischer Bericht, Cognitive Science Research Report RP-94-21. School of Computer Science and Cognitive Science Research Centre. University of Birmingham, 1994.
- [WSB96] WRIGHT, I. P., A. SLOMAN und L. BEAUDOIN: *Towards a design-based Analysis of Emotional Episodes*. In: *Philosophy Psychiatry and Psychology*, Seiten 101–126, 1996.
- [WW95] WALKER, A. und M. WOOLDRIDGE: *Understanding the Emergence of Conventions in Multi-Agent Systems*. In: *Proceedings of International Conference on Multiagent Systems - ICMAS*, Seiten 384–389. MIT Press, 1995.
- [ZS00] ZENTNER, M. R. und K. R. SCHERER: *Partikuläre und integrative Ansätze*. In: OTTO J., EULER H. A. UND MANDL H. (Herausgeber): *Handbuch Emotionspsychologie*, Seiten 151 – 164. Weinheim: Beltz, PsychologieVerlagsUnion., 2000.
- [ZS09] ZHAO, R. und D. SZAFRON: *Learning Character Behaviors using Agent Modeling in Games*. In: *roceedings of the Fifth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE-09), Stanford, USA*, Seiten 179–185, 2009.
- [ZS11] ZHAO, R. und D. SZAFRON: *Generating Believable Virtual Characters Using Behavior Capture and Hidden Markov Models*. In: *Advances in Computer Games 13 Conference (ACG), Tilburg, The Netherlands*, 2011.

Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, den 2. Dezember 2012