

Universität Trier
FB I – Philosophie

Magisterarbeit

Konnektionistische Systeme

Simuliertes „Bewußtsein“

oder

Bewußtsein selbst?

vorgelegt im August 1999

von

Claus-Christian Carbon

Betreuung:

Prof. Dr. Klaus Fischer

Prof. Dr. Karl F. Wender



Frontispiz:

Komposition aus einer
Zeichnung des Alchemisten Robert Fludd
und einem
Ausschnitt eines neuronalen Netzes

INHALTSVERZEICHNIS

VORWORT

1	Einleitung: Ziel der Arbeit	1
1.1	Eingrenzung der zu behandelnden Problematik	1
1.2	Interdisziplinarität	2
1.2.1	Syntopie	3
1.2.2	Neurophilosophie	4
1.3	Intelligibilitätslücke	5
2	Begriffsklärungen	7
2.1	Bewußtsein	7
2.1.1	Begriffsgeschichte	7
2.1.2	Definitions- und Abgrenzungsversuche	9
2.1.3	Spezifische Schwierigkeiten des Terminus	11
2.1.3.1	Subjektivität	11
2.1.3.2	Deutscher Sprachraum	12
2.2	Konnektionismus	12
3	Klassische KI	15
3.1	Einleitung	15
3.2	Paradigmen	16
3.2.1	Ideal des rational Handelnden	16
3.2.2	Symbolverarbeitung/Funktionalismus	17
3.2.3	Maschinen-Metapher	18
3.3	Unterscheidung starke/ schwache KI	20

3.4	Grenzen der KI	21
3.4.1	Praktische Leistungen aktueller KI-Systeme	21
3.4.2	Technische Begrenzungen	22
3.4.3	Theoretische Probleme	22
3.4.3.1	<i>Formalismus-Problematik</i>	22
3.4.3.2	<i>Explizitheitsanspruch</i>	23
3.4.3.3	<i>Kontextproblem und fehlende Ganzheitlichkeit</i>	24
3.4.3.4	<i>Homunculus-Fehlschluß</i>	24
3.4.3.5	<i>Gödelsches Unvollständigkeitstheorem</i>	25
4	EXKURS: STRUKTUR DES GEHIRNS	26
4.1	Hochkomplexes neuronales System	26
4.2	Entwicklungsbedingungen des Gehirns	28
4.2.1	Ontogenese	28
4.2.2	Phylogenese	28
4.3	Fähigkeiten und Möglichkeiten des Gehirns	29
4.4	Schlußfolgerungen	30
5	KONNEKTIONISMUS	32
5.1	Theorie	32
5.1.1	Theoretische Vorüberlegungen	32
5.1.2	Historie	33
5.1.3	Theorie	34
5.1.3.1	<i>Grundtypen</i>	34
5.1.3.2	<i>Typen von Lernregeln</i>	34
5.1.4	Der Vergleich zu natürlichen neuronalen Netzen	35
5.1.5	Probleme konnektionistischer Systeme	36
5.2	Erfolgreiche Implementationen	37
5.2.1	Muster- und Zeichenerkennung	37
5.2.2	Sprachemulation	37
5.2.3	Robotik	38
5.2.4	Qualitätskontrolle	39
5.2.5	Erkennung von dreidimensionalen Körpern	39
5.2.6	Neuroprothetik	40
5.2.7	Börsenvorhersage	40
5.2.8	Kreditzuschreibung	41
5.3	Was ist das neue an künstlichen neuronalen Netzen?	42
5.4	Ressentiments gegenüber künstlichen neuronalen Netzen	43
5.4.1	Kritiker aus der klassischen KI-Gemeinde	44
5.4.2	Symbolverarbeitung als Eckpfeiler westlicher Philosophie	45
5.4.3	Schwere „Kontrollierbarkeit“	47
5.4.4	Fehlende „Einheit“ bei neuronalen Netzen	48
5.4.5	Konfundierung mit klassischer KI	48
5.5	Neueste und zukünftige Entwicklungen	49
5.5.1	Abschätzung des Potentials	49
5.5.2	Praktische Perspektiven	49
5.5.3	Auflösung einer alten Feindschaft	51

6	BEWUßTSEIN	52
6.1	Das Phänomen „Bewußtsein“	52
6.1.1	Einzig, einzigartig oder einsam und alleine?	52
6.1.2	“Handicap” Subjektivität	55
6.1.3	Komplexität des Phänomens	56
6.1.4	Exkurs: Möglicher Nutzen durch Bewußtsein	56
6.2	Anhaltspunkte für Bewußtsein	57
6.2.1	Konzeptuelle Anhaltspunkte	57
6.2.1.1	<i>Gedächtnis</i>	57
6.2.1.2	<i>Analyseebene</i>	58
6.2.2	Biologische Anhaltspunkte	58
6.2.3	Neurobiologische Anhaltspunkte	58
6.2.3.1	<i>Cell assemblies</i>	59
6.2.3.2	<i>Neuronale Muster</i>	59
6.3	Theoretische Ansätze	60
6.3.1	Psychismus–Theorien	60
6.3.1.1	<i>Panpsychismus</i>	60
6.3.1.2	<i>Vitalismus</i>	60
6.3.2	Dualistische Theorien	61
6.3.2.1	<i>Interaktionismus</i>	61
6.3.2.2	<i>Parallelismus</i>	62
6.3.2.3	<i>Epiphänomenalismus</i>	63
6.3.2.4	<i>Exkurs: Lokalisations–Problematik</i>	63
6.3.2.5	<i>Zusammenfassung</i>	65
6.3.3	“Chauvinistische” Theorien	67
6.3.3.1	<i>Kohlenstoff–Chauvinismus</i>	67
6.3.3.2	<i>Sprachfähigkeit = Bewußtsein</i>	69
6.3.3.3	<i>Humanoide Exklusivität</i>	69
6.3.4	Materialismus/Physikalismus	70
6.3.4.1	<i>Elimination des mentalistischen Vokabulars</i>	71
6.3.4.2	<i>Eliminativer Materialismus</i>	72
6.3.4.3	<i>Das Autocerebroskop</i>	76
6.3.5	Quantenmechanische Erklärungen	77
6.3.6	Bewußtsein als Vetoinstanz	79
6.3.7	Funktionalismus	80
6.3.8	Emergenztheorie	80
6.4	Entscheidungskriterien für Bewußtsein	82
6.4.1	Unklare Kriterien	82
6.4.2	Neurophysiologie	83
6.4.3	Simulationen durch Computer und Roboter	83
6.4.4	Turing–Test	84
6.4.5	Gibt es überhaupt eine Entscheidungsgrundlage?	86
7	CONCLUSIO	87
8	LITERATURVERZEICHNIS	91

„Wie auch immer es zustande kam, das Endergebnis war eine maschinelle Intelligenz, die sämtliche Tätigkeiten des menschlichen Gehirns ausführen oder — wie manche es auszudrücken vorzogen — imitieren konnte, allerdings ungleich schneller und verlässlicher. Die Herstellung war außerordentlich kostspielig, und bisher hatte man nur einige Einheiten der *Hal-9000*-Serie gebaut, aber der alte Scherz, es wäre billiger, menschliche Gehirne mit der beliebten alten Methode zu erzeugen, begann schal zu klingen“

Arthur C. Clark: 2001 — Odyssee im Weltraum

Vorwort

Mein besonderer Dank gilt Herrn Prof. Dr. Klaus Fischer und Herrn Prof. Dr. Karl F. Wender, die sich beide spontan als Betreuer dieser Arbeit zur Verfügung stellten. Dies hat mich vor allem deswegen sehr gefreut, da ich durch diese beiden Herren, Experten aus verschiedenen Wissenschaftsbereichen vor mir hatte. Bei einer interdisziplinären Arbeit wie der vorliegenden ist dies von unschätzbarem Vorteil: Während mir Herr Prof. Dr. Klaus Fischer wertvolle philosophische und wissenschaftstheoretische Hinweise liefern konnte, vermittelte mir Herr Prof. Dr. Karl F. Wender wesentliche psychologische und kognitionswissenschaftliche Grundlagen.

Des Weiteren will ich allen danken, die mir zwischen den langen Nächten intensiver Arbeit den nötigen Rückenhalt gaben und mich allzu oft liebevoll umsorgten.

Zum Schluß will ich ganz herzlich meinem Vater, Herrn Dipl.-Ing. Theo R. Carbon danken, der stets fest an mich glaubte und mich in großzügiger Weise unterstützte.

Ich wünsche, ich könnte auch meiner Mutter für die so wichtigen Orientierungsgespräche danken, ohne die ich mich niemals in so konsequenter Weise dem Studium der Philosophie gewidmet hätte.

1 Einleitung: Ziel der Arbeit

1.1 Eingrenzung der zu behandelnden Problematik

Die Technik neuronaler Netze wird heute in vielen Bereichen der Informationsverarbeitung erfolgreich eingesetzt. Exemplarisch sei an das System *NETtalk* erinnert, welches als Antwort auf das von der Firma DEC entwickelte Programm *DECtalk*, verstanden werden muß (Churchland, 1997, S.100 ff.). Der Computerhersteller DEC hatte mit *DECtalk* ein sehr zuverlässiges Aussprachemodul auf der Basis konventioneller Programmierverfahrensweisen für die englische Sprache entwickelt. Im Jahre 1986 ging Terry Sejnowski der Frage nach, ob nicht auch ein konnektionistisches System diese Aufgabe ebenso gut beherrschen könne. Mußte für das Hinzufügen neuer Ausspracheregeln das System von DEC ständig umprogrammiert und adaptiert werden, so benötigte das System von Sejnowski lediglich einige zusätzliche Trainingsdurchläufe, um vergleichbare Resultate zu erzielen. Nach kurzer Entwicklungszeit war *NETtalk* ähnlich leistungsfähig wie das Konkurrenzprodukt von DEC, das nach klassischen Mustern und dadurch weitaus aufwendiger programmiert worden war (vgl. Frey & Schach, 1996).

Ein anderes Beispiel für die hohe Funktionalität und Praxistauglichkeit konnektionistischer Systeme ist ein Netzwerk zur Identifikation von Sonarsignalen. Obwohl es Sonarexperten nicht möglich ist, ihr Fachwissen über die Vielschichtigkeit von Sonargeräuschen explizit zu machen, können neuronale Netze schon nach wenigen Trainingsdurchgängen mindestens ebenso akkurate Kategorisierungen durchführen, wie humanoide Experten¹. Dazu sind lediglich eine geeignete Netzstruktur zu wählen und

¹ das System wurde von Terry Sejnowski und Paul Gorman entwickelt (Churchland, 1997, S.93ff.).

bereits korrekt kategorisierte Sonarsignal–Beispiele (Input: Sonarsignale; Output: Kategorisierungen) dem konnektionistischen System als Trainingsbeispiele vorzugeben. Hervorragende Arbeit leisten konnektionistische Systeme außerdem im Bereich des räumlichen Sehens², der Gesichtswahrnehmung³ oder der Zeichenerkennung (Barr, 1991, S.78ff.; Mallot, 1995, S.845ff.). Gerade der Bereich der Handschriftenerkennung wird heute vor allem von Programmen mit neuronalen Netzwerken dominiert. Der Nachweis genereller Praxistauglichkeit konnektionistischer Systeme ist durch die gerade erwähnten Beispiele überzeugend erbracht. Dennoch bleibt die Frage, wie man solche Systeme qualitativ einstufen muß. Sind es de facto nur komplexe Systeme, die auch mit herkömmlichen Programmierwerkzeugen auf Von-Neumann-Maschinen programmiert werden könnten oder stellen sie eine qualitativ von diesen Systemen unterscheidbare Methode dar, mit der weiterreichende Probleme gelöst werden können? Die im nächsten Abschnitt näher erläuterten Eigenschaften wie *Selbstprogrammierung* durch „Lernen“, *Autoassoziation* und *Generalisation* sind auf jeden Fall durchaus neue Konzepte. Ob diese Eigenschaften jedoch schon alleine einem System *Intelligenz* oder gar *Bewußtsein* oder *Selbstbewußtsein* verleihen können, soll die Grundfrage dieser Arbeit sein.

Die Frage nach der Möglichkeit von künstlichem Bewußtsein ist dabei aus einer ganzen Reihe von Gründen interessant. Zum einen verlangt sie eine genaue Definition dessen, was wir oft sehr unscharf als *Bewußtsein* resp. *bewußt* kennzeichnen. Dies führt zwangsläufig zu Kriterien, die in einem System obligatorisch erfüllt sein müssen, um diesem ein Bewußtsein zuschreiben zu können. Zum anderen muß zur Behandlung dieser Frage die explanatorische Basis von Bewußtsein eingegrenzt werden, d.h. es muß untersucht werden, in welchen Gegenstandsbereich das zu behandelnde Phänomen fällt (Metzinger, 1996c, S.643). Kann es alleine durch die Neurobiologie erklärt werden? Benötigen wir Computersimulationen aus dem Fachbereich der Informatik? Ist es ein rein philosophisches Problem, oder macht die Komplexität des Problems gar eine ausgefeilte Syntopie vieler einzelner Wissenschaftszweige notwendig?

1.2 Interdisziplinarität

Gerade im Bereich der Philosophie des Geistes scheint für ein sinnvolles Arbeiten eine interdisziplinäre Arbeitsweise (*crossing discipline boundaries*) immer mehr ein Muß darzustellen (Lyons, 1996, Einleitung LXIII). Diese Art der Zusammenarbeit kann sowohl temporär in Form von gemeinsamen Forschungsprojekten, als auch im pro-

² z.B. mit dem System *Fusion.net* (nach Churchland, 1997, S.83ff.).

grammatischen Aufheben der Disziplin- und Methodengrenzen, oder auch in der Gründung grundlegend neuer Wissenschaftszweige geschehen. In gewisser Weise ist die Funktion des Gehirns dabei selbst eine „Metapher für die moderne Forschung selbst, denn die integrierte Aktivität des Gehirns ist eine Verwirklichung von Syntopie⁴“ (Pöppel, 1996, S.18), da unterschiedliche Funktionen zwar an verschiedenen Orten verankert sind, sie jedoch in irgendeiner Weise integriert werden müssen, damit überhaupt etwas Sinnvolles geschehen kann.

1.2.1 Syntopie

Der Begriff Interdisziplinarität ist mittlerweile zu einem geflügelten Wort und oft blind benutzten Modebegriff im Alltagsverständnis von Wissenschaftlichkeit avanciert, das fast ebenso oft wie sinnlos für effektives und effizientes Arbeiten ins Feld geführt wird. Tatsächlich entstehen in interdisziplinär agierenden Forschungsgruppen oft existentielle Probleme, die durch die innerhalb jedes einzelnen Forschungszweiges sonst üblichen und durchaus erfolgreichen Bewältigungsstrategien, nicht zu überwinden sind. Dies liegt zum einen daran, daß unterschiedliche, disziplinär sozialisierte Normalwissenschaftler aufeinander treffen, die sich auf dem interdisziplinären Terrain nicht mehr auf ihre Fachselbstverständlichkeiten verlassen können (siehe Hastedt, 1993, S.5). Zum anderen scheinen einzelne Wissenschaftler ihr historisch gewachsenes Expertenwissen und die sich daraus ergebende Macht verlustieren zu sehen, sobald sich ihre *Kollegen* zu sehr den bisher von ihnen selbst exklusiv beforschten Themen widmen. In einem späteren Kapitel soll hierzu ein konkreter Vorfall aus den frühen Zeiten der KI angeführt werden (Dreyfus & Dreyfus, 1996).

Da jeder einzelne Forschungsbereich jeweils nur einen mehr oder minder begrenzten Methodenapparat und ein klar abgegrenztes Fachwissen besitzt, scheinen interdisziplinäre Ansätze trotz dieser Schwierigkeiten dennoch erfolgversprechend. Hier kann gerade ein Perspektivenwechsel entscheidende positive Impulse bereitstellen. Zudem interessieren sich verschiedene Fachgruppen auch für ganz unterschiedliche Fragen, so daß durch eine interdisziplinäre Arbeitsweise eine breitere Abdeckung des Fragenhorizonts ermöglicht wird. Während sich z.B. die Philosophie inhaltlich eher darum bemüht, Probleme in ihrem Wesen zu verstehen, setzen sich die Naturwissenschaften als Ziel die Lösung von naturwissenschaftlichen und mathematischen Problemen (siehe auch Husserl, 1965; Wundt, 1906). Philosophische Probleme betreffen also im Gegen-

³ vgl. Churchland, 1997, S.46ff.

satz zu naturwissenschaftlichen Problemen, Grundfragen unseres menschlichen Selbstverständnisses, die als solche erst allen anderen Wissenschaften eine systematische Einheit geben können (siehe Kant in Weischedel, 1983b, A24). Immanuel Kant verurteilt aufs Strengste den voreiligen Gebrauch von Einzelwissenschaften ohne kritische Philosophie: „Die Kritik der Vernunft führt also zuletzt notwendig zur Wissenschaft; der dogmatische Gebrauch derselben ohne Kritik dagegen auf grundlose Behauptungen“ (Kant in Weischedel, 1983a, B22).

Auch in Bezug auf die logische Struktur ergeben sich eklatante Unterschiede: Naturwissenschaftliche Aussagen haben idealiter einen deskriptiv–sachhaltigen Charakter, philosophische Sätze liegen dagegen den einzelwissenschaftlichen Aussagen systematisch voraus: „Philosophie hat seit jeher auch und insbesondere die Aufgabe, die Begriffe zu klären, die in den Wissenschaften Verwendung finden“ (Wolters, 1993, S.21). Philosophische Sätze wollen also gar keine definitiven Antwort suchen, vielmehr stellen sie den Versuch dar, die Bedeutung der gestellten Fragen zu analysieren und deshalb besser zu verstehen.

Ein bekanntes Beispiel interdisziplinärer Forschungsweise zum Thema Bewußtsein, stellen die gemeinsamen Arbeiten von Sir Karl R. Popper und Sir John C. Eccles dar. Gerade weil diese beiden Persönlichkeiten aus unterschiedlichen Wissenschaftsbereichen kommen — Popper war hauptsächlich Philosoph und Wissenschaftstheoretiker, wohingegen Eccles seines Zeichens Gehirnphysiologe war — ergeben sich gewinnbringende und vielschichtige Diskussionen (z.B. Popper & Eccles, 1997). Tatsächlich ist die Hirnforschung vermutlich die am stärksten durch Syntopie gekennzeichnete Wissenschaftssparte (Pöppel, 1996). Einen besonders aktiven Bereich bildet dabei eben die Schnittstelle zur Philosophie (Metzinger, 1996d).

Diesem Ideal folgend soll in der vorliegenden Arbeit nicht nur auf philosophische Argumente eingegangen werden, sondern darüber hinaus auch wichtiges technisches und wissenschaftliches Grundlagenwissen zu den Themen Konnektionismus und Bewußtsein vermittelt werden.

1.2.2 Neurophilosophie

Einen Schritt weiter als das Konzept der Interdisziplinarität geht die Forderung nach einer eigenständigen Wissenschaftsrichtung, welche gerne unter den Begriff *Neurophilosophie* subsumiert wird. Patricia Churchland prägte in enger wissenschaftlicher Zusammenarbeit mit ihrem Mann Paul Churchland diesen Begriff und verfolgt damit

⁴ Pöppel (1996) schlägt den Begriff *Syntopie* anstatt *Interdisziplinarität* vor, da dieser nicht nur die Zusammenarbeit unterschiedlicher Fachrichtungen in *methodischen* Fragen erfaßt, sondern zudem *in-*

das Programm, auf reduktionistische Weise der Erklärung psychischer Fähigkeiten näherzukommen. Ihres Erachtens ist das Verständnis neurobiologischer Mechanismen nicht bloß Ausschmückung, sondern eine Notwendigkeit. Wesentlich ist jedoch, daß man nicht nur eine reine *bottom-up*-Strategie verfolgt, sondern statt dessen auf verschiedenen Ebenen forscht. Im Gegensatz zu ebenfalls neurobiologisch orientierten Kollegen vertritt Churchland (1996b) außerdem die Ansicht, daß Beschreibungen und Fähigkeiten auf einer höheren Ebene als der neuronalen, nicht als unwissenschaftlich abzuwerten sind. Besonders erfolgversprechend erscheint ihr die Vorgehensweise von Llinás (siehe Churchland, 1996b, S.482ff.), der sowohl phänomenologische als auch neurobiologische Unterschiede zwischen Zuständen des Schlafens, Träumens und Wachseins systematisch untersucht. Tatsächlich ergeben sich durch diese Methode sehr interessante Phänomene, die jedoch an anderer Stelle der vorliegenden Arbeit intensiv diskutiert werden sollen (siehe Kapitel 6.3).

1.3 Intelligibilitätslücke

Zwar scheint es sehr wichtig, sich mit den neuronalen Korrelaten resp. Ursachen für Bewußtheitsphänomene zu beschäftigen (vgl. Churchland, 1996b), jedoch mahnt Kirk (1996) wohl zu Recht, daß man zu allererst ein gutes Verständnis davon benötigt, „*was es bedeutet*, daß etwas bewußt ist“ (Kirk, 1996, S.642). Denn eine bloße Beobachtung der internen Abläufe im Gehirn, verhilft uns noch nicht automatisch zu einem echten Verständnis davon, ob ein bestimmter Zustand bewußt ist oder nicht. Somit entsteht eine *Intelligibilitätslücke* zwischen dem rein physikalischen System und dem bewußten Subjekt. Schon Leibniz machte auf diese Problematik aufmerksam:

„Was jetzt das Denken anbelangt, so ist es sicher, und der Verfasser erkennt es mehr als einmal an, daß es keine begreifliche Modifikation der Materie sein kann, d.h. daß das empfindende oder denkende Wesen nicht eine Maschine wie eine Uhr oder eine Mühle ist, so daß man die Größen, Gestalten und Bewegungen begreifen könnte, deren mechanische Verknüpfung etwas Denkendes und selbst nur Empfindendes in einem Stoffe hervorbrächte, in dem sonst nichts der Art wäre, und was denn auch von selbst durch die Unordnung dieser Maschine aufhören würde. Es ist also der Materie nicht natürlich zu fühlen und zu denken...“ (Leibniz, 1904, S.26).

Diese *Intelligibilitätslücke* werden wir jedoch nicht überbrücken können, indem wir alleine Phänomene beobachten und untersuchen, sondern dadurch, daß wir genau

analysieren, was es bedeutet *bewußt* zu sein. Daher wird uns dieses Thema ausführlich in einem separaten Kapitel (siehe Kapitel 6) beschäftigen.

2 Begriffsklärungen

2.1 Bewußtsein

Dennett (1994) bezeichnet das Phänomen des menschlichen Bewußtseins als beinahe letztes Geheimnis. Seines Erachtens gab es zwar sehr viele „Geheimnisse“ zu Themen wie z.B. Fortpflanzung, Entstehung des Universums oder Gravitation, die bis heute nicht vollständig und vollkommen schlüssig gelöst werden konnten. Jedoch wissen wir wenigstens, wie wir uns diesen Themen nähern können. Dies ist im Falle des Bewußtseins gänzlich anders:

„Unser Bewußtsein stellt sich heute als ein Problem dar, das oft die besten Denker sprach und ratlos macht. Und wie im Falle aller früheren Geheimnisse gibt es viele Menschen, die darauf hoffen, daß es nie entmystifiziert wird, ja, die darauf bestehen, dieses letzte Tabu unangetastet zu lassen“ (Dennett, 1994, S.37f.).

Zusätzlich zum Problem, daß wir uns nicht sicher sind, wie wir uns dem Bewußtsein nähern können, wird der Begriff sehr uneinheitlich gebraucht. Charles Mingus sagte einmal treffend über den Jazz: „Man kann nicht aus nichts improvisieren, man muß aus etwas improvisieren“ (nach Calvin, 1998, S.47). Da wir uns in einer ähnlich unangenehmen Situation befinden, in der wir nicht so recht wissen, *was* wir eigentlich über Bewußtsein wissen, soll zuerst die Begrifflichkeit als solches näher analysiert werden.

2.1.1 Begriffsgeschichte

Der Begriff Bewußtsein hat sehr viele unterschiedliche Bedeutungen, obwohl wir dennoch ein Gefühl dafür haben, was es heißt, bewußt zu sein resp. ein Bewußtsein zu haben (vgl. Anonymus, 1999a). Selbst wenn religiöse und weltanschauliche Fragen ausgeklammert werden, bleibt eine undurchschaubare Vielfalt an Bedeutungen zurück (vgl. Calvin, 1998; Zoglauer, 1998). Dies wird alleine schon aus der Begriffsgeschichte des Wortes heraus verständlich. So sind alleine in der philosophischen Tradi-

tion drei Arten des Bewußtseinsbegriffs verbreitet (siehe Metzinger, 1996a, S.58): Erstens die Denkfigur der *conscientia* als höherstufige Form des Wissens, die sich sowohl auf das Wissen und die Vorstellung als auch auf das Begehren und Wollen ausdehnt. Bewußtsein ist dabei eine spezifische Form des Wissens, die geistige Prozesse begleiten kann. Zweitens der Begriff *synaesthesia*, welcher die vereinheitlichen Leistungen des aktiven Wahrnehmens charakterisiert. Drittens existiert der mittlerweile in philosophischen Disputen am häufigsten verwendete Bewußtseinsbegriff im Sinne von *cogitatio*, der sich bei Descartes konstituierte. In die deutsche philosophische Terminologie wurde der substantivierte Infinitiv „Bewust seyn“ schließlich durch Christian Wolff im Zuge seiner deutschen Schriften im Jahre 1719 eingeführt⁵. Im weiteren Verlauf der Begriffsgeschichte nahmen die Bedeutungen schnell zu, wobei immer wieder auf eine der drei *klassischen* Denkfiguren rekurriert wurde.

Interessanterweise taucht das Wort Bewußtsein auch in anderen Sprachen erst recht spät auf (Preilowski, 1993), was vielleicht als Indiz dafür gewertet werden muß, daß Bewußtsein als etwas sehr Selbstverständliches gegolten hat, über das man nicht explizit zu reden hatte.

Die Etymologie verrät uns, daß das Adjektiv *bewußt* ein Partizip des frühneuhochdeutschen Wortes *bewissen* resp. des mittelhochdeutschen Wortes *beweten* darstellt (Kluge, 1989, S.82). Beide Verben bedeuten „wissen, sich zurechtzufinden“ und wurden in den allgemeinen deutschen Sprachgebrauch durch Luthers Einfluß in Form des Partizips *bewußt* aufgenommen.

In der englischen Sprache läßt sich ebenfalls eine ganze Reihe von Bedeutungen des Adjektivs *bewußt* (*conscious*) finden. Gängige Synonyme sind *aware*, *awake*, *deliberate*, *sensitive* und *arousable*. Auch hier zeigt sich, daß sich schon alleine aufgrund der mangelnden Gleichartigkeit der Konstrukte im Kopfe einzelner Menschen, Kommunikationsschwierigkeiten einstellen können (Calvin & Ojemann, 1994). So schafft oft erst der Kontext die Rahmenbedingungen dafür, was überhaupt mit Bewußtsein gemeint ist (vgl. Bieri, 1996).

Oeser & Seitelberger (1995) versuchen, das Phänomen Bewußtsein in drei verschiedene Zugangsebenen zu unterteilen. Sie unterscheiden zwischen einem introspektiven Bewußtseinsbegriff, einem empirisch operational definierbaren Bewußtseinsbegriff und einem nicht-empirischen intentionalen Bewußtseinsbegriff. Der introspektive

⁵ Christian Wolff: *Vernünfftige Gedanken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt*, Band I, Kapitel 3, § 194 [nach Metzinger (1996a, S.57)].

Zugang ist dabei der ursprüngliche Begriff, da er den individuellen Alltagszugang beschreibt, welcher die Quelle aller unauflösbaren Paradoxien darstellt. In der medizinischen, neurologischen und experimental-psychologischen Forschung wird die empirisch-operationale Zugangsebene benutzt. Der nicht-empirische intentionale Bewußtseinsbegriff schließlich dient der Erkenntnistheorie als apriorische *Bedingung der Möglichkeit der Erfahrung* (siehe Kant in Weischedel, 1983a, B162) und wird wie der empirische Bewußtseinsbegriff indirekt aus den Funktionen der Erkenntnis erschlossen (Oeser & Seitelberger, 1995, S.162f.).

2.1.2 Definitions- und Abgrenzungsversuche

Calvin (1994) ist der Meinung, daß jeglicher universelle Definitionsversuch in Bezug auf den Begriff Bewußtsein zwangsläufig scheitern muß, da es sich um einen *open-ended*-Begriff handelt. Noch pessimistischer zeigt sich Marvin Minsky (nach Calvin & Ojemann, 1994), indem er zynisch den üblichen und unbedachten Gebrauch des Wortes Bewußtsein verspottet: “[consciousness is] one of the words we have for things we don’t understand”. Auch Searle (1996) gibt zu bedenken, daß sich für Bewußtsein keine Definition mit Hilfe hinreichender und notwendiger Bedingungen angeben läßt, und daß es auch nicht möglich sei, diesen Begriff in aristotelischer Manier durch Angabe von *Genus* und *Differentia specifica* zu definieren. In einem anderen Teil der Arbeit (siehe Kapitel 6) werden wir trotz dieser entmutigenden Statements noch genauer auf einige bekannte Definitionsansätze zurückkommen.

An dieser Stelle sollen hingegen nur Zitate aus zwei philosophischen und einem allgemeinwissenschaftlichen Lexikon einen Eindruck von der speziellen Problematik des Bewußtseinsbegriffes vermitteln.

Ein sehr knapper und klarer Definitionsversuch aus dem *Philosophischen Wörterbuch* von Schischkoff (1991) entspricht dem Stil und dem Anspruch eines kompakten Wörterbuchs, sollen hier doch viele verschiedene Facetten eines Begriffs angesprochen werden, ohne sich zu sehr mit den Details zu beschäftigen. Um den Umfang der Arbeit nicht unnötig zu erhöhen, soll hier lediglich der Basisteil, ohne Berücksichtigung der Begriffsgeschichte und ohne Beschreibung der Verwendung in einzelnen philosophischen Strömungen, wiedergegeben werden:

„Bewußtsein [ist] in der Psychologie das Insgesamt der Bewußtseinsinhalte, die normalerweise begleitet sind von einem mehr oder weniger deutlichen Wissen (Schischkoff, 1991, S.76, Erg. v. Verf.).

Sehr viel spezieller geht Lorenz (1973) mit seiner transzendentallogischen Analyse von Bewußtsein vor. Obwohl die Analyse sehr genau gestaltet ist und somit erheblichen Umfang besitzt, erscheint mir aufgrund der Schärfe der Analyse eine ausführliche Zitation dennoch angebracht:

„Bewußtsein stellt immer etwas vor. Ein Bewußtsein, das nichts vorstellt, ist ein Vorstellen, das nichts vorstellt, ein hölzernes Eisen. Wenn eine Vorstellung nicht als der Phantasie entsprungen bewußt ist, so nimmt das vorstellende Subjekt an, daß das vorgestellte Objekt so vorgestellt sei, wie es in Wahrheit ist. Diese Annahme ermöglicht erst die Täuschung. Wäre nicht bei einer, sich später als Täuschung zeigenden Vorstellung angenommen, vorgestellt sei, was wirklich wäre, so hätte sich das Subjekt in kritischer Distanz zu seiner Vorstellung gehalten oder die Täuschung durchschaut: jedenfalls wäre es nicht getäuscht worden. Wenn auch das Faktum der Täuschung die Tatsache der naiven Annahme, das Vorgestellte sei so, wie es wirklich ist, bestätigt, so stellt es doch den Inhalt der Annahme in Frage; denn der Getäuschte stellt ja gerade etwas vor, was nicht wirklich ist. Die Täuschung stellt die Selbstverständlichkeit des Vorurteils, das Vorgestellte sei vorgestellt, wie es in Wahrheit wirklich sei, in Frage. Somit fragt sich, ob das, was im Bewußtsein erscheint, so erscheine, wie es ist.

Wenn die Frage so gestellt wird, so ist damit zugleich gesagt, daß das Bewußtsein, falls es in Wahrheit erkenne, dasjenige zur Erscheinung bringe, was ist, anders gewendet: daß das Bewußtsein das abbilde, was ist. Das, was ist, ist das Urbild, welches das Bewußtsein wiederzugeben hat, wie es an sich ist. Damit gibt das, was ist, in seinem An–Sich–Sein das Maß, nach dem das Bewußtsein sich richten muß, wenn es sich nicht täuschen will. Das Bewußtsein wird vom An–Sich des Gegenstandes bestimmt, wenn es wahr vorstellt.

Wenn aber das An–Sich als Maßstab der Vorstellung behauptet wird, so behauptet dies das vorstellende Subjekt. Das An–Sich kann also nicht anders in die Theorie des Bewußtseins eingehen, denn als bewußtes. Eine Theorie, die das An–Sich prinzipiell jenseits des Bewußtseins ansiedeln wollte, vergäße, daß ihr das An–Sich, von dem sie redet, bewußt ist; die sich hier zeigende Immanenz des An–Sich im Bewußtsein wird bei jeder Auflösung einer Täuschung deutlich. Eine Ent–Täuschung geschieht ja nicht durch ein reines An–Sich, sondern durch das erscheinende An–Sich, das den täuschenden Schein eines An–Sich verdrängt. Die durchgeführte Überlegung hebt die Differenz von bewußtem Gegenstand und An–Sich–Sein dieses Gegenstandes nicht auf, aber sie erweist sie als eine im und vom Bewußtsein selbst gezogene Differenz“ (Lorenz, 1973, S.234).

Der letzte Definitionsversuch aus dem Brockhaus–Universallexikon bezieht sich speziell auf die Bedeutung von Bewußtsein für die Erkenntnistheorie:

„Bewußtsein eines Objekts ist die Zusammenfassung von Vorstellungen zur Einheit. Diese Zusammenfassung unterliegt hinsichtlich ihrer Aktualisierung psychologischen Gesetzen...“ (Anonymus, 1953, S.86).

Alleine an diesen drei Definitionen läßt sich bereits ablesen, daß sich das Bewußtsein vor allem dadurch auszeichnet, daß es Inhalte in irgend einer Weise zusammenfaßt und daß sich das Bewußtsein dabei Täuschungen ausgesetzt sieht.

2.1.3 Spezifische Schwierigkeiten des Terminus

Wie schon die vorherigen Punkte aufzeigen, ergeben sich spezifische Schwierigkeiten bei der Verwendung des Terminus *Bewußtsein*.

2.1.3.1 Subjektivität

Eines der Hauptprobleme ist, daß Bewußtsein immer subjektiver Natur ist. Es ist stets *jemandes Bewußtsein von etwas*. Dies bedeutet aber zugleich, daß Bewußtseinsphänomene einen ganz anders gearteten Charakter als Naturphänomene besitzen. Daß die Ontologie des Geistes eine Ontologie der ersten Person ist, wurde aber beständig von der Wissenschaft übersehen (vgl. Searle, 1996). Metzinger (1996b) stellt daher die berechnete Frage

„Sind subjektives Empfinden und das Entstehen einer Innenperspektive überhaupt als Bestandteil der natürlichen Ordnung der Dinge *denkbar* — oder werden wir an dieser Stelle mit einem letztlich unauflösbaren Mysterium konfrontiert, mit einem weißen Fleck auf der Landkarte des wissenschaftlichen Weltbildes, der vielleicht aus *prinzipiellen* Gründen immer ein weißer Fleck bleiben muß?“ (Metzinger, 1996b, S.15).

Es ist wichtig, an dieser Stelle festzuhalten, daß Subjektivität hier nicht als erkenntnistheoretischer Modus, sondern als ontologische Kategorie zu verstehen ist.

Dies hat mehrere wichtige Implikationen. Zunächst ist ein Bewußtseinszustand nicht jedem Beobachter gleichermaßen zugänglich. So, wie man selbst in einer ganz besonderen Beziehung zu seinen Bewußtseinszuständen steht, die nicht so wie die eigenen Beziehungen zu den Bewußtseinszuständen anderer Menschen sind, so besitzen auch andere Menschen wiederum eine ganz eigene Beziehung zu ihren Bewußtseinszuständen. Der Zugang zur Welt über das Bewußtsein ist daher immer zwingend perspektivisch.

Wenn es nun aber diesen irreduziblen Faktor Subjektivität gibt, wie soll man dann jemals zu einem kohärenten Weltbild gelangen, das von jedermann in gleicher Weise

akzeptiert werden könnte? Und wie könnte eine Methode zur Kommunikation subjektiver Bewußtseinsinhalte aussehen, um auch anderen Personen unsere Bewußtseins Erfahrungen mitteilen zu können?

2.1.3.2 Deutscher Sprachraum

Ein weiteres Problem besteht darin, daß das Wort *Bewußtsein* in seinem ganzen Bedeutungsumfang scheinbar nur in der deutschen Sprache anzutreffen ist. Somit scheint auch die *Philosophie des Bewußtseins* eine nur dem deutschen Sprachraum typische Angelegenheit darzustellen (vgl. Wiener, 1990). In manchen Sprachen existiert noch nicht einmal ein Äquivalent (siehe Bieri, 1996). Ein Sprechen oder Philosophieren über Bewußtsein kann daher möglicherweise alleine aus sprachlichen und kulturellen Unterschieden zu Mehrdeutigkeiten und Verständnisproblemen führen.

2.2 Konnektionismus

Der Begriff Konnektionismus ist im Deutschen bis heute kaum gebräuchlich, abgesehen vielleicht von Bereichen der Informatik, den Neurowissenschaften und der Psychologie. Daher sei kurz die genauere Begrifflichkeit eingegrenzt.

In technischem Kontext versteht man unter Konnektionismus normalerweise eine komplexe parallele Informationsverarbeitung innerhalb eines Netzwerkes miteinander verschalteter Informationseinheiten. Solch ein Netzwerk wird wegen der Analogie zu Gehirnstrukturen auch als *neurales Netz* resp. *neuronales Netz* bezeichnet (siehe Brunak & Lautrup, 1993).

Maßgebliche Pionierarbeit in der Erforschung konnektionistischer Systeme leisteten Rosenblatt in den 50er Jahren und David E. Rumelhart und James L. McClelland mit ihrem 1986 erschienenen Buch *Parallel Distribute Processing*⁶, das von Konnektionismus-Kritikern sogar abschätzig als „Manifest der Konnektionisten“ (Papert, 1996, S.2) bezeichnet wird.

⁶ deutscher Titel: „Parallelverarbeitung“.

Allgemein zeichnet sich der konnektionistische Ansatz durch drei Hauptcharakteristika aus:

1. Informationen werden nicht mehr wie bei herkömmlichen Computersystemen symbolisch gespeichert, sondern *subsymbolisch*, in Form von verteilten Informationsträgern, sogenannten *Knoten*, die untereinander verschaltet sind. Die Stärke dieser Verschaltungen kann dabei numerisch festgelegt werden.
2. Der Prozeß der Informationsverarbeitung in einem solchen Netz verläuft parallel, also nicht mehr Von-Neumann⁷-Maschinen⁸-konform.
3. *Selbstlernen* durch Auto-Assoziationen: Die Software-Programmierung muß nicht alle Sonderfälle und Sonderregeln einer zu lernenden Struktur explizit berücksichtigen. Statt dessen wird das Netz mit Beispielen „trainiert“.

Dabei gilt zu beachten:

„Der Konnektionismus [ist] keine Theorie und auch kein ideologischer Ansatz in der Kognitionswissenschaft, sondern umfaßt eine Gruppe von Theorien und Ansätzen, die ein gemeinsames Interesse an komputationalen Systemen teilen, die durch *konnektionistische* Basisoperationen gekennzeichnet sind“ (Hatfield, 1993, S.138).

Obwohl konnektionistische Systeme bereits in vielen Maschinen, wissenschaftlichen Simulationen und Computerprogrammen Anwendung finden, ist der Gebrauch des Wortes „Konnektionismus“ noch nicht sehr weit verbreitet. Dies zeigt u.a. eine Analyse der Titel von Neuerscheinungen. Hierzu wurden von mir alle per EDV (*OPAC-System*) erfaßten Buch- und Zeitschriftenbestände der Staatsbibliothek in Berlin nach sämtlich vorkommenden Stichworten *Konnektionismus* resp. *connectionism* im Titel durchsucht. Es fanden sich lediglich 35 Titel, in denen mindestens eines der gesuchten Stichworte vorkam, wobei die erste Fundstelle das Jahr 1987 als Publikationszeitpunkt auswies.

Ursprünglich geht der Begriff Konnektionismus auf Thorndike⁹ zurück. In der Thorndikeschen Definition wird angenommen, daß verhaltensmäßige *responses* auf spezifi-

⁷ Von Neumann, John, geb. 1903, gest. 1957, U.S.-amerikanischer Mathematiker. Grundlegende Arbeiten auf fast allen Gebieten der modernen Mathematik.

⁸ Von Neumann-Maschine: eine rein seriell arbeitende(r) Maschine/Computer, d.h. die erfolgreiche Abarbeitung eines Arbeitsschrittes kann erst dann erfolgen, wenn der vorhergehende Arbeitsschritt bis zur Gänze abgeschlossen worden ist.

sche Stimuli durch einen Prozeß von Versuch und Irrtum (*trial and error*), welcher über neuronale Verknüpfungen zwischen den Stimuli und den am besten passenden *responses* vermittelt werden, zustande kommen (Anonymus, 1999c).

Schefe (1993) versucht herauszustellen, daß konnektionistische Systeme noch viel mehr leisten können, als nur eine theoretische Basis für die *black-box* des Gehirns zu liefern. Unterschied man bislang zwischen *analysierend* und *synthetisierend*¹⁰, so könnte der Konnektionismus einer dritten Dimension Relevanz verschaffen: Als *konstruierende* Wissenschaft könnte der Konnektionismus im Sinne künstlichen Lebens „Ernst machen mit der Vorstellung, daß nur dasjenige wissenschaftlich verstanden sei, was man auch selber konstruieren kann“ (Schefe, 1993, S.4).

⁹ Thorndike, Edward Lee, geb. 1874, gest. 1949, U.S.–amerikanischer Psychologe, dessen Arbeit über Tierverhalten und Lernverhalten zur Theorie des Konnektionismus führten.

¹⁰ z.B. ist das Gebiet der Informatik neben ihrem analysierenden Gestus zusätzlich synthetisierend.

3 Klassische KI

3.1 Einleitung

Unter klassischen KI-Systemen¹¹ wollen wir im folgenden KI-Implementationen im Sinne des Von-Neumannschen Ansatzes bezeichnen, die ihre Anweisungen seriell abarbeiten und deren Programme alleine auf explizitem Wege initialisiert werden. Alle „Entscheidungen“, die der Computer trifft, werden *explizit* in Form von „Wenn-dann“-Anweisungen programmiert. Von-Neumann-Systeme sind daher regelbasiert. Daten werden in diskreten Strukturen gespeichert, die optimalerweise keine Redundanz aufweisen, d.h. ein Datum wird immer nur genau in *einer* Struktur an *einem* bestimmten Ort repräsentiert.

Klassische KI-Implementationen wie etwa der *General Problem Solver* von Newell und Simon aus dem Jahre 1956 (siehe Mayer, 1992, 177f.) basieren auf solchen herkömmlichen Computersystemen. Von deren Verfechtern wird die Ansicht vertreten, daß

„das menschliche Gehirn und der digitale Computer, obwohl völlig verschieden in Struktur und Mechanismus, auf einem gewissen Abstraktionsniveau eine gemeinsame, funktionale Beschreibung aufweisen“ (Newell & Simon nach Graubard, 1996, S.16).

Newell und Simon, zwei der Hauptvertreter der KI, sahen im Jahre 1958 die Zukunft der KI sehr optimistisch. So behaupteten sie, daß „Intuition, Einsicht und Lernen nicht länger ausschließlich menschliche Eigenschaften seien“ und daß „jeder große Com-

¹¹ John Haugeland bezeichnet solche auf dem klassischen Symbolverarbeitungs-Paradigma basierenden Systeme auch als *GOF AI* (Good Old Fashioned Artificial Intelligence); nach Zimmerli & Wolf (1994, S.6).

puter mit hoher Geschwindigkeit so programmiert werden kann, daß er entsprechend über sie verfügt“¹².

Historisch gesehen lassen sich bereits seit Ende des 2. Weltkrieg ernsthafte Versuche erkennen, denkende Maschinen zu bauen. Entscheidende Impulse kamen in den Anfangsjahren von der Kybernetik, die mit Hilfe elektronischer Schaltkreise Nervensysteme nachzuahmen versuchte (siehe Moravec, 1996). Die etwas später initiierte Forschungsrichtung der Künstlichen Intelligenz nutzte anfangs dagegen die Rechenkapazität, um Lehrsätze aus der Logik zu beweisen (siehe Moravec, 1996; Siekmann, 1994).

3.2 Paradigmen

Die Bandbreite der Zielvorstellung der einzelnen KI-Gemeinden differiert derart, daß es notwendig ist, die gemeinsamen Ursprünge und Basisforderungen vorzustellen, als auch eine kurze Unterscheidung von einzelnen KI-Strömungen aufzuzeigen. Nur so können die Gegenpositionen, die von der konnektionistischen Seite angeführt werden, sinnvoll eingeordnet und interpretiert werden (vgl. Tetens, 1994).

Die folgenden Abschnitte geben daher einen kurzen Einblick darüber, welche Geistes-Strömungen und welche darin enthaltenen Prämissen als Grundlage für die Forschung der Künstlichen Intelligenz anzusehen sind.

3.2.1 *Ideal des rational Handelnden*

Einen wesentlichen Anteil an den Ursprüngen der Künstlichen Intelligenz besitzt das Ideal des rational handelnden Menschen. Als unwichtig wird dagegen der Rückbezug auf den Körper gesehen, einzig alleine von Interesse ist der reine Geist, dessen Intelligenz, so die Meinung der KI, vollständig durch Informationsverarbeitungsprozesse verstanden werden könne (vgl. Christaller, 1996; Siekmann, 1994). Ursache für die Verwendung dieses Paradigmas ist vermutlich das Faktum, daß sich in der gesamten Wissenschaftsgeschichte „das schlußfolgernde Denken als außerordentlich leistungsfähige Methode auf dem Weg vom Bekannten zum Unbekannten erwiesen hat“ (Bharati, 1996, S.201).

Mit diesem sehr eingeschränkten, alleine auf Rationalität abzielenden, Intelligenzbegriff, der als *Arbeitshypothese* jedoch durchaus sinnvoll ist, wurden vor allem Anfang der 70er Jahre bis in die Mitte der 80er Jahre eine Reihe von beeindruckenden technischen Ergebnissen erzielt. Einen Hauptteil davon machten die sogenannten *Experten-*

¹² Herbert Simon & Allen Newell: Heuristic Problem Solving. The Next Advance in Operations Research, in: *Operations Research* 6 (Januar-Februar 1958), S.6; nach: Dreyfus & Dreyfus (1996, S.19).

systeme aus, deren Aufgabe darin bestand, das Wissen hochspezialisierter Fachleute in Form von regelbasierten Abläufen zu repräsentieren (vgl. Siekmann, 1994). Vorbedingung dafür ist, daß sich das Wissen der Experten explizit machen läßt und daß es klaren Regeln folgt (*Ideal des rational Handelnden*). Die erfolgreichsten Expertensysteme sind daher Klassifikationsprogramme, welche hauptsächlich Verwendung in der Medizin und in geologischen Erkundungen finden (vgl. Zimbardo, 1992).

3.2.2 Symbolverarbeitung/Funktionalismus

Der Funktionalismus in der Philosophie des Geistes ist eine Theorie, die explizit erstmals Anfang der 60er Jahre formuliert worden ist: „Die funktionalistische Auffassung des Mentalen entwickelte sich schnell zur offiziellen Hintergrundmetaphysik der KI“ (Keil, 1993, S195), da sie geradezu auf das Computermodell zugeschnitten ist: „Gedanken verhalten sich aufgrund eines Symbolismus parallel zur Realität“¹³. Einfach ausgedrückt verhält sich gemäß dem Ansatz des Funktionalismus der Geist zum Gehirn wie ein Computerprogramm zur Computer-Hardware. Da Computer hierbei im herkömmlichen Von-Neumannschen Sinne zu verstehen sind, werden entsprechend der Vorstellung von Fodor und Pylyshyn psychologische Prozesse mit Symbolverarbeitungsprozessen gleichgesetzt (siehe Bechtel, 1993b; Hatfield, 1993, S.146; Lyons, 1996). Das physiologische Substrat, das in biologischen Systemen auch gerne als *wetware* (so in Keil, 1993, S.196) bezeichnet wird, spielt dabei keine essentielle Rolle, und wird dadurch austauschbar:

„Das physiologische Substrat stellt lediglich das Medium dar, in dem kognitive Funktionen realisiert werden; in Ermangelung einer klar umrissenen Identität zwischen der kognitiven Funktion und ihrer physiologischen Realisierung [...] ist dieses Medium für die Formulierung genuin psychologischer Theorien unbedeutend“ (Hatfield, 1993, S.146).

¹³ Craik, K.J.W. (1943). *The Nature of Explanation*. Crambridge: Cambridge Univ. Press; zit. nach Cowan & Sharp (1996, S.119).

Mentale Zuschreibungen können daher auch jenseits unserer Spezies legitim sein, da mentale Ereignisse lediglich funktional definiert werden. Nur mit Hilfe dieser Prämisse kann Pylyshyn (1989) auch *computations*¹⁴ mit *cognitions* gleichsetzen¹⁵.

Auch die moderne Kognitionswissenschaft benutzt als Arbeitsgrundlage die einfache Annahme, daß Kognitionen mit der Speicherung und Manipulation von Informationen gleichzusetzen sind (vgl. Bechtel, 1993a; Dörner, 1994). Die Art der Informationen ist allerdings nicht alleine auf satzartige Strukturen beschränkt (z.B. Pylyshyn, 1981), sondern kann z.B. auch bilderähnlich aufgebaut sein (z.B. Kosslyn & Pomerantz, 1977). Bechtel (1993a) sieht einen Grund für die Konzentration auf die sprachliche Repräsentation darin, daß Menschen die meisten ihrer Gedanken in Worte kleiden.

3.2.3 *Maschinen-Metapher*

Die Arbeitshypothese der „harten KI“ geht von einem generellen Mensch–Maschine–Isomorphismus aus, d.h. der Computer wird dem Menschen als grundsätzlich gleichwertig entgegengestellt¹⁶. Somit „erscheint es möglich, alle menschlichen Eigenschaften und Fähigkeiten im Computer zu simulieren...“ (Foerst, 1995). Das „pfiffige“ an diesem verwendeten Paradigma ist, daß schließlich auch mit Hilfe der Reiz–Reaktions–Muster des Computers Rückschlüsse auf Verhaltens–Prozesse des Menschen möglich sein sollen (vgl. Görz, 1995). So können Modelle menschlicher Leistungen gebildet werden und auf Computern realisiert werden. Um diese Modelle zu verifizieren, genügt alleine der Vergleich maschinell erzeugter Verhaltensmuster mit denen des Menschen. Sind keine signifikanten Abweichungen zwischen modellierten und menschlichen Reaktionsmustern erkennbar, so wird davon ausgegangen, daß man den entsprechenden Bereich menschlicher Fähigkeit adäquat abgebildet hat. Dieses Verfahren wird vor allem im Bereich der *Cognitive Science* häufig angewendet:

¹⁴ als *Standarddefinition der Computation* sieht Searle (1996) die typische Operationalität einer Turingmaschine: „Sie kann eine 0 auf ihrem Band durch eine 1 ersetzen; sie kann eine 1 auf ihrem Band durch eine 0 ersetzen; sie kann das Band um ein Feld nach links und um ein Feld nach rechts verschieben. Sie wird durch ein Programm von Anweisungen gesteuert; jede Anweisung spezifiziert eine Bedingung und eine Handlung, die in dem Fall auszuführen ist, in dem die Bedingung erfüllt ist“ (Searle, 1996, S.231)

¹⁵ dazu muß freilich angenommen werden, daß es sich bei *computation* um ein intrinsisches, also nicht–beobachter–relatives Merkmal handelt. Ein ausführliches Für und Wieder dieser Position findet sich z.B. in Searle (1996, S.9ff.)

¹⁶ siehe Simon, Herbert A. (1969). *The Sciences of the Artificial*. Cambridge, Mass.: MIT Press; nach Floyd (1994, S.259).

„Die wesentliche Eigenschaft wissensbasierter Systeme ist ihre Fähigkeit, über ihrem internen Modell einer Anwendungsdomäne Operationen der Informationsverarbeitung durchzuführen und auf diese Weise die zu erklärenden Verhaltensphänomene erzeugen zu können“ (Opwis & Plötzner, 1996, S.5).

Der große Vorteil dieser Modelle ist ihr inkrementeller Charakter, da durch differenzierte Analysen der Abbildungsqualität des Modells, gezielte Verbesserungen und Optimierungen an den internen Prozessen und ihrer Datenbearbeitung durchgeführt werden können. Das Modell hat so die Eigenschaft, wenigstens auf der Basis extern kontrollierbarer Variablen, mit den abzubildenden Geistesprozessen immer weiter zu konvergieren.

Die reine Modulation des *Verhaltens* eines Systems, legt nahe, daß sich die Qualität einer traditionellen KI-Implementation am Turing-Test¹⁷ messen lassen kann. Letztendlich entscheidet also nur das Reiz-Reaktionsmuster darüber, ob eine Implementation als sinnvoll und gelungen erscheint. Wie diese jedoch genau intern realisiert wird, ist von zweitrangigem Interesse. Näheres zum Turing-Test findet der Leser in Kapitel 6.4.4.

Üblicherweise wird der Arzt und Philosoph Julien Offroy de La Mettrie mit seiner zur damaligen Zeit anonymisiert erschienen Publikation *Der Mensch eine Maschine* als *Vater der Maschinenmetapher* angesehen (siehe Calvin, 1998). Er verglich den Menschen mit einer rein mechanisch funktionierenden Maschine:

„Der Mensch ist eine Maschine, welche so zusammengesetzt ist, dass es unmöglich ist, sich zunächst von ihr eine deutliche Vorstellung zu machen und folglich sie zu definieren. Deshalb sind alle Untersuchungen theoretischer Natur, welche die grössten Philosophen angestellt haben, das heisst, indem sie gewissermassen auf den Flügeln des Geistes vorzugehen versuchten, vergeblich gewesen. Also kann man nur practisch, oder durch einen Versuch der Zergliederung der Seele, nach Art der Aufklärung über die körperlichen Organe, ich will nicht sagen mit Sicherheit die Natur des Menschen enträthseln, aber doch wenigstens den möglichst höchsten Grad von Wahrscheinlichkeit über diesen Gegenstand erreichen“ (La Mettrie, 1875, S.21).

¹⁷ benannt nach Alan Mathison Turing, geb. 1912, gest. 1954, brit. Mathematiker und Informatiker. Ein 1951 entwickelter Test zum Nachweis von Maschinenintelligenz. Sollten Dritte das Antwortverhalten von Maschinen in einem Gespräch mit diesen nicht von denen normal intelligenter und sprachbegabter Menschen unterscheiden können, so wird nach Turing diesen Maschinen Intelligenz zugesprochen (siehe Churchland, 1997, S.267).

Obgleich La Mettrie seiner Zeit zweifellos voraus war, hat er dennoch die Maschinenmetapher nicht erfunden. René Descartes hatte sie bereits ein Jahrhundert zuvor in seinem Werk *De Homine* gebraucht (siehe Calvin, 1998, S.14), wobei er durch diese Metapher den Geist nicht völlig verbannen wollte; vielmehr versuchte er dadurch den *Sitz der Seele* genau zu lokalisieren (siehe Kapitel 6.3.2.4).

3.3 Unterscheidung starke/ schwache KI

Bisher ist in dieser Arbeit nur ganz allgemein von *Künstlicher Intelligenz* die Rede gewesen. Bezogen auf die Theorie des Geistes existieren aber mindestens zwei unterscheidbare theoretische Ansätze innerhalb der klassischen KI, die üblicherweise als *starke KI (strong AI)* und *schwache KI (weak AI)* bezeichnet werden (siehe Searle, 1993; Searle, 1997)¹⁸. Die Position der starken KI geht zurück auf Alan Turing (siehe Eysenck & Keane, 1995); den Standpunkt der schwachen KI nehmen dagegen üblicherweise die modernen Kognitionswissenschaften ein. Sie verwenden als wesentliches Forschungsparadigma *Simulationen* geistiger Vorgänge, die lediglich als *mögliche* Implementationen von kognitiven Prozessen angesehen werden. Das Ziel ist, ein Modell zu entwickeln, das ein konsistentes Zusammenspiel einzelner kognitiver Leistungen widerspiegelt. Als wichtigste Parameter bei den Modellierungen werden die Latenzzeiten¹⁹ kognitiver Prozesse angesehen.

Searle (1996) charakterisiert die Grundideen der beiden Strömungen kurz und bündig folgendermaßen:

„Als starke KI bezeichne ich folgende Auffassung: Einen Geist haben heißt ein Programm haben, und mehr ist am Geist nicht dran. Als schwache KI bezeichne ich die Auffassung, daß Hirnvorgänge (und geistige Vorgänge) mittels eines Computers simuliert werden können“ (Searle, 1996, S.227, Klammern i. Original).

Da aus dieser Übersicht deutlich wird, daß die schwache KI lediglich eine Art Simulation geistiger Vorgänge darstellen will, also gar nicht den hehren Anspruch hat, „wahren“ Geist oder Bewußtsein zu schaffen, wird in den folgenden Kapiteln der Begriff *Künstliche Intelligenz* im Sinne von *starker KI* verwendet.

¹⁸ Diese Unterscheidung wird analog zur Unterscheidung eines starken resp. schwachen Funktionalismus durchgeführt (siehe Oeser & Seitelberger, 1995).

¹⁹ die Methode der Erfassung von Latenzzeiten beruht auf einem Ansatz von Joseph Jastrow, nach dem bestimmte psychische Prozesse jeweils spezifische Zeitressourcen benötigen (siehe Luce, 1986).

3.4 Grenzen der KI

Der Mathematiker Alan Turing bewies, daß es zu jeder arithmetischen Funktion, die überhaupt berechenbar ist, eine Maschine gibt, die diese Funktion berechnet, und daß es eine *universelle* Maschine gibt, die den Wert *jeder beliebigen* berechenbaren Funktion für jedes beliebige Argument ermittelt (nach Beckermann, 1994, S.73f.).

Die praktische Umsetzung einer solchen universellen Turing-Maschine führt zu mehreren Problemen: Erstens ist die Frage zu stellen, ob alle Vorgänge und Handlungen auf der Welt und die Welt an sich als Funktionen darstellbar sind. Zweitens, ob solch eine *universelle Maschine* mit technischen Mitteln überhaupt jemals konstruierbar sein wird. Zusätzlich ereilen klassische KI-Systeme noch eine Reihe anderer Probleme, die sich teilweise gegenseitig bedingen.

Um systematisch die Grenzen der KI aufzuspüren, wollen wir uns zunächst die Leistungsfähigkeit aktueller KI-Systeme ansehen. Anschließend sollen technische und theoretische Grenzen des KI-Ansatzes aufgezeigt werden.

3.4.1 *Praktische Leistungen aktueller KI-Systeme*

Warwick (1998) gibt einen extensiven Überblick über die Leistungsfähigkeiten aktueller Robotersysteme. Dabei zieht er die Schlußfolgerung, daß heutige Computer- und Robotersysteme zwar ganz erstaunliche Leistungen vollbringen, jedoch nach wie vor eine dem Menschen wesentliche Funktion vermissen lassen, nämlich Flexibilität und Vielseitigkeit. Seiner Meinung nach zeichnet den Menschen nicht die hohe Akkuratheit bei der Ausführung einer Aufgabe aus, sondern die Fähigkeit, auch auf neue Probleme sinnvoll zu reagieren:

”Simply because a human has been doing a job for many years does not mean that it is the best design to actually perform the task in question. In fact the opposite is probably true, in that the advantage of humans, at present, is our versatility” (Warwick, 1998, S.64f.)

Die von ihm besprochenen Robotersysteme sind allesamt sehr stark auf die ihnen zugedachten Aufgaben spezialisiert, dadurch werden Abweichungen von den Vorgaben häufig durch Fehler, Fehlinterpretationen oder gar Totalausfälle quittiert.

Dies ist auch eines der Hauptprobleme, mit denen die weiter oben erwähnten Expertensystemen (siehe Kapitel 3.2.1) zu kämpfen haben, sobald sie mit Kontexten und Inhalten konfrontiert werden, die vom Programmierer nicht explizit vorgesehen wurden. Churchland & Churchland (1990) kommen zu dem Schluß, daß zwar viele Einzel-„Fähigkeiten“ von klassischen Computersystemen, insbesondere deren Rechengeschwindigkeit, enorm zugenommen haben, ein Mehr an *kognitiver* Leistung jedoch bisher nicht zu beobachten sei.

3.4.2 Technische Begrenzungen

Eine wesentliche, wenn auch rein technische Begrenzung, erfahren serielle Systeme, wenn ein Ausbau ihrer Rechenkapazität gefordert wird. Wohingegen bei parallelen Systemen normalerweise lediglich die Anzahl der Prozessoren erhöht werden muß, ohne daß dabei eine Überarbeitung der Struktur notwendig ist, verlangt eine Kapazitätserhöhung klassischer Computersysteme multiple strukturelle Veränderungen wie z.B. Signalwegeverkürzungen²⁰ oder die Entwicklung noch schneller reagierender Materialien. Dies führt zu einer immer stärkeren Miniaturisierung einzelner Komponenten, welche schließlich auf atomarer Ebene ihre Grenze finden wird.

Nach Einschätzung von David L. Waltz, sind „die derzeitigen seriellen Computer an Grenzen gestoßen, jenseits derer ihre Schnelligkeit nicht mehr im Rahmen angemessener Kosten gesteigert werden kann“ (Waltz, 1996, S.185).

Somit begrenzen alleine schon die zur Zeit zur Verfügung stehenden Computer die problemlösenden Algorithmen.

3.4.3 Theoretische Probleme

3.4.3.1 Formalismus–Problematik

In Bezug auf die für KI–Wissenschaftler selbstverständlichen Formalismus–Annahme ergeben sich schwerwiegende Probleme. Christiane Floyd faßt dies treffend zusammen:

„Die KI stützt sich auf ein spezifisches Menschenbild, das sie vergegenständlicht und in Technik umsetzt. Dieses Menschenbild befaßt sich — nur — mit *rationalem, regelgeleitetem Denken* und *funktionalen, zweckgerichtetem Handeln*, löst diese Fähigkeiten aus dem gesamt–menschlichen Zusammenhang von *Bewußtsein, Erlebnis* und *Erfahrung*, isoliert sie und bildet sie ausschnittsweise nach“ (Floyd, 1994, S.259, Herv. v. Verf).

Im Laufe der KI-Forschung zeigten sich aber auch eine Reihe anderer Probleme. Zu nennen sind vor allem die mangelnde Generalisationsfähigkeit klassischer KI-Verfahren und die fehlenden oder nur schwer realisierbaren Erweiterungsmöglichkeiten eines bestehenden programmierten Systems. Klassische KI–Verfahren verhalten sich weder situationsbezogen noch adaptiv. Daraus ergeben sich zwangsläufig “rigid and brittle systems” (Smolensky, 1989, S.52). Dies zeigt sich z.B. darin, daß KI–Sy-

²⁰ letztendlich setzt die Lichtgeschwindigkeit der internen Kommunikationsgeschwindigkeit und damit der gesamten Verarbeitungsgeschwindigkeit eines Computersystems eine obere Grenze.

steme von Menschenhand gebaut werden und deren späteren Wissenserweiterungen explizit von menschlichen Benutzern programmiert werden müssen (Steels, 1996).

Auch Dreyfus & Dreyfus (1996) sind der Meinung, daß die KI auf der Basis formaler Beschreibungen nicht das Wesen der geistigen Leistungen eines Menschen erfassen kann (siehe auch Wiener, 1990). Zu wirklich *kreativen* Lösungen scheinen KI-Systeme nicht fähig zu sein:

“...a production-system framework²¹ can account for how people transfer a *general skill* they learned in one domain to another but *not for how creative solution is invented* ” (Mayer, 1992, S.201, Herv.v.Verf.).

3.4.3.2 Explizitheitsanspruch

Ein weiterer entscheidender Nachteil klassischer KI-Verfahren ist ihr “need for explication”: Alle Fakten müssen dem Programm „hart“ einprogrammiert werden. Zu diesem Zwecke müssen die nötigen Fakten und Regeln aber überhaupt explizierbar sein. Matlin (1989) ist der Meinung, daß wir vielleicht Möglichkeiten besitzen, um uns im Klaren über die *Produkte* unserer Hirnaktivitäten zu werden, daß wir aber keinesfalls alle unseren Gehirnprozesse *per se* bewußt machen können:

“We may be fully conscious of the *products* of our thought processes [...], but we are usually not conscious of the *process* that created the product” (Matlin, 1989, S.62, Herv. i. Original).

Ohne vollständige Angabe all dieser Prozesse, Fakten und Regeln ist ein Produktionssystem aber nicht vollständig lauffähig, denn “the symbolic level that implements knowledge structures is alleged to be *exact* and *complete* ” (Smolensky, 1989, S.52, Herv. v. Verf.). Paul Churchland äußert sich ebenfalls kritisch über die Möglichkeit einer vollständigen Explikation. Er erläutert dies am Beispiel der Gesichtsidentifikation: „...genauso wie bei den einfacheren Sinneswahrnehmungen können wir verschiedene Gesichter wesentlich besser analysieren als *verbal charakterisieren* “ (Churchland, 1997, S.32), was einer Explikation gleichkommt. Paul Churchland bezweifelt auch, ob man, selbst wenn man über solche explizierten Strukturen und Prozesse verfügen würde, diese in „exakte“ KI-Systeme umsetzen könne:

²¹ Arbeitsansatz mit der Verwendung von *Produktionen* , d.h. einem wohldefinierten Regelsystem in Form von „Wenn-dann-Regeln“.

„...in den letzten Dekaden des 20. Jahrhunderts ist wiederholt die Meinung vertreten worden, daß rein mathematische Prozesse ungeeignet sind, um die volle Bandbreite des menschlichen Verstandes zu erklären“ (Churchland, 1997, S.23).

In diesem Zusammenhang stellt sich zwangsläufig die Frage, was alles in einem KI-System repräsentiert werden muß, damit es sinnvoll funktionieren kann (siehe Churchland & Churchland, 1990). Nach Turing sollte eigentlich das gesamte Wissen über die Welt in eine solche Maschine eingespeist werden, da zu einer Turing-Maschine ein *unendliches Band* gehört, von dem die Programme gelesen und geschrieben werden (nach Waltz, 1996, S.186). Jedoch kann rein theoretisch *keine* Maschine über einen unendlichen Speicher verfügen. Menschen sind von dieser Beschränkung ebenfalls betroffen: Auch sie verfügen letztendlich nur über begrenzte Speicherungs- und Wissensfähigkeiten.

3.4.3.3 Kontextproblem und fehlende Ganzheitlichkeit

Gesetzt den Fall, man könnte das gesamte Weltwissen in einem KI-System repräsentieren, so würde immer noch das sogenannte *Qualifikations-* oder *Rahmenproblem* bestehen, nämlich die Entscheidungsschwierigkeit, welche Reize verarbeitet werden sollen und welche nicht (vgl. Churchland & Churchland, 1990). Das Problem erweitert sich in unüberschaubarer Weise, wenn man noch zusätzlich alle Konsequenzen aus erfolgten und unterlassenen Handlungen betrachten will, um analysieren zu können, welchen Einfluß diese auf die geforderte Zielhandlung haben oder haben könnten (siehe Beckermann, 1994, S.76ff.; Dennett, 1996c, S.278f.; Tetens, 1994, S.98).

Pöppel (1996) gibt daherzu bedenken, daß Computersysteme stets nur als *selektive* Intelligenzverstärkungen auftreten können, da sie jeweils nur Teilleistungen der Intelligenz abbilden, nicht jedoch das gesamte Wirkungsgefüge unseres intelligenten Handelns.

3.4.3.4 Homunculus-Fehlschluß

Searle (1996) kommt zu der Überzeugung, daß es zwangsläufig zu einem Homunculus-Fehlschluß kommen muß, wenn man, wie in der KI geschehen, das Gehirn für einen Computer hält (vgl. Manhart, 1991). Solange man Computer als reine Rechenmaschinen betrachtet, die von Menschen bedient werden, entsteht dieses Problem nicht, da hier der „Homunculus“ mit dem menschlichen Bediener identisch ist.

Ganz anders sieht es jedoch innerhalb der computationalen Theorie des Geistes aus, deren „Idee immer ist, das Hirn so zu behandeln, als wäre da jemand drin, der es zum Rechnen benutzt“ (Searle, 1996, S.238). Zwar gibt es zahlreiche Versuche, diesem Homunculus zu entkommen, jedoch können all diese Argumentationen nicht überzeugen, da sie ab einem bestimmten Punkt der logischen Kette ebenfalls dogmatisch ab-

brechen. So werden z.B. von Daniel Dennett die computationalen Tätigkeiten des Computers in immer kleinere Fähigkeiten zerlegt, d.h. von komplexen Berechnungen hin zu basalen Flip-Flop-Mustern. Dennett will mit Hilfe dieser Zerlegung zeigen, daß höherstufige Homunculi lediglich aus einfachen Schaltern bestehen (siehe Searle, 1996, S.239). Diese rekursive Dekomposition schlägt jedoch fehl, da letztendlich auch die Zustände dieser einfachen Schalter in sinnvoller Weise interpretiert werden müssen, da es nach Searle (1996) ohne die zusätzliche Annahme eines höheren Geistigen noch nicht einmal *Syntax* gibt.

3.4.3.5 Gödelsches Unvollständigkeitstheorem

Stark vereinfacht, besagt das Gödelsche Unvollständigkeitstheorem (*Erster Gödelscher Satz*), daß sich arithmetische Sätze von verhältnismäßig elementarer Natur angeben lassen, die offensichtlich richtig und doch innerhalb des Formalismus nicht abgeleitet werden können (Weyl, 1966, S.279). Somit ist der Formalismus nicht vollständig, d.h. der Allwissenheitsanspruch der Mathematik ist gebrochen.

Die Möglichkeiten algorithmischer maschineller Systeme scheinen dadurch wesentlich eingeschränkt, nicht jedoch die des Mathematikers oder des Menschen im allgemeinen, der die Existenz solcher Beweisbarkeitsgrenzen ja einsehen kann (siehe Penrose, 1991). Inwieweit jedoch durch die Unvollständigkeit formalistischer Systeme auch die Bewußtseinsfähigkeit eines maschinellen System betroffen ist, bleibt ungewiß.

4 Exkurs: Struktur des Gehirns

Da sich der Konnektionismus zumindest von seiner Grundkonzeption her an vernetzten neuronalen Hirnstrukturen orientiert, soll hier kurz sowohl auf die Charakteristika des menschlichen Gehirns als auch auf dessen Entwicklungsbedingungen und Lernvorgänge eingegangen werden. Patricia Churchland (1996b), die einen reduktionistischen Standpunkt einnimmt, ist sogar der Meinung, daß „das Verständnis neurobiologischer Mechanismen nicht bloß Ausschmückung, sondern eine Notwendigkeit darstellt“ (Churchland, 1996b, S.464).

Letztendlich stellt sich in diesem Kontext auch die interessante Frage, warum das Gehirn ein solch besonderes Organ ist, das u.a. Bewußtsein entwickeln kann (vgl. McGinn, 1996).

4.1 Hochkomplexes neuronales System

Das menschliche Hirn besteht etwa aus 100 Milliarden Nervenzellen, wobei jedes einzelne Neuron wiederum zwischen 1.000 und 10.000 Synapsen besitzt²². Schwartz (1996) schätzt die gesamte Langzeitspeicherkapazität des Gehirns auf ca. zehn Billionen Bytes (also etwa 10.000 Terabytes)²³. Die Bandbreite des Gehirns, also die Leistungsfähigkeit pro Zeiteinheit, wird von Schwartz mit 1.000.000 Billionen arithmeti-

²² die Schätzungen über die wahre Neuronenanzahl im menschlichen Gehirn schwanken von 100 Milliarden (Churchland, 1997; Warwick, 1998) bis hin zu 1.000 Milliarden (Schwartz, 1996; Tetens, 1994). Auch die genaue Gesamtanzahl der Synapsen ist nicht bekannt: Während Churchland (1997), Thompson (1990) und Kotulak (, 1996) von 100 Billionen Synapsen sprechen, vermutet Hillis (1996) 1.000 Billionen Synapsen, und Schwartz (1996) schließt nicht aus, daß es sich gar um bis zu 100 Billionen handeln könnte.

²³ andere Schätzungen wie etwa von Waltz (1996) gehen von ungefähr der gleichen Größenordnung aus. Waltz spricht von ca. 4×10^{16} Bits, was etwa 5×10^{15} Bytes, gleich 5000 Terabytes, entspricht.

schen Operationen pro Sekunde (entspricht 1.000.000 Teraflops²⁴) beziffert (Schwartz, 1996, S.124)²⁵. Diese extrem hohe Leistungsfähigkeit wird trotz niedriger Leitungsgeschwindigkeit²⁶ (Saum-Aldehoff, 1993; Thompson, 1990) durch parallel ablaufende Prozesse erreicht (Saum-Aldehoff, 1993).

Wenn man heutige Supercomputersysteme analysiert, so dringt die Technik derzeit gerade einmal in den Teraflop-Bereich vor (siehe Reiter, 1996b); in der Speichertechnik ist man ebenfalls erst im unteren Terabyte-Bereich angelangt:

„Die Herausforderung bis zum Ende der 90er Jahre lautet TTT: ein Teraflops (sic!) Rechenleistung, ein Terabyte Speicher, ein Terabyte/s Kommunikationsbandbreite.“ (Zobel, 1993, S.68).

Solche Hochleistungsrechner sind jedoch nach Expertenmeinung allenfalls mit parallelverarbeitenden Systemen zu realisieren:

„In den letzten Jahren hat sich die Meinung durchgesetzt, daß nur massiv parallele Rechner (Massively Parallel Processor, MPP) mit mehreren hundert oder tausend Prozessoren das angestrebte Ziel im Teraflops-Bereich (= 1 Million MFLOPS) erreichen können. Die Idee besteht darin, hundert oder tausend Prozessoren gleichzeitig an einer Problemstellung arbeiten zu lassen [...]. Wesentliche Komponente eines jeden Parallelrechners ist ein Kommunikationsnetzwerk, über das die Prozessoren miteinander verbunden sind, um Abläufe zu koordinieren und Daten auszutauschen. Das Netzwerk muß ausreichend schnell sein, damit der Zeitgewinn durch die verteilte Bearbeitung nicht durch den Zeitbedarf für die Kommunikation und das Zusammenfassen der Ergebnisse aufgehoben wird“ (Zobel, 1993, S.68).

Auch nach Meinung von Daniel Hillis, dem Erfinder der Connection Machine²⁷ und somit einem Pionier der parallelen Supercomputertechnik, ist die hohe Bandbreite des Gehirns kaum durch serielle Maschinen abbildbar (siehe Hillis, 1996).

²⁴ die Leistungsfähigkeit eines Computersystems wird üblicherweise in MFLOPS oder MIPS angegeben, wobei MFLOP für „eine Million Fließkomma-Operationen pro Sekunde“ und MIPS für „eine Million Instruktionen pro Sekunde“ steht.

²⁵ dabei wird im Gegensatz zu anderen Forschern die logische Aktivität eines Neurons als Prozeß aufgefaßt, der ungefähr 10000 Eingabe-Bytes mit ca. 40000 Synapsenstatus-Bytes kombiniert; als Geschwindigkeit wird eine Neuronen-Feuerrate von 100 mal pro Sekunde angenommen.

²⁶ Churchland und Churchland (1990) sprechen davon, daß die Leitungsgeschwindigkeit des Gehirns etwa 1.000.000 mal niedriger als die von Computersystemen ist.

²⁷ die *Connection Machine* ist ein hochparallel verarbeitendes Supercomputersystem, welches schon 1985 mit Hilfe von bis zu 65536 Prozessoren mit einer Zugriffsrate von 10^{11} Bits pro Sekunde betrieben werden konnte (nach Hillis, D.W. (1985). *The Connection Machine*, Cambridge: MIT Press; Quelle: Hillis (1996, S.178)).

4.2 Entwicklungsbedingungen des Gehirns

4.2.1 Ontogenese

Bei der Geburt eines Menschen sind zwar bereits alle Nervenzellen voll entwickelt, jedoch verändern sich deren Verbindungen untereinander während des weiteren Lebens noch ständig. Dafür verantwortlich sind drei verschiedene Prinzipien (siehe Thompson, 1990, S.244ff.). Auf das Prinzip des trophischen Wachstums, welches ein Konzept von chemischen Signalen darstellt, und auf das Prinzip des „Bootstrappings“ soll hier nicht weiter eingegangen werden.

Viel interessanter für unsere weiteren Betrachtungen ist ein Prinzip, das bereits 1949 von Hebb (siehe Cowan & Sharp, 1996, S.88; Dreyfus & Dreyfus, 1996, S.18; Schwartz, 1996, S.131f.; Smolensky, 1989, S.57) entdeckt wurde. Diesem Prinzip zufolge besteht ein immerwährender Wettstreit einzelner Neuronenverbindungen, bei dem jeweils diejenige Verbindung bestehen bleibt, welche am häufigsten innerviert wurde. Weniger oft erregte Axonenleitungen ziehen sich dagegen zurück und sterben schließlich ab. Dadurch entsteht ein, auf die Synchronisation von sensorischen und motorischen Neuronen bezogen, hochoptimiertes Nervenfasernetz.

Nur durch dieses Prinzip der Assoziation funktional miteinander verbundener Neuronen, das die gesamte Lebensspanne eines biologischen Systems andauert, ist ein immerwährendes Lernen gewährleistet. Stets werden Nervenfasernverbindungen neu optimiert und verschaltet. Das Gehirn besitzt dadurch einen Lernmechanismus ohne daß es irgendeine explizite Anweisung benötigt (siehe Hinton, 1992).

4.2.2 Phylogenese

Nur kurz sei auf einen interessanten Umstand der phylogenetischen Entwicklung des menschlichen Hirns hingewiesen: Die biologische Evolution läßt Zeiten relativer Stabilität und dann wiederum Perioden plötzlicher Sprünge erkennen. So läßt sich bezogen auf die Entwicklung des menschlichen Gehirns ein sprunghafter Anstieg der Gehirngröße vor ca. 200000 Jahren erkennen (siehe Steels, 1996). Die oft angeführten Klimaschwankungen können jedoch nicht für diese Vervierfachung des Hirnvolumens innerhalb von relativ kurzer Zeit verantwortlich gemacht werden (siehe Calvin, 1998). Als populäre Theorie wird deshalb angeführt, daß sich erst zu dieser Zeit wesentliche soziale, sprachliche und bewußtseinsmäßige Fähigkeiten herausgebildet haben.

4.3 Fähigkeiten und Möglichkeiten des Gehirns

Wie schon in Abschnitt 4.1 erläutert wurde, besitzt das Gehirn die Eigenschaft, ohne explizite Anweisungen zu lernen. Dieses *implizite* Lernen kommt in vielen verschiedenen Bereichen unseres Lebens vor (siehe Roediger & McDermott, 1993). Beispiele dafür sind unsere Fähigkeiten, Sinneseindrücke zu memorieren, Assoziationen zwischen einzelnen Eindrücken und Gedächtnisinhalten aufzubauen oder motorische Feinabstimmungen durchzuführen (vgl. Churchland, 1997).

Der spezifische Aufbau des menschlichen Gehirns bedingt, daß Gedächtnisinhalte nicht symbolisch und diskret, d.h. lokal, gespeichert, sondern subsymbolisch und verteilt repräsentiert werden. Diese Speicherungsart macht eine sehr wichtige und beim Menschen hochentwickelte Eigenschaft erklärbar: Selbst wenn ein Reiz nämlich nur unvollständig oder in veränderter Art und Weise gegenüber einem früher vorgegebenen *Referenzmuster* präsentiert wird, kann, über autoassoziative Prozesse, dennoch der dem Organismus bereits bekannte Stimulus rekonstruiert werden. Die Fähigkeit zur Reizvervollständigung und –generalisation ist somit, aufgrund der spezifischen Organisationsstruktur des Gehirns, bereits von Geburt an vorhanden. Es handelt sich dabei um eine lebenswichtige Fähigkeit, bedenkt man, daß es praktisch keine Reize gibt, die von unveränderlicher Gestalt sind. Oder ist die beruhigende Stimme der Mutter etwa an jedem Tag gleich, der Geruch des Essens stets derselbe oder die Unterschrift auf den eigenen Schecks immer identisch?

Grush & Churchland (1996) gehen hier noch einen wesentlichen Schritt weiter. Für sie könnte man auch Einsicht und Intuition im Sinne von komplexen Mustererkennungsvorgängen verstehen:

„...das, was wir ‚Einsicht‘ und ‚Intuition‘ nennen, könnte sehr gut komplexe Mustererkennung durch rückgekoppelte neuronale Netze sein“ (Grush & Churchland, 1996, S.228).

Neben reinen Mustererkennungsleistungen wird auch die Fähigkeit zur Abstraktion mit der Struktur neuronaler Netze erklärbar. So lernt z.B. ein Kind allmählich den visuellen Unterschied zwischen weiblichen und männlichen Wesen, indem immer wieder, durch Rückmeldungen der richtigen Zuordnung, Assoziationen zwischen dem „generellen“ visuellen Eindruck eines Geschlechtstyps und dem Begriff gebildet werden. Am Ende dieses Prozesses muß das Kind dann nicht mehr mühsam nach einzelnen visuellen Bestimmungsstücken suchen und diese analysieren, denn die Assoziationen zwischen einzelnen Neuronenpopulationen sind derart stark geworden, daß die Geschlechtsbezeichnung quasi „auf der Zunge“ liegt. Anders ausgedrückt werden durch Autoassoziationen bereits bestimmte Bereiche im Hirn selbständig erregt, deren Aktivitätspotentiale eine bestimmte Geschlechtsbezeichnung nahelegen.

Das Prinzip der Autoassoziation ist jedoch nicht nur für die Reizdarbietung von unschätzbarem Wert. Da es immer wieder zu „Unfällen“ im Gehirn kommt (Einblutungen bei hypertensiven Krisen, Quetschungen durch mechanische Einwirkungen, Zellsterben durch Sauerstoffmangel, natürliche Alterungserscheinungen, u.a.), sterben ab und zu einzelne Neuronen oder sogar ganze Gehirnpartien ab. Bei einer konventionellen symbolischen Speicherungsstrategie, wie sie etwa auf üblichen Computern benutzt wird, wären nach einem solchen Neuronenausfall sofort komplette Datenbereiche unwiederruflich verloren. Die Folge wäre, daß komplexe Fähigkeiten oder Erinnerungsmuster dem Neuronentod anheimfallen würden. Tatsächlich kommt es in unserem Kopf jedoch meist nicht zu solchen „Totalausfällen“. Dadurch, daß unsere Fähigkeiten und Erinnerungen durch viele verschiedene Neuronen und deren spezifische Verschaltungen repräsentiert werden, können meist auch mit weniger verbleibenden Neuronen weiterhin sinnvolle Inhalte assoziiert werden, wodurch ein störungsfreies Weiterarbeiten gewährleistet wird (siehe Flohr, 1993; Thompson, 1990).

Ein weiteres Charakteristikum des menschlichen Gehirns ist dessen rekurrenter Verschaltungsaufbau: durch diese rücklaufende Neuronenverbindungen kann eine einfache Art von Kurzzeitgedächtnis ausgebildet werden. Dies impliziert, daß dadurch auch zeitliche Informationen kodiert werden können (vgl. Churchland, 1997).

4.4 Schlußfolgerungen

Das Gehirn stellt sich uns als ein höchst komplexes und nichtlineares System dar, das mit Hilfe von paralleler Verschaltungsweise, autoassoziativer Neuronenverbindungen und rekurrenter Netzstruktur sehr interessante Eigenschaften besitzt (vgl. Warwick, 1998): Es besitzt sowohl eine enorm hohe Arbeitsgeschwindigkeit, die Befähigung zur Fehlertoleranz, Generalisierungskompetenzen als auch die Fähigkeit zur Ausbildung eines Gedächtnisses. Ein ungeheuer vielfältiges Leistungsspektrum steht damit gegenwärtig den relativ eingeschränkten und bescheidenen Ergebnissen der KI gegenüber.

Vielleicht liegt diese deutliche Divergenz an den wesentlichen strukturellen Unterschieden zwischen natürlichen neuronalen Systemen und den Von-Neumannschen-Maschinen. So sind natürliche neuronale Systeme im Gegensatz zu herkömmlichen Computersystemen parallel aufgebaut, die Leitungsinformationen werden analog statt digital überliefert und durch Rückprojektionen existiert eine Modulation sensorischer Verarbeitung (siehe Churchland & Churchland, 1990).

Ob diese Unterschiede jedoch tatsächlich unüberbrückbare Grenzen darstellen zwischen natürlichen Biosystemen auf der Grundlage neuronaler Netze, denen wir gemeinhin Intelligenz und ein Bewußtsein zuschreiben, und Computersystemen, welche auf Von-Neumann-Maschinen basieren, bleibt letztendlich ungewiß. In jedem Fall

läßt die genaue Betrachtung und Analyse des menschlichen Gehirns erahnen, daß dieses Organ vermutlich um einiges komplexer ist, als dies von Vertretern der harten KI allgemein angenommen wird, indem sie postulieren, daß unser Denken im Grunde dasselbe wie die Tätigkeit eines komplizierten Computers sei.

5 Konnektionismus

5.1 Theorie

Aufbauend auf dem durch die Erläuterungen zu traditionellen Computersystemen vermittelten Wissen, wird in den folgenden Abschnitten die Theorie des Konnektionismus vorgestellt und diese mit der Theorie klassischer KI-Systeme kontrastiert. Zu diesem Zweck wird auch auf Teile des Exkurses über den Aufbau und die Funktionsweise des Gehirns rekuriert (siehe Kapitel 4).

5.1.1 Theoretische Vorüberlegungen

Der Konnektionismus baut im Gegensatz zur klassischen KI nicht auf formalen Ansätzen auf, sondern geht eher pragmatisch vor. Während traditionelle KI'ler erst mühsam Modelle für Prozesse bei Phänomenen wie Wahrnehmung, Gedächtnis oder Denken entwerfen müssen, orientieren sich Konnektionisten an generellen Funktionsweisen biologischer Systeme. Sie übernehmen deren grundsätzliche Funktionsweise als Basis für ihre Modellvorstellungen:

“With neural networks an attempt is made to understand roughly, how a brain works from the inside rather than the outside; or perhaps one should say at a much lower level” (Warwick, 1998, S.152).

Oder wie es Frank Rosenblatt, der als erster in der Praxis die Nützlichkeit und Tauglichkeit neuronaler Netze demonstrieren konnte²⁸, formulierte:

„Es ist sowohl einfacher als auch nützlicher, das physikalische System zu axiomatisieren und dann dieses analytisch zu untersuchen, um sein Verhalten zu bestimmen, als das Verhalten zu axiomatisieren und dann ein physikalisches System zu entwerfen mit Hilfe von Techniken logischer Synthesebildung“²⁹

Die Analyseebene in der konnektionistischen Theorie ist somit viel basaler angesetzt als bei klassischen KI-Systemen: Das Interesse gilt der spezifischen Verschaltung relativ primitiver Verrechnungseinheiten, die in Anlehnung an die Gehirnstruktur, *Neuronen* genannt werden. Neuronen werden als die fundamentalen Funktionseinheiten angesehen, die das Gehirn befähigen, so zu operieren, wie wir es gewohnt sind (vgl. Goschke & Koppelberg, 1993; Warwick, 1998). Sie alleine sind hinreichend, um die wesentlichen Eigenschaften des Gehirns erklären zu können. Weiterführende Betrachtungen, wie z.B. die von quantenmechanischen Effekten (siehe Penrose, 1991) scheinen daher überflüssig (vgl. Churchland, 1997).

5.1.2 Historie

Als Pioniere der künstlichen neuronalen Netze gelten Warren McCulloch und Walter Pitts, die 1943 theoretisch zeigten, daß mit Hilfe von Netzwerken, die aus einer Vielzahl von untereinander verschalteten Einheiten bestehen, im Prinzip jede logische oder arithmetische Funktion berechnet werden kann (siehe Crick, 1997; Manhart, 1991). Sie abstrahierten mit ihren Netzen das, was in den 40er und 50er Jahren vom Gehirn bekannt war (siehe Bechtel, 1993a).

In der Folge konnte Frank Rosenblatt zeigen, daß selbst sehr einfache Arten von Netzen, sogenannte *Perzeptronen*, erstaunliche Leistungen in Form von Mustererkennungen vollbringen können (nach Dreyfus & Dreyfus, 1996). Rosenblatt demonstrierte, daß Perzeptronen eine bestimmte Klasse von einfachen Problemen, die *linear separabel* genannt werden, in einer endlichen Anzahl von Lernschritten lösen können.

Der nächste Schritt in der Entwicklung konnektionistischer Systeme gelang John Hopfield mit seinen rekurrenten Netzen, die auch *Hopfield-Netze* genannt werden (Hoffmann, 1992). Das besondere an den Hopfield-Netzen ist, daß sie weder unbe-

²⁸ dies gelang ihm mit seinem *Perzeptron*, das später noch näher ausgeführt wird.

²⁹ Rosenblatt, F. (1962). *Strategic Approaches to the Study of Brain Models*, in: Heinz von Foerster (Hrsg.): *Principles of Self-Organization*, S.386; zit. nach Dreyfus & Dreyfus (1996, S.18).

stimmt umherwandern noch oszillieren, sondern sehr schnell einen stabilen Zustand erreichen (nach Crick, 1997).

Obwohl neuronale Netze zu dieser Zeit bereits theoretisch und praktisch eifrig erforscht wurden, gelang der eigentliche Durchbruch erst mit der Veröffentlichung des Buchs *Parallel Distributed Processing* von David Rumelhart, James McClelland und der PDP-Gruppe im Jahre 1986. In diesem zweibändigen Werk wurde erstmals der *Backpropagation*-Algorithmus vorgestellt, bei dem Fehler in einem *Rückwärtsverfahren* korrigiert werden (Rumelhart, McClelland, & PDP-Group, 1986). Auch die heutigen konnektionistischen Systeme verwenden meist dieses rückwärtsgerichtete Verfahren (siehe Barr, 1991; Kirchner, 1990; Sarnow, 1990b). Näheres zu diesem Algorithmus unter Abschnitt 5.1.3.2.

5.1.3 Theorie

5.1.3.1 Grundtypen

Mallot (1995) unterscheidet grundsätzlich zwischen drei verschiedenen *Grundtypen* von neuronalen Netzen, wobei er zu bedenken gibt, daß sich eine einheitliche Einteilung bisher noch nicht durchgesetzt hat. Sein Unterscheidungskriterium bezieht sich auf die Dimensionalität von Input und Output.

Demnach stellen *klassifizierende Netzwerke*, die auch als *Perzeptronen* bezeichnet werden, den einfachsten Typ neuronaler Netze dar: Sie teilen den Raum der Inputvektoren lediglich in Klassen ein und können dementsprechend nur die Outputzustände 0 oder 1 detektieren.

Die Gruppe der *Assoziatoren* ordnen bestimmten Inputvektoren spezifische Outputvektoren zu, wobei die Anzahl der assoziierten Paare relativ klein ist.

Schließlich implementieren *abbildende Netzwerke* eine Zuordnung zwischen beliebigen Inputvektoren und zugehörigen Outputvektoren. Mit solchen Zuordnungsvorschriften sind bereits Approximationen von Funktionen und Operatoren möglich.

Zusätzlich existieren zu diesen drei Haupttypen weitere Untertypen, die hier aber nicht einzeln besprochen werden sollen (weiteres siehe Mallot, 1995).

5.1.3.2 Typen von Lernregeln

Ein anderes wichtiges Unterscheidungsmerkmal stellt der Typ der implementierten Lernregel dar. Im Laufe der Entwicklungsgeschichte neuronaler Netze existierten viele verschiedene Ansätze für die Implementation von Lernregeln (siehe Smolensky, 1989); heute wird meist die sogenannte *Delta-Regel* eingesetzt. Eine Generalisation dieser Regel stellt der *Backpropagation*-Algorithmus dar (vgl. Frey & Schach, 1996; Sarnow, 1990b). Zur weiteren Vertiefung zum Thema Lernregeln sei auf die entspre-

chende Fachliteratur verwiesen (z.B. Crick, 1997; Mallot, 1995), da eine nähere Erläuterung den Rahmen dieser Arbeit sprengen würde.

5.1.4 Der Vergleich zu natürlichen neuronalen Netzen

Reeke und Edelman (1996) drücken die Anlehnung konnektionistischer Systeme an das Gehirn folgendermaßen aus:

„Weil konnektionistische Modelle ihre Inspiration und einen großen Teil ihrer Terminologie aus den Neuronennetzen lebender Organismen ziehen, sagen wir, daß sie ‚seitwärts nach der Biologie schielen‘, aber sie stellen keine Modelle von Neuronennetzen dar (und sie sind auch nicht als solche gedacht)“ (Reeke & Edelman, 1996, S.149, Klammern i. Original).

Natürliche neuronale Systeme sollen also durch künstliche neuronale Netze nicht genau *imitiert* werden; insbesondere auf eine chemisch–physikalische Imitation soll verzichtet werden. Vielmehr soll die Arbeitsweise der biologischen Systeme als Vorbild für eigenständige Entwicklungen benutzt werden (vgl.Hoffmann, 1992).

Konnektionistische Systeme weisen, wie auch ihre biologischen Verwandten, Redundanz auf: Ein Bit wird nicht mehr nur von einem, sondern gleich durch viele verschiedene Neuronen repräsentiert, welche jedoch jeweils nicht ausschließlich nur ein *bestimmtes* Bit speichern, (Cowan & Sharp, 1996, S.87). Dies hat zur Folge, daß lediglich größere Ansammlungen von Neuronen sinnvoll interpretiert werden können, einzelne Neuronen bleiben dagegen in ihrem Repräsentationsgrad unbestimmt. Das Prinzip der Redundanz verhilft dem Gesamtsystem zu einer hohen Verlässlichkeit, selbst wenn einzelne Neuronen ausfallen sollten.

Da man aufgrund der verteilten Speicherung nicht mehr von gespeicherten *Symbolen* oder *diskreten Daten* sprechen kann, bezeichnet man neuronale Netze auch gerne als *subsymbolische* Systeme. Auf dieser subsymbolischen Ebene sind die einzigen sinnvoll interpretierbaren semantischen Entitäten „Muster von Aktivationen“ (Smolensky, 1989, S.54), wohingegen die Entitäten, die durch formale Regeln manipuliert werden, die individuellen Aktivationen der Zellen innerhalb des Netzwerks darstellen.

In einem künstlichen neuronalen Netz wird jeder Verbindung von zwei Neuronen eine bestimmte Verbindungsstärke in Form eines „*Verbindungsgewichts*“ zugesprochen, wobei ein hohes Gewicht eine enge Verbindung, und ein niedriges Gewicht eine entsprechend geringe Assoziation symbolisiert. Eine Null als Verbindungsgewicht würde einer nicht vorhandenen Verbindung zwischen zwei natürlichen Neuronen entsprechen. Die *Aktivationsstärke* eines künstlichen Neurons kommt nun durch ein summatorisches Verrechnen aller auf dieses Neuron eintreffenden Einzelaktivitäten zustande. Die Einzelaktivitäten wiederum werden üblicherweise berechnet durch eine einfache

Multiplikation von Verbindungsgewicht und dem efferenten Aktivitätsniveau. Ist die Summe aus allen auf afferenten Bahnen einströmenden Aktivitäten berechnet, so wird auf diese die sogenannte *Transferfunktion*³⁰ angewendet. Sie entscheidet letztendlich darüber, wie stark das Neuron aktiviert wird.

Diese Auflistung von Eigenschaften künstlicher neuronaler Netze macht indessen deutlich, daß sie lediglich eine Abstraktion natürlicher Systeme darstellen. So werden einige Aspekte biologischer Systeme in vereinfachter Weise umgesetzt, um die Kontrollierbarkeit und Handhabbarkeit zu erhöhen: Während z.B. die Anzahl und die Verbindungsbedingungen der Neuronen in einem biologischen System über die Lebensspanne hinweg variabel sind, wird bei konnektionistischen Modellen normalerweise eine feste Verschaltungsstruktur vorgegeben, bei der lediglich die Verbindungsgewichte geändert werden können (vgl. Tetens, 1994).

5.1.5 Probleme konnektionistischer Systeme

Trotz der vielen Vorteile, die konnektionistische Systeme gegenüber klassischen KI-Maschinen besitzen, sei jedoch auch auf typische Beschränkungen solcher Systeme hingewiesen.

Zum einen sind alle bisher implementierten Netze äußerst einfach gehalten, d.h. sie erreichen nicht annähernd die Komplexität von natürlichen Systemen (vgl. Dreyfus & Dreyfus, 1996; Tetens, 1994). Zum anderen scheint ein paralleles Verarbeitungsverfahren nicht für alle Berechnungsarten ideal zu sein. Bei Aufgaben nämlich, die nur kleine Eingabevektoren besitzen, dafür aber Millionen von schnell iterierten rekursiven Berechnungen erfordern, scheinen konnektionistische Systeme überfordert zu sein. Hier können herkömmliche serielle Maschinen eine deutlich bessere Performance erreichen (vgl. Churchland & Churchland, 1990). Da die Klasse dieser Aufgabe sehr umfangreich und wichtig ist, scheinen demnach klassische Computersysteme weiterhin unverzichtbar; konnektionistische Systeme können in diesem Bereich kaum überzeugen.

Andererseits handelt es sich dabei auch genau um diejenigen Probleme, bei deren Lösung auch das menschliche Gehirn Schwierigkeiten hat. Insofern scheint das Aufzeigen der Grenzen konnektionistischer Systeme gleichzeitig eine Untermauerung der Ähnlichkeit zu biologischen Systemen darzustellen.

³⁰ die Transferfunktion wird entweder in Form eines Schwellenwertes ausgedrückt, oder aber ist linearer oder sigmoider Art, wobei sigmoide Funktionen dem Aktivitätsverhalten natürlicher Neuronen noch am nächsten kommen; aber auch sigmoide Abbildungen stellen ebenfalls nur eine grobe Annäherung an natürliche Bedingungen dar.

5.2 Erfolgreiche Implementierungen

Um die weitreichende Praxistauglichkeit neuronaler Netze zu demonstrieren, sollen exemplarisch einige Anwendungsbeispiele genannt werden.

5.2.1 Muster- und Zeichenerkennung

Die erste praktische Demonstration der Nützlichkeit und Tauglichkeit eines künstlichen neuronalen Netzes führte Frank Rosenblatt im Jahre 1956 durch. Sein System, das er *Perzeptron* nannte, war so eingerichtet, daß es bestimmte Arten von Mustern als ähnlich klassifizieren und von anderen Typen unterscheiden konnte (Dreyfus & Dreyfus, 1996, S.20). Rosenblatt konnte mit seinem Perzeptron zeigen, daß typische Probleme menschlicher Kognition wie das der Wahrnehmung mit statistischen Mitteln effektiver abbildbar sind als mit den Methoden der damaligen „konservativen“, alleine mit logischen Prinzipien operierenden, KI-Forschung (wie etwa Newell und Simon).

Tatsächlich stellen bei heutigen Schrifterkennungssystemen neuronale Netze einen integralen Bestandteil³¹ dar, die bedingt durch ihre Fehlertoleranz für das Erkennen von nicht konstantem Reizmaterial geradezu prädestiniert sind (vgl. Beneke & Schwippert, 1995; Hoffmann, 1992; Sarnow, 1990a; Sarnow, 1996).

Komplizierte Erfassungssysteme zur Gesichtererkennung (z.B. Malsburg, 1996) oder zum Erfassen von KFZ-Nummernschildern sind ebenfalls bereits in der Praxis erfolgreich erprobt worden (vgl. Bertuch, 1992). So erkennt ein System zur Gesichtskontrolle in über 90 % der Fälle berechnigte Personen beim Zugang in einen berechnigungsbedürftigen Bereich. Selbst „Brille noch der 3-Tage-Bart können das System irritieren“ (Bertuch, 1992). Auch ein KFZ-Erfassungssystem, bei dem über eine Videokamera und ein nachgeschaltetes neuronales Netz, KFZ-Nummernschilder erkannt werden, ist bereits erfolgreich in einem Parkhaus installiert worden (Bertuch, 1992).

5.2.2 Sprachemulation

Ein immer wieder gern zitiertes Beispiel für erfolgreiche Umsetzung neuronaler Netz-Strategien ist das System *NETtalk*, das bereits in Kapitel 1 kurz charakterisiert wurde. Dieses System wurde mit einem willkürlichen englischen Text von 1000 Wörtern Länge trainiert. Nach einer nur zehnstündigen (!)³² Lernphase konnte das System den vorgegebenen Text mit einer Genauigkeit von ca. 95% lesen und sprechen, d.h. nur 5% der Wörter wurden nicht völlig korrekt ausgesprochen (siehe Churchland, 1997;

³¹ Beispiel z.B. in: Barr (1991).

Cowan & Sharp, 1996). Selbst beim Lesen von neuen Texten machte das System nur 22% Fehler. Die meisten Fehler waren zudem nur kaum wahrnehmbare phonetische Unkorrektheiten, so daß der Sinn des Textes im Gesamtzusammenhang recht gut verstanden werden konnte. Nachfolgendes Training mit Hilfe längerer Texte steigerte schließlich die Leistungsfähigkeit von *NETtalk* auf 97,5% korrekt ausgesprochene Worte.

Dagegen programmierten die Entwickler des Konkurrenzproduktes *DECtalk*, welches ebenfalls für die Aussprache englischer Texte konzipiert worden war, explizit jedes einzelne Ausspracheproblem und jede Sonderregel in jahrelanger Kleinarbeit (vgl. Görz, 1995). Letztendlich verfügte *DECtalk* jedoch dennoch nicht über die von *NETtalk* systemimmanenten Generalisierungsfähigkeiten. Auch unvollständige Reizdarbietungen konnten von *NETtalk* noch einigermaßen verständlich weiterverarbeitet werden. *DECtalk* war dagegen auf vorangegangene explizite isomorphe Beispiele angewiesen (siehe Crick, 1997)³³.

5.2.3 Robotik

Grundsätzliche Probleme in der Robotersteuerung stellen die Ungenauigkeiten der Positionssteuerung und deren Detektion dar. Obwohl der Mensch bestimmte Positionen nur sehr ungefähr mit seinen Gliedmaßen „ansteuern“ kann, gelingt es ihm dennoch, mit Hilfe von schnellen Strategiewechseln und feinmotorischen Adaptionprozessen, sehr genaue handwerkliche Arbeiten durchzuführen. So stellt z.B. ein Steckspiel für ein Kind oder das Zusammenbauen eines Motors für eine gelernte Fachkraft im Prinzip keine große Schwierigkeit dar.

Demgegenüber verfügt ein Roboter zwar theoretisch über viel genauere Positioniermöglichkeiten, jedoch können störende Einflüsse wie mechanische oder thermische Einwirkungen von außen (z.B. durch im Weg stehende Gegenstände oder Veränderung der Raumtemperatur), sowie systemimmanente Variablen, wie Steifigkeit des Materials der Greiferarme oder wechselnde Reibungskoeffizienten durch ungleiche Schmierung, schnell ein rein mechanisch arbeitendes System ohne Eigenintelligenz in Positionierungsschwierigkeiten bringen.

Daher muß ein für paßgenaues Arbeiten eingesetzter Roboter über Kraft- und Drehmomentsensoren verfügen, mit deren Hilfe die Bewegungsabweichungen korrigiert werden können (Barr, 1991). Jedoch haben die Sensorendaten einen komplizier-

³² um so erstaunlicher bei dieser so kurzen Lernphase ist der Umstand, daß das Einstellen der Verbindungsstärken von einem aus heutiger Sicht äußerst leistungsarmen System durchgeführt wurde, welches höchstens in der Leistungsstärke eines heute üblichen Heimcomputers liegt.

ten, nichtlinearen Einfluß auf die Geschwindigkeit und den Drehwinkel des Greiferarmes. Dies führt zu komplexen, schwer berechenbaren Gleichungen, die mit großem mathematischen Aufwand und Know-how berechnet und den Robotern in Form von Korrekturalgorithmen zur Verfügung gestellt werden müssen.

Ein neuronales Netz dagegen kann durch

„Erlernen einiger ausgewählter Wertepaare für Ein- und Ausgabewerte auf den allgemeinen Zusammenhang zwischen Bewegung und Bewegungskorrektur verallgemeinern und damit Steuerungsvorgänge *in den Griff* bekommen“ (Barr, 1991, Herv. i. Original).

Dadurch kann auf den aufwendigen Zwischenschritt der Arithmetrisierung vollständig verzichtet werden.

5.2.4 Qualitätskontrolle

Der Sinn von Qualitätskontrollen besteht darin, gute und schlechte Ware möglichst scharf kategorisieren zu können. Eine effiziente Möglichkeit, z.B. gute von schlechten Elektromotoren zu unterscheiden, besteht in der Analyse ihrer Laufgeräusche. Die für Menschen sehr streßreiche und ermüdende Arbeit, wurde am amerikanischen Forschungszentrum der Siemens AG in Princeton einem neuronalen Netz überantwortet (Eckmiller, 1994). Die aufgezeichneten Schallwellen werden in einzelne Frequenzbänder zerlegt und in die Eingabeknoten eines konnektionistischen Systems eingespeist. Nach nur einem Tag Training leistet das Computersystem mindestens gleichwertige Leistungen wie ein ca. sechs Wochen ausgebildeter menschlicher Tester.

5.2.5 Erkennung von dreidimensionalen Körpern

Das System *Fusion.net* ist ein auf das räumliche Sehen spezialisiertes Netzwerk. Es besitzt 36.000 Zellen und insgesamt 50.400 Verbindungen (nach Churchland, 1997). Als Eingabevektor verlangt es zwei 60 x 60 große Graustufen³⁴-Matrizen, die das zu beurteilende stereoskopische Bild-Paar repräsentieren. Diese Eingabevektoren entsprechen in sehr vereinfachter Weise den beiden retinalen Bildmustern eines Menschen. Der von *Fusion.net* generierte Output besteht aus drei Schwarz-Weiß-Matrizen, die jeweils verschiedene Entfernungsbereiche repräsentieren (Nah-, Fixierungs- und Fernbereich).

Mit Hilfe dieses Netzaufbaus gelingt, trotz der hohen Körnigkeit der Ausgangsbilder, eine recht exakte Auswertung der Dreidimensionalität (vgl. Churchland, 1997). Ein

³³ Weitere Beispiele zum Thema Spracherkennung finden sich in Hoffmann (1992).

³⁴ Es werden 10 verschiedene Graustufen verwendet (nach Churchland, 1997).

ähnliches konnektionistisches System zur Rekonstruktion von dreidimensionalen Objekten aus reinen Schattierungsinformationen, das interessanterweise eine hohe Ähnlichkeit zum visuellen Areal V1 im menschlichen Gehirn aufweist, findet sich in Crick (1997).

5.2.6 Neuroprothetik

Der Bereich der Neuroprothetik ist noch recht jung und derzeitige Forschungsvorhaben fokussieren vor allem den Bereich der Hör- und Seh Wahrnehmung. Allerdings liegen noch nicht genügend lauffähige Systeme zur Verfügung, um sich ein umfassendes Urteil über diesen Bereich zu erlauben (vgl. Hamilton, 1996).

Generell besteht der Zweck dieser Forschungsrichtung im Ausgleichen von Defekten des auditorischen und visuellen Apparat durch implantierte Elektronik (vgl. Reiter, 1996a). Nach Aussage von Patrick Hamilton wurden „beim Ohr schon gute Erfahrungen gesammelt“ (Hamilton, 1997). Aufgrund der um einiges komplexeren Struktur des visuellen Apparats steht ein durchgreifender Erfolg von Augenprothesen vermutlich allerdings noch länger aus (siehe Ebeling, 1995).

5.2.7 Börsenvorhersage

Die Erstellung von Prognosen im Finanzbereich basiert auf Auswertungen von Informationen verschiedenster Quellen. Wesentliche Indikatoren stellen die Meinungen von Börsenexperten dar, deren Wissen allerdings nicht vollständig formalisierbar ist, da diese oft rein intuitiv vorgehen.

Börsenprognosen zählen deswegen eindeutig zu den sogenannten *nicht-konservativen* Domänen, die durch prozedurale oder wissensbasierte Programme kaum verbessert werden können. Daher werden in diesem Bereich verstärkt neuronale Netze eingesetzt, die nach Mustern in den ihnen präsentierten Daten suchen (vgl. Birkelbach, 1995; Hoffmann, 1992):

„Diese [künstlichen neuronalen Netze] können sich beispielsweise aus historischen Kursen, technischen Indikatoren, fundamentalen Daten, Informationen aus anderen Markt Bereichen oder weiterem, für eine Prognose relevantem (sic!) Material zusammensetzen. Sind solcherart Daten für einen hinreichend großen Zeitraum verfügbar, dann läßt sich mit ihnen ein KNN [künstliches neuronales Netz] trainieren“ (Kühn, 1991, Erg.v.Verf.).

Gegenüber klassischen Verfahren zeigen sich verschiedene positive Merkmale von neuronalen Netzen (nach Siebers & Weigert, 1998):

- Die Ursache–Wirkungszusammenhänge müssen nicht bekannt sein; es genügt, das Netz mit Beispielen zu trainieren;
- Bei den denkbaren Einflußfaktoren können technische und fundamentale Faktoren beliebig gemischt werden;
- Es können auch hochgradig nichtlineare Strukturen erkannt werden.

Trotz einer Reihe noch nicht optimal gelöster Probleme bei der Erstellung von Prognosesystemen mit Hilfe künstlicher neuronaler Netze sind die bereits erstellten Anwendungen deutlich leistungsstärker als ihre statistischen Vorläufer (siehe Kühn, 1991; Siebers & Weigert, 1998).

5.2.8 Kreditzuschreibung

Als letztes Beispiel soll ein System zur automatischen Entscheidung über die Vergabe von Krediten vorgestellt werden (vgl. Eckmiller, 1994), das von allen hier vorgestellten Beispielen wohl das heikelste ist. Bei den meisten anderen Entwicklungen ging es mehr oder weniger um die Verbesserung stupider und fehlerträchtiger Arbeitsmethoden, die in ihrer direkten Verwendung kaum zur Sorge aufrufen können, wenn man von dem schwerwiegenden Problem der Einsparung von menschlichen Arbeitskräften durch solche Systeme einmal absieht. Ein Computersystem im Bereich der Kreditzuschreibung dringt dagegen direkt in soziale und moralische Bereiche ein.

Bei der Vergabe von Krediten interessiert von unternehmerischer Seite aus gesehen vor allem, ob der Kreditnehmer den aufgenommenen Kredit tatsächlich wieder zurückzahlen wird. Zur besseren Entscheidung für die Kreditvergabe werden daher üblicherweise Informationen wie Kreditbedarf, Kreditwürdigkeit und Familienverhältnisse analysiert.

Die amerikanische Firma Nestor hat zum Zweck der Automatisierung dieses Verfahrens ein neuronales Netz mit 200 bereits vollzogenen Kreditvorgängen trainiert. Es stellte sich heraus, daß das System, das lediglich diese 200 Datenpakete als Wissensbasis besaß, in etwa 25% der Fälle zu besseren Entscheidungen kam, als es seinerzeit die zuständigen Kreditberater taten. Einige Kreditanträge, die sich später als monetärer Verlust für die Bank herausgestellt hatten, wurden „richtigerweise“ vom Computersystem zurückgewiesen³⁵. Der materielle Schaden wäre bei Verwendung dieses Computersystems beträchtlich geringer ausgefallen. Allerdings entstehen große Probleme,

³⁵ Leider schweigt sich die Quelle allerdings über die Basisrate der jeweils zurückgewiesenen Kreditanträge aus; sollte das konnektionistische Kreditvergabesystem lediglich weniger Anträge akzeptieren, so könnte der erzielte Effekt ein reines Artefakt der Basisrate darstellen.

wenn nach der Begründung, der durch das neuronale Netz erzeugten Ergebnisse, gefragt wird. Die Akzeptanz solcher Systeme scheint alleine dadurch schon stark gefährdet (Frey & Schach, 1996).

Zu diesem Beispiel sei außerdem noch kritisch angemerkt, daß es zu bedenken gilt, ob erstens die Eingangsparameter wirklich die aussagekräftigsten Informationen zur sicheren Vergabe von Krediten bereitstellen und zweitens, ob sich eine Bank tatsächlich bei so einem heiklen Bereich wie der Kreditvergabe auf ein Computersystem stützen sollte. Immerhin können von Krediten nicht nur die Überlebensfähigkeit von Firmen und damit das Wohl ihrer Angestellten, sondern auch das Schicksal einzelner Menschen abhängen, die aus, durch das System nicht erfaßten Gründen, in eine monetäre Zwangslage gekommen sind. Zweifellos sollte die Beachtung solcher Kriterien einen wesentlichen Teil des Vergabeverfahrens darstellen.

5.3 Was ist das neue an künstlichen neuronalen Netzen?

Die wesentlichen Unterschiede von künstlichen neuronalen Netzen gegenüber klassischen KI-Systemen wurden in dieser Arbeit bereits intensiv erörtert: Nur kurz zu erinnern sei an die starke Anlehnung an das Gehirn³⁶, an den parallelen Aufbau (vgl. Mallot, 1995), die Lernfähigkeit (siehe Crick, 1997), Generalisierungsbefähigung, Vereinfachungstendenzen (siehe Smolensky, 1989), Fehlertoleranz und Assoziationsfähigkeit (siehe Churchland, 1997; Churchland & Churchland, 1990).

Es wurde dargestellt, daß das Konzept der neuronalen Netze einen „Abschied von alten Freunden“ bedeutet: Informationen werden nicht mehr diskret und symbolisch, sondern *verteilt* und *subsymbolisch* repräsentiert (vgl. Bechtel, 1993a). Außerdem benötigen konnektionistische Systeme keine expliziten Regeln, sondern lediglich Beispiele von Input-Output-Paaren, um prozedurale Informationen zu verwerten (siehe Bechtel, 1993a; Searle, 1996). Dieses fallbezogene und nicht regelbasierte Lernverhalten kommt dem Lernverhalten des menschlichen Gehirns näher. Churchland & Churchland (1990) erachten deswegen die regelgesteuerte Symbolmanipulation lediglich als *eine von vielen* kognitiven Fähigkeiten, die ein neuronales Netz erlernen kann: „Sie [= regelgeleitete Symbolmanipulation] ist jedenfalls nicht sein [= Gehirn] grundlegender Modus operandi“ (Churchland & Churchland, 1990, S.53, Erg.v.Verf.).

All diese Erfahrungen, die Wissenschaftler mit konnektionistischen Systemen machen konnten, ermöglichen es, sich den Begriffen *Regel* und *Repräsentation* neu zu nähern.

³⁶ Die starke Anlehnung an das Gehirn geht dabei deutlich weiter als der Ansatz der *fuzzy logic*, da bei konnektionistischen Systemen nicht nur die „unscharfe Logik“ des Gehirns Anwendung findet, sondern auch dessen spezifische Speicherungsart verwendet wird (vgl. Crick, 1997).

Ungeprüfte Selbstverständlichkeiten, wie z.B., daß Regeln explizierbar sein müssen, um sie zu verwenden, oder die Ansicht, daß Repräsentationen stets symbolisch sind, müssen daher neu reflektiert werden (vgl. Hatfield, 1993; Searle, 1996). Auch die oft als zwingend erachtete Verbindung zwischen *Denken* und *Sprache* wird durch die Verwendung konnektionistischer Systeme aufgehoben (siehe Bechtel, 1993a).

5.4 Ressentiments gegenüber künstlichen neuronalen Netzen

Trotz der von mir zahlreich aufgezählten Vorteile und sinnvollen Einsatzgebiete neuronaler Netze, war die Aufnahme solcher Systeme in den allgemeinen Wissenschaftsapparat bis zur Mitte der 80er Jahre nur sehr bescheiden. Interessanterweise sind die Gründe hierfür hauptsächlich *nicht* in möglichen technischen Unzulänglichkeiten konnektionistischer Systeme begründet. Vielmehr schien der Konnektionismus lange Zeit einfach nicht in die Wissenschaftslandschaft des beginnenden Computerzeitalters und der in den 50er Jahren aufkommenden KI zu passen. Außerdem entstanden durch die Theorie des Konnektionismus Divergenzen zur etablierten symbolorientierten philosophischen Tradition. Neben diesen Hauptfaktoren waren allerdings auch wissenschaftspolitische Rängeleien und Handhabungsschwierigkeiten der neuronalen Netze dafür verantwortlich.

Nur die Summe all dieser Faktoren kann das äußerst retardierende Moment in der Entwicklungsgeschichte der neuronalen Netze erklären.

5.4.1 Kritiker aus der klassischen KI-Gemeinde

Schon im Jahre 1956, in der Zeit als Newell und Simon ihren GPS (General Problem Solver) vorstellten (nach Beckermann, 1994), konnte Frank Rosenblatt ein funktions-tüchtiges neuronales Netz zur Mustererkennung vorweisen. Euphorisch bemerkte er 1958:

„Es scheint klar, daß... das Perzeptron eine neue Art von informati-onsverarbeitendem Automat einführt: zum ersten Mal haben wir eine Maschine, die in der Lage ist, neuartige Ideen zu entwickeln [...] Als Konzept scheint das Perzeptron zweifellos die Durchführbarkeit und das Prinzip der nicht-menschlichen Systeme zu besitzen, die mensch-liche, kognitive Funktionen verkörpern können.“³⁷

Nach Einschätzung von Dreyfus & Dreyfus (1996, S.21) sahen beide Ansätze, der parallele von Rosenblatt und der serielle von Newell und Simon, zu dieser Zeit noch gleich vielversprechend aus. Beide Ansätze erhoben übertriebene Ansprüche und waren so gleichermaßen angreifbar. Aber ein entscheidender kritischer und letztendlich vernichtender Artikel gegenüber Rosenblatts Konzeption, verfaßt von Marvin Minsky und Seymour Papert, veränderte nachhaltig negativ die Bedingungen für Forschungen über parallelverarbeitende Systeme. Schon um das Jahr 1965 begannen Minsky und Papert, die selbst Leiter eines KI-Laboratoriums am MIT waren, das sich dem symbolmanipulierenden Ansatz widmete, eifrig gegen den Konnektionismus zu agitieren. So werteten sie den Großteil der Publikationen über neuronale Netze pauschal mit der Aussage „ohne wissenschaftlichen Wert“³⁸ ab.

Tatsächlich konnten sie zwar in einer ihrer Publikation beweisen, daß einschichtige Perzeptronen, so wie sie Rosenblatt unter anderem verwendete, schon einfache Probleme wie das XOR-Problem³⁹ nicht lösen können. Jedoch ignorierten sie sowohl Rosenblatts Konzeption mehrschichtiger neuronaler Netze, als auch seinen Beweis der Konvergenz eines auf Wahrscheinlichkeit beruhenden Lernalgorithmus. Zusätzlich wurde die Möglichkeit der Rückwärtspropagierung nicht in diese Analyse mit einbezogen (siehe Bechtel, 1993a).

Die zu kurzgegriffene und polemische Kritik von Minsky und Papert wurde jedoch trotz ihrer augenfälligen Schwächen schnell populär und in der Folgezeit immer wie-

³⁷ Rosenblatt, F. (1958). *Mechanisation of Thought Processes*, in: Proceedings of a Symposium held at the National Physical Laboratory, London: Her Majesty's Stationery Office, Vol.1, S.449; zit. nach Dreyfus und Dreyfus (1996, S.20).

³⁸ Minsky, M., Papert, S. (1969). *Perceptrons*. Cambridge: MIT Press, S.4; zit. nach Dreyfus & Dreyfus (1996, S.22).

³⁹ näheres zum XOR-Problem in Papert (1996).

der für das Argument kolportiert, daß neuronale Netze generell unfähig seien, mathematische Funktionen wie Parität zu berechnen. Damit waren sie wertlos für die Wissenschaft der künstlichen Intelligenz; ihre weitere Erforschung blieb dementsprechend fast komplett aus:

„Um 1970, soweit die KI betroffen war, waren die neuronalen Netze tot [...] Rosenblatt wird in... [wichtigen] historischen Darstellungen des KI-Feldes noch nicht einmal erwähnt“ (Dreyfus & Dreyfus, 1996, S.25).

Die heftigen und oftmals nicht besonders sachlich geführten Angriffe der KI-Gemeinde gegen konnektionistische Ansätze scheinen aber nicht nur aus fachlichen Gründen geführt worden zu sein. Vielmehr spielten nicht allzu selten wissenschafts- und machtpolitische Gründe eine wesentliche Rolle: Etablierte KI-Forscher wendeten sich alleine schon deswegen von der konnektionistischen Herangehensweise ab, da sie befürchteten, ihre Förderungsmittel und wissenschaftlichen Positionen könnten dadurch gefährdet werden. So hatten auch Minsky und Papert zur Zeit der größten Verunglimpfungen neuronaler Netze, leitende Stellungen in einem klassischen KI-Institut inne; die Vermutung scheint daher nicht fern zu liegen, daß auch sie nicht vollständig unabhängig von jenen macht- und forschungspolitischen Beweggründen gewesen sind (siehe Dreyfus & Dreyfus, 1996).

5.4.2 Symbolverarbeitung als Eckpfeiler westlicher Philosophie

Heidegger zufolge ist die „traditionelle Philosophie von Anfang an durch ihre Konzentration auf Tatsachen in der Welt definiert, während sie die Welt als solche ‚übergeht‘“⁴⁰. Durch Theorienbildung soll die Welt verstanden werden, wobei jede Theorie Beziehungen zwischen kontextfreien und objektiven Elementen in Form von abstrakten Prinzipien formuliert. Die Praxis der Theorienbildung, die es als selbstverständlich ansieht, den alltäglichen Zusammenhang menschlicher Aktivität zu ignorieren und zu verzerren, indem sie versucht, deren abstrakte Prinzipien zu extrahieren, ist durch Sokrates und Plato als fester Bestandteil in unser westliches Denken eingegangen. Descartes und Leibniz schließlich erweiterten das Prinzip der Symbolverarbeitung über den Bereich der *apriorischen* Ideen hinaus, indem sie den Bereich der Alltagspraxis ebenfalls als generell theoretisierbar charakterisierten:

⁴⁰ Heidegger, Martin: *Sein und Zeit*; zit. nach Dreyfus & Dreyfus (1996, S.26).

„Die wichtigsten Beobachtungen und Veränderungen der Geschicklichkeit in allen Arten von Handel und Berufen sind bis jetzt ungeschrieben. Diese Tatsache wird durch die Erfahrung bestätigt, daß man von der Theorie zur Praxis übergeht, etwas *hinzufügen wünscht*. Natürlich wir können auch diese Praxis beschreiben, insofern sie *im Grunde nur eine andere, komplexere und besondere Theorie* ist“⁴¹.

Diese Selbstsicherheit in der Möglichkeit der erfolgreichen Anwendung von Theorien auf alle möglichen Gegenstandsbereiche ist nur durch die damaligen dramatischen und bahnbrechenden Erfolge der Naturwissenschaften zu begreifen. In dieser „Aufbruchstimmung der Wissenschaften“ schienen plötzlich alle möglichen Bereiche formalisierbar zu sein. Der Weg dorthin besteht darin, die kontextfreien Elemente und Prinzipien zu isolieren und eine formale, symbolische Repräsentation auf dieser theoretischen Analyse aufzubauen. Diese Denkweise ist bei führenden KI-Wissenschaftlern, wie z.B. Terry Winograd, wiederzuerkennen:

„Wir sind damit beschäftigt, einen Formalismus zu entwickeln oder eine *Repräsentation*, mit der man [...] Wissen beschreiben kann. Wir suchen nach den *Atomen*, und *Partikeln*, aus denen es zusammengesetzt ist und den *Kräften*, die auf sie einwirken“⁴².

Das übliche Vorgehen in der KI bestand und besteht darin, relativ simple und isolierte Teilsysteme zu betrachten, und diese zu axiomatisieren. Auf dieser Grundlage des Forschens wurde jeder Versuch, die Welt holistisch zu interpretieren, *a priori* abgelehnt. Der holistisch operierende Konnektionismus kam daher in die breite Schußlinie traditioneller Forschergruppen. Als Rechtfertigung ihres abweisenden Verhaltens, wurden oft alltagspsychologische Argumente angeführt. So sind mentale Zustände in der Alltagspsychologie grundsätzlich semantisch, also propositional, individuiert (siehe Keil, 1993) und Denken wird mit Sprache gleichgesetzt (siehe Penrose, 1991). In KI-Fachkreisen wurde in diesem Zusammenhang auch gerne auf Wittgensteins *Tractatus logico-philosophicus* verwiesen, welches einen atomistischen Standpunkt einnimmt. Was diesen Forschern jedoch scheinbar nicht bekannt war, ist der Umstand, daß Wittgenstein, nachdem er jahrelang vergebens nach den atomistischen Tatsachen und grundlegenden Gegenständen gesucht hatte, seinen *Tractatus* und damit auch seine rationalistische Philosophie selbst aufgab! Die „vernichtende Kritik seines Tractatus“ (Dreyfus & Dreyfus, 1996, S.28), seine *Philosophischen Untersuchungen*, erschien 1953, also zu jener Zeit, als sich die klassische KI gerade zu formieren begann.

⁴¹ Leibniz: *Selections*; zit. nach. Dreyfus & Dreyfus (1996, S.27), Herv.v.Verf.

⁴² Winograd, T. (1976). Artificial Intelligence and Language Comprehension, in: *Artificial Intelligence and Language Comprehension*, S.9; zit.nach. Dreyfus & Dreyfus, 1996, S.27, Herv. v. Verf.

Wittgensteins späterer Ansicht nach ist die Zerlegung von alltäglichen Situationen in Tatsachen und Regeln nur in einigen bestimmten Kontexten und für einige wenige eng umgrenzte Zwecke sinnvoll:

„Wenn wir versuchen, die letzten kontextfreien, zweckfreien Elemente zu finden, — entsprechend der Notwendigkeit, die elementaren Symbole zu ermitteln, die wir benötigen, um einen Computer zu füttern —, sind wir in Wirklichkeit damit beschäftigt, Aspekte unserer Erfahrung von genau der pragmatischen Organisation zu befreien, die es möglich macht, sie für den Umgang mit alltäglichen Problemen auf intelligente Weise zu gebrauchen“ (zit. nach Dreyfus & Dreyfus, 1996, S.28).

Gerade das Durchbrechen festgefahrener, auf Symbole und Theorienbildung fixierter Wissenschaftsstrukturen (vgl. Hatfield, 1993), scheinen Rosenblatt und seine Kollegen in wissenschaftlichen Mißkredit gebracht zu haben. Selbst heute noch setzt man die als legitim erachtete Zugehörigkeit zur *scientific community* leichtsinnig aufs Spiel, wenn man den Symbolverarbeitungsansatz nicht als Forschungsparadigma verwendet (siehe Searle, 1993).

Anders ist es kaum zu erklären, daß Rosenblatts Ideen erst Mitte der 80er Jahre wieder ernsthaft aufgenommen und weiterentwickelt wurden⁴³.

5.4.3 Schwere „Kontrollierbarkeit“

Als weiteren Grund für die zurückhaltende Aufnahme des konnektionistischen Ansatzes wird oft die mangelnde Kontrollierbarkeit neuronaler Netze angegeben. Zwar scheint die hohe Praxistauglichkeit neuronaler Netze unbestritten, jedoch schleicht sich ein gewisses Unbehagen ein, in Anbetracht der Tatsache, daß man die in einem solchen System vor sich gehenden Prozesse, nicht mehr vollständig erfassen kann. Die Kontrollierbarkeit komplexer neuronale Netze entzieht sich dem menschlichen Betrachter, da er nicht mehr zu jeder Zeit alle neuronalen Zustände gleichzeitig erfassen kann (vgl. Lyons, 1996). Dieses Problem existiert nicht bei streng seriell arbeitenden Maschinen, da hier immer nur eine Operation zu einer bestimmten Zeit ausgeführt wird.

Aus diesem Grunde wird auch von einigen Autoren zum besonnenen Einsatz neuronaler Netze in sensiblen Bereichen wie der Flugsicherung, bei militärischen Entscheidungen oder beim Ausüben von Regierungsgewalt gemahnt (Hoffmann, 1992).

⁴³ wesentliche neue Impulse kamen von Rumelhart, McClelland und der PDP-Gruppe (1986).

5.4.4 Fehlende „Einheit“ bei neuronalen Netzen

Das Phänomen Bewußtsein wird gemeinhin als Einheit gesehen. Dies begründet sich in der vermeintlichen Tatsache, daß ein Mensch zu *einer bestimmten* Zeit immer nur *einen bestimmten* Gedanken denken kann. Auch wenn spätestens durch psychologische Experimente zu Doppel- und Mehrfachaufgaben (siehe Baddeley, 1990) hinreichend bewiesen wurde, daß ein Mensch tatsächlich mehrere Aufgaben in sinnvoller und koordinierender Weise gleichzeitig vollbringen kann, hält sich hartnäckig die Mär der Einzelaufgaben-Beschränkung des Bewußtseins (z.B. Penrose, 1991).

Da konnektionistische Systeme jedoch parallele eine Daten- und Speicherungsverarbeitung als *modus operandi* verwenden, sieht z.B. Penrose (1991) die Möglichkeit von einer Bewußtseinsfähigkeit bei solchen Systemen gefährdet. Die Argumentation von Penrose ist allerdings nicht ganz stichhaltig, denn selbst wenn das menschliche Bewußtsein nur zu Einzelaufgaben fähig wäre, kann dies noch kein echtes Argument dafür sein, daß deswegen parallel arbeitende Systeme kein Bewußtsein entwickeln können (vgl. Tholey, 1992).

Grush & Churchland (1996) machen in diesem Zusammenhang noch auf einen weiteren wichtigen Punkt aufmerksam. Ihres Erachtens wird der Konnektionismus schon alleine deswegen oft schwer attackiert, da er kontrainduktiv zur Psyche des Menschen ist: Der Mensch trägt insgeheim stets die Vorstellung einer *Einheit* des Bewußtseins als etwas höherem und mystischem in sich. Da sich der Konnektionismus aber deutlich von einer homunculus'schen Bewußtseinsvorstellung distanziert, kann er auch nicht dem üblichen Verständnis von Bewußtsein entsprechen.

5.4.5 Konfundierung mit klassischer KI

Der Ansatz der klassischen KI scheint, durch viele praktische Beispiele (siehe Abschnitt 3.4.1) und durch theoretische Überlegungen (siehe Abschnitt 3.4.3), deutlich in seine Schranken verwiesen worden zu sein. Manche sprechen sogar davon, daß die Forschungsrichtung der KI zwingend ins Lehre laufen müsse (z.B. Dreyfus & Dreyfus, 1996) bzw., daß Künstliche Intelligenz auf Basis von Von-Neumann-Maschinen gar nicht wirklich existiere, da sie überhaupt nicht fähig ist, *Intelligenz* hervorzubringen (siehe Reeke & Edelman, 1996).

Aus diesen Gründen, wird von vielen der Versuch, künstliche Intelligenz zu schaffen, als *insgesamt* gescheitert angesehen. Diese Kollektiv-Verurteilung wird aber der KI, wie sie sich heute darstellt, in keinster Weise gerecht. Zu viele neue Ansätze, die durchaus Potential zur Nachbildung menschlicher Fähigkeiten besitzen, sind seit den frühen Tagen der KI entwickelt worden. Der Konnektionismus ist dabei nur einer unter vielen erfolgversprechenden Ansätzen. Deswegen muß genau unterschieden werden, welcher Art von KI man die Fähigkeit zum Entwickeln von Intelligenz oder Be-

wußtsein abspricht. Zu oft ist jedoch Konnektionismus mit klassischer KI gleichgesetzt worden, woraus eine voreilige Verdammung dieses Forschungsbereichs resultierte.

Hat man aber erst einmal diese Konfundierung aufgelöst, lohnt es sich, frei nach dem Motto „Die KI ist tot — lange lebe die KI!“, die Ansätze der modernen KI näher zu untersuchen.

5.5 Neueste und zukünftige Entwicklungen

5.5.1 Abschätzung des Potentials

Große Potentiale werden konnektionistischen Systemen durch die Entwicklung effizienterer Lernalgorithmen zugebilligt. Alle bisher gängigen Modelle gebärden sich äußerst rechen- und zeitintensiv und sind damit meist nur in sehr simpler Art und Weise zu implementieren.

Weitere Verbesserungsmöglichkeiten bestehen in der Auflösung der symmetrischen Aufbaustruktur heutiger neuronaler Netze (siehe Warwick, 1998) und der Anpassung der verwendeten Transferfunktionen, jeweils in Anlehnung an biologische Netze. Diese Weiterentwicklungen führen zu einer Veränderung der Kommunikationsstruktur zwischen einzelnen Neuronen. Bisher gibt es noch einen wesentlichen funktionalen Unterschied zwischen natürlichen und künstlichen neuronalen Netzen, da natürliche Neuronen schubweise über Impulspakete (sog. *Spiketrains*) kommunizieren, künstliche Neuronen jedoch über digitale Aktivitätszustände (vgl. Hamilton, 1997).

Neuronale Netze der *dritten Generation* nehmen sich genau dieser Thematik an, indem sie mit neu entwickelter analoger Hardware analoge Kommunikationsformen zulassen. Ziel der neuen Hardware ist es, „analoge Werte sehr schnell in fehlertoleranter Weise über größere Entfernungen zu übermitteln“ (Hamilton, 1997, S.138).

Auch Jacob T. Schwartz hebt die Vorteile analoger Netze, die er in ihrer „großartigen Simplizität und Kompaktheit im Unterschied zu vergleichbaren, funktionalen, digitalen Systemen“ (Schwartz, 1996, S.136) sieht, hervor.

Allerdings befindet sich der Forschungsbereich der analog operierenden neuronalen Netzen erst im Anfangsstadium.

5.5.2 Praktische Perspektiven

Wesentliche Impulse erwarten sich moderne KI-Forscher von sogenannten rekonfigurierbaren Systemen, deren Schaltkreisstrukturen vollautomatisch auf die an sie gestellten Aufgaben adaptieren (Fremerey, 1999). Diese Adaptions-Eigenschaft befähigt diese Computer, sich schnell und effektiv auf neue Probleme einzustellen. Intelligenz in Form von Flexibilität wäre damit möglich.

Eine weitere Forschungsrichtung moderner neuronaler Forschung besteht in der Konstruktion von Biologischen Computern, die als Hardware⁴⁴ biologische Systeme verwenden. So wird z.B. in Schulemann (1999) ein biologischer Computer vorgestellt, der aus Blutegel-Neuronen besteht. Noch beschränkt sich die Leistungsfähigkeit solcher Systeme auf das Lösen einfacher Rechenaufgaben und ist somit recht bescheiden, doch scheinen die Zukunftsaussichten durchaus vielversprechend:

„In Zukunft will man eine neue Generation flexibler und schnellerer Computer auf Basis der Forschungsergebnisse entwickeln, deren Funktionsweise der des menschlichen Gehirns ähneln soll. Im Gegensatz zu gängigen Rechnern soll der Bio-Computer Antworten auch aus Teilinformationen entwickeln können, indem er Lücken selbständig füllt“ (Schulemann, 1999, S.28).

Einen großen Stellenwert innerhalb der neueren Forschung nimmt der Bereich der *AL* (*Artificial Life*) ein, deren Ziel es ist, nicht nur kognitive Leistungen in punktueller Weise hervorzubringen, sondern ein Gesamtsystem zu konstruieren, das in einer realen Umgebung in sinnvoller Weise flexibel auf natürliche Probleme reagieren kann. Sperlich (1993) beschreibt die Rolle der Neuroinformatik innerhalb der *Artificial-Life*-Forschung folgendermaßen:

„Es besteht jedoch kein Zweifel daran, daß das Prinzip der verteilten und dynamisch interaktiven Systeme, wie es die Neuroinformatik präsentiert, den ‚Real-World‘-Problemen wesentlich näher kommt, als alle Ansätze, die die Wissenschaft bisher verfolgt hat. Sie sind die logische Antwort auf die Erkenntnis, daß das menschliche Denken, ebenso wie die Welt selbst, kein simples, deterministisches Räderwerk ist, sondern ein komplexes, vernetztes System, das man nicht beliebig vereinfachen kann, ohne seine charakteristischen Eigenschaften zu verlieren. Und gerade diese Eigenschaften sind es, die neuronale Netze, künstliche wie natürliche, so faszinierend machen“ (Sperlich & Möcke, 1993).

Weitere eindrucksvolle Beispiele von konnektionistischen Forschungsvorhaben gibt Churchland (1997) resp. Werntges & Eckmiller (1988).

⁴⁴ In diesem Zusammenhang ist statt *Hardware* wohl besser der Begriff *Wetware* zu verwenden (siehe Keil, 1993, S.196).

5.5.3 Auflösung einer alten Feindschaft

Goschke & Koppelberg (1993) kritisieren, daß konnektionistische und symbolverarbeitende Ansätze meist als sich wechselseitig ausschließende Forschungsvorhaben angesehen werden. Sie plädieren für eine integrierte Perspektive.

Der Konnektionismus sollte nicht als Substitut für klassische Ansätze der KI benutzt werden, sondern als ergänzende Komponente, um menschliches Verhalten besser erklären zu können. So könnte auch das fundamentale Dilemma aufgelöst werden, daß sich üblicherweise jedem Lebewesen in irgendeiner Form stellt: Biologische Systeme müssen einerseits Kovariationen und Invarianzen ihrer Umwelt entdecken, andererseits müssen sie auf die Muster der gefundenen Umwelteinflüsse adaptiv und möglichst direkt reagieren.

Goschke & Koppelberg (1993) sehen daher symbolorientierte Verarbeitungsprozesse nicht als eigenständige Operationsmodi, sondern als Manifestationen hochkomplexer neuronaler Verschaltungen.

Konsequenterweise wird deshalb von Goschke & Koppelberg (1993) angenommen, daß Symbolsysteme zwar durchaus von zentraler Bedeutung für das Verständnis menschlicher Kognition sind, daß dies jedoch nicht unbedingt impliziert, daß man auch von der Existenz eines inneren Repräsentationssystems mit einer kombinatorischen Syntax und Semantik ausgehen muß.

6 Bewußtsein

Der Traum, ein künstliches Wesen mit Wissen, Intelligenz und letztendlich auch Bewußtsein zu schaffen, ist alt. Schon

„Hobbes und Leibniz versuchten, auf sehr verschiedene Weise, die Implikationen der Idee zu erforschen, den Geist in kleine, letztlich mechanische Operationen zu zerlegen“ (Dennett, 1996c, S.269).

Wie kann nun aber ein künstliches „Wesen“ ein Bewußtsein oder ein Selbst entwickeln? Muß ihm dieses Bewußtsein praktisch quasi als „Zugabe“ mit einprogrammiert werden, ist es unter bestimmten Umständen *a priori* vorhanden, oder reicht es, eine bestimmte Komplexität des Systems zu überschreiten, um ein Bewußtsein von selbst generieren zu lassen? Historische und theoretische Überlegungen sollen im folgenden Kapitel etwas Licht in dieses Dunkel bringen.

Außerdem soll der Leser für das besondere Problem der Subjektivität sensibilisiert werden, das durch Searle (1997) ironisch charakterisiert wird: “If it consciously seems to me that I am conscious, then I am conscious.” (Searle, 1997, S.213).

6.1 Das Phänomen „Bewußtsein“

6.1.1 *Einzig, einzigartig oder einsam und alleine?*

Ist das Bewußtsein *einzig*, d.h. besitzt nur der Mensch Bewußtsein? Ist dieses Bewußtsein *einzigartig* in dem Sinne, daß kein anderes Phänomen so außergewöhnlich ist? Ist das Bewußtsein insofern *einsam*, als es eine Autarkie des Geistes gibt?

In jedem Fall stellt sich uns ein extraordinär komplexes Themenfeld dar, das Popper (1997) folgendermaßen beschreibt: „Die Emergenz eines vollen, der Selbstreflexion fähigen Bewußtseins, [...] ist eigentlich eines der größten Wunder“ (Popper & Eccles, 1997, S.167). Und selbst Karl H. Pribram meint in Bezug auf das Phänomen *Bewußt-*

sein: “ghosts really exist, and [...] they too are important” (nach Dewan et al., 1976, S.317).

Demgegenüber bestreitet McGinn (1996) die von vielen anderen Wissenschaftlern vermeintlich entdeckte hohe Komplexität des Phänomens *Bewußtsein*. Seines Erachtens ergibt sich die spezifische Problematik nicht so sehr aus der *Komplexität* der Probleme, als vielmehr aus dem *Typ* der Probleme:

“This reflection should make us receptive to the idea that it is something about the tracks of our thought that prevents us from achieving a science that relates consciousness to its physical basis: the enemy lies within the gates” (McGinn, 1996, S.287).

McGinn (1996) kommt zum Schluß, daß das, was Bewußtsein ausmacht, zwar nie vollständig spezifiziert werden kann, daß dies aber in jedem Fall nichts Geheimnisvolles (*miraculous*) ist. Zwar weiß er, daß er seine These nicht begründen kann, jedoch gibt er zu bedenken, daß der Mensch aufgrund seiner kognitiven Geschlossenheit (*cognitive closure*) niemals die Grenzen der Realität erfahren kann. Deshalb mag der Mensch vielleicht viele Phänomene nicht hinreichend erklären können, geheimnisvoll sind sie deshalb aber noch lange nicht. Mit dieser Meinung begibt sich McGinn (1996) auf ein gefährliches Terrain: So fordert er eine ganz neue Herangehensweise an die Probleme des Geistigen, die er folgendermaßen umreißt:

“We need to cultivate a vision of reality (a metaphysics) that makes it truly independent of our given cognitive powers, a conception that includes these powers as a proper part” (McGinn, 1996, S.289, Klammern i. Original).

Tatsächlich erinnern diese Zielvorgaben an vorkantische Wissenschaftsideen von der systematischen Erforschung der *Dinge-an-sich*, die spätestens seit Erscheinen von Kants *Kritik der reinen Vernunft* obsolet erscheinen.

Untersucht man die Frage, ob ein Mensch nur ein einziges Bewußtsein oder mehrere Bewußtseinsströme gleichzeitig ausbilden kann, fällt sofort auf, daß der Plural zum Begriff Bewußtsein in einer Vielzahl von Sprachen gar nicht vorgesehen ist. Die Verwendung einer Mehrzahlform würde vermutlich allgemein auf wenig Verständnis stoßen, da gemeinhin angenommen wird, daß man stets nur *ein* Bewußtsein zu einer *bestimmten* Zeit besitzen kann.

Allerdings kann die Neurobiologie mit sogenannten *split-brain*-Experimenten eindrucksvoll demonstrieren, daß Bewußtseinsvorgänge nicht zwingend singulär sind:

Spaltet man nämlich das *Corpus callosum* (Gehirnbalken), also die Nervenfasern, die die linke und die rechte Hirnhälfte miteinander verbinden⁴⁵, so erreicht man eine Isolierung der beiden Gehirnhemisphären. Die Folge ist, daß Realitäten von den beiden Hirnhälften jeweils unterschiedlich interpretiert werden⁴⁶: Da jede der beiden Hemisphären auf bestimmte Aufgaben spezialisiert ist⁴⁷, können nur die jeweils für eine bestimmte Hirnhälfte spezifischen Fähigkeiten eingesetzt werden, d.h. *split-brain*-Patienten nehmen ein und dieselbe Sache, je nachdem welcher Gehirnhälfte sie präsentiert wird, vollkommen unterschiedlich wahr⁴⁸. Nach Ansicht des Gehirnforschers Sperry könnte dies ein Indiz dafür sein, daß bei solchen Patienten *Bewußtsein* tatsächlich in Form zweier separater Einheiten existiert (siehe auch Wimsatt, 1976):

„Alles [...] deutet darauf hin, daß der Geist dieser Personen durch den chirurgischen Eingriff zweigeteilt worden ist, daß sie also zwei getrennte Bewußtseinssphären besitzen. Das, was in der rechten Hirnhälfte erfahren wird, scheint vollständig außerhalb des Erfahrungsbereichs der linken Gehirnhälfte zu liegen. Diese geistige Dimension ist für Wahrnehmung, Denken, Wollen, Lernen und Gedächtnis nachgewiesen worden“ (zit. nach Thompson, 1990, S.301).

Demgegenüber vertritt der Neurophysiologe John Eccles die Meinung einer strikten Einheit des Bewußtseins. Da für ihn das Bewußtsein von gesunden Menschen immer mit Sprache und verbalen Fähigkeiten assoziiert ist, lokalisiert Eccles den Sitz des Bewußtseins in der linken Hirnhälfte (siehe Kapitel 6.3.2.4). Beispiele von Patienten, deren rechte Hirnhälfte zusätzlich zur dominanten linken Hemisphäre beachtliche verbale Leistungen entwickelte, sprechen jedoch gegen den Ansatz von Eccles (siehe Thompson, 1990).

Erkenntnisse aus der Erforschung der Tierwelt legen zudem nahe, daß die Idee der Singularität des Bewußtseins vermutlich zu kurz gegriffen ist. So kann man mittels

⁴⁵ Dieser schwerwiegende chirurgische Eingriff wird freilich nur bei pathologischen Zuständen des Gehirns (z.B. bei schweren Formen von Epilepsie) durchgeführt.

⁴⁶ Nur eine bestimmte Verschaltungseigenart des Gehirns läßt diese Effekte überhaupt in der beschriebenen Weise so leicht erkennen: Da nämlich die *linke* Augennetzhaut nur mit dem visuellem Areal der *rechten* Hirnhälfte verbunden ist (und die *rechte* Netzhaut mit der *linken* Hirnhälfte), kann man sicherstellen, daß durch eine Präsentation auf nur einem Auge, auch nur die jeweils gegenüberliegende Hirnhälfte Signale vom visuellen Apparat erhält.

⁴⁷ Normalerweise besitzt die linke Hemisphäre vorrangig sprachliche Fähigkeiten, der rechten Hirnhälfte werden dagegen eher Gedächtnis- und Formerkennungsfähigkeiten zugesprochen (nach Silberman & Despopoulos, 1988).

⁴⁸ Diese Dissonanz geht teilweise so weit, daß mit der einen Hirnhälfte Gegenstände ohne Verzögerung benannt werden können, mit der anderen Hirnhälfte dagegen nur sinnlose oder vollkommen falsche Bezeichnungen generiert werden (siehe Birbaumer & Schmidt, 1991).

EEG–Daten von Delphinen⁴⁹ erkennen, daß sich jeweils immer nur höchstens eine der beiden Gehirnhälften in einem „Schlafzustand“ befindet. Es ist daher durchaus denkbar, daß die beiden Hemisphären eines Delphins relativ unabhängig voneinander sind. Dies könnte aber weiter bedeuten, daß in der Phase der Wachheit beider Gehirnhälften auch ein zweigeteilter Bewußtseinszustand möglich ist.

6.1.2 “Handicap” Subjektivität

Der Mensch ist sich der Existenz seines eigenen Bewußtseins absolut sicher; darüber hinaus ist er der Meinung, daß andere Menschen ebenfalls über ein Bewußtsein verfügen (vgl. Hoffmann, 1992; Tetens, 1993). Von seiner eigenen Bewußtseinsfähigkeit auf die anderer zu schließen ist zwar eine gute begründete Annahme, bleibt jedoch letztendlich unüberprüfbar, da es sich hier um ein *intrinsisches* Phänomen handelt.

Dieses Problem spiegelt sich auch bei der Abbildung von Bewußtsein in künstlichen Systemen wider. Da man immer nur selbst weiß, welche Qualitäten die eigenen Empfindungen haben, bleibt das Wissen über die mentale Erlebniswelt stets unvollständig (vgl. Zoglauer, 1998): Subjektives Empfinden läßt sich objektiv nicht beschreiben!

Dieses Faktum disqualifiziert auch die Methode der Introspektion, insofern sie sich als adäquate Beschreibungsebene von Erlebnisinhalten begreift, da sich die von introspektiven Verfahren okkupierte Metapher der visuellen Wahrnehmung nicht auf Denkkakte anwenden läßt:

„Im Falle der visuellen Wahrnehmung haben wir eine klare Unterscheidung zwischen dem gesehenen Gegenstand und dem visuellen Erlebnis, das der Wahrnehmende hat, wenn er den Gegenstand wahrnimmt. Diese Unterscheidung können wir jedoch beim Akt der Introspektion eigener bewußter Geisteszustände nicht machen“ (Searle, 1996, S.167).

Dies bedeutet, daß man, wenn man die Aufmerksamkeit nach innen richtet, eben nicht einen Schritt zurücktreten kann, um besser oder objektiv zu sehen (Introspektionsproblematik siehe auch Kapitel 2.1.3.1).

⁴⁹ Das Großhirn von Delphinen ist ebenso groß wie das von Menschen und besitzt eine äußerst komplexe Verschaltungsstruktur (vgl. Penrose, 1991). Ein Vergleich zwischen Menschen und Delphinen in Bezug auf Bewußtseins- und Kognitionsfähigkeiten ist daher nicht uninteressant.

6.1.3 Komplexität des Phänomens

Mittlerweile ist es in den allermeisten Wissenschaftskreisen unstrittig, daß das Auftreten von phänomenalen Zuständen durch bestimmte Vorgänge im Gehirn bedingt wird. Dennoch ist bis heute nicht klar, wie derartige Zustände hervorgebracht werden (siehe Flohr, 1993). Das Gehirn und sein Aufbau ist zwar relativ weit erforscht, aber die genauen Faktoren und Umstände, die Geistiges aus ihm entwachsen lassen, bleiben immer noch ungeklärt:

“The brain has physical properties we can grasp, and variations in these correlate with changes in consciousness, but we cannot draw the veil that conceals the manner of their connection” (McGinn, 1996, S.288).

In Anbetracht der Komplexität⁵⁰ des Problems, versuchen einzelne Forscher nicht *das Bewußtsein* zu erforschen, sondern lediglich Teilbereiche wie etwa das *visuelle Bewußtsein* (so z.B. Crick, 1997). Allerdings wird von anderer Seite davor gewarnt, das Phänomen des Bewußtsein zu einseitig bzw. zu eingeschränkt zu betrachten: So mahnt Calvin (1989b), Bewußtsein nicht nur als Aufmerksamkeit (*awareness/attention*) anzusehen, und Penrose (1991) zeigt auf, daß Bewußtsein mehr leistet, als nur mathematische Urteile zu bilden.

6.1.4 Exkurs: Möglicher Nutzen durch Bewußtsein

Eine nicht uninteressante Frage, die jedoch die Grundfrage der Arbeit nur am Rande streift, und deswegen nur kurz behandelt werden soll, betrifft den möglichen Nutzen von Bewußtsein: Wozu ist Bewußtsein eigentlich gut?

Betrachtet man Lebewesen, denen man gemeinhin Bewußtsein zuspricht, so zeigt sich ein wesentlicher Selektionsvorteil durch deren Möglichkeit, Entscheidungen bewußt zu treffen (vgl. Penrose, 1991; Popper & Eccles, 1997). Zwar können die meisten Lebenssituationen durch Routine gelöst werden; sobald aber unerwartete Ereignisse auftreten, ist Bewußtsein *sensu* Aufmerksamkeit und Bewußtsein in Form von kreativen Entscheidungsfindungen von großem Überlebenswert⁵¹. Weitere Vorteile ergeben sich

⁵⁰ Nach Searle (1997) sollte man freilich, ehe man den Begriff Komplexität verwendet, genau definieren, wie man diese überhaupt messen kann.

⁵¹ Schrödinger ging in diesem Punkt noch einen Schritt weiter, da er meinte, daß jedes für ein Organismus neue Problem, einen bewußten Lösungsversuch verlangt. Peter Medawar konnte jedoch zeigen, daß z.B. das Immunsystem fortgesetzt neuen Problemen gegenübersteht, die es jedoch vollständig „unbewußt“ löst (nach Popper (1997, S.163)).

durch die größere Unterscheidungsfähigkeit, durch Aufmerksamkeitsfokussierung, erhöhte Empfindungsfähigkeit und höhere Flexibilität (vgl. Searle, 1996).

6.2 Anhaltspunkte für Bewußtsein

Der folgende Abschnitt will Anhaltspunkte dafür aufspüren, wann und wie Bewußtseinszustände entstehen können. Theoretische Begründungen für diese Indizien spielen dabei noch keine Rolle, da diese erst in Kapitel 6.3 besprochen werden sollen.

6.2.1 Konzeptuelle Anhaltspunkte

6.2.1.1 Gedächtnis

Karl Popper ist der festen Ansicht, daß das Vorhandensein eines Gedächtnisses ein notwendiger Faktor zum Ausbilden eines Bewußtseins darstellt:

„Ich glaube nämlich, [...] daß das Bewußtsein aufhört zu existieren, wenn das Gedächtnis so unterbrochen wird, daß es oft und nachhaltig genug in kurze Abschnitte zerlegt wird“ (Popper & Eccles, 1997, S.526).

Zwar akzeptiert Popper die Meinung, daß durch kurzfristige Unterbrechung des Gedächtnisses möglicherweise Zwischenzustände von Bewußtheit eintreten, jedoch führe dies bei einer weitergehenden Atomisierung des Gedächtnisses zwangsläufig zum Verlust des Bewußtseins (vgl. Dennett, 1996b; Rosenfield, 1996). Gemäß dieser Gedächtnis–Bewußtseins–Beziehung wird von Popper auch Tieren eine Art Bewußtheit zugesprochen, die auf Erinnerung beruht.

Auch Churchland (1997) nimmt an, daß es Bewußtsein ohne gewisse Gedächtnisleistungen nicht geben kann (siehe auch Calvin, 1998):

„Typischerweise ermöglicht uns Bewußtsein wahrzunehmen, wie sich der eigene psychische und physische Zustand im Laufe der Zeit entwickelt. Ein derartiger Sinn erfordert wenigstens eine gewisse kognitive Erfassung der unmittelbar vorangegangenen Geschehnisse und damit eine Form von Gedächtnis, wenigstens ein Kurzzeitgedächtnis“ (Churchland, 1997, S.251).

Die enge Verwandtschaft der Konzepte *Bewußtsein* und *Gedächtnis* läßt sich auch am üblichen Sprachgebrauch ablesen: So wird oft der Ausdruck „er kann sich nicht daran erinnern“ mit „er ist sich nicht darüber bewußt“ gleichgesetzt.

6.2.1.2 Analyseebene

Ein wesentliches Kennzeichen wissenschaftlichen Arbeitens ist das Erkennen und Verwenden von verschiedenen Erklärungsebenen, je nach dem welches Phänomen man beschreibt. So ist es zwar eine ausgezeichnete wissenschaftliche Strategie, die Dinge auf ihre Grundlagen zurückzuführen, jedoch muß dafür Sorge getragen werden, daß sich diese Grundlagen auf dem richtigen Organisationsniveau befinden. Daher scheint es wenig sinnvoll zu sein, Bewußtsein vollständig erfassen zu wollen, indem man einzelne Aktivationspotentiale von Gehirnneuronen analysiert. Der Kognitionswissenschaftler Douglas Hofstadter macht dies am Beispiel eines Verkehrsstaus deutlich: Auch hier erkennt man schnell, daß die Ursache für das dynamische Verhalten eines Verkehrsstaus nicht in einem einzigen Verkehrsteilnehmer zu finden ist. Gelegentlich kann zwar ein defektes Auto der Auslöser einer Verkehrsverdichtung sein, jedoch ist ein eingebrauntes Auslaßventil kaum die richtige Analyseebene; jedenfalls nicht im Vergleich zum allgemeinen Verkehrsaufkommen, den Witterungsverhältnissen, den Reaktionszeiten der einzelnen Fahrer oder der spezifischen Bremskraft der Fahrzeuge (vgl. auch Calvin, 1998; siehe Hofstadter, 1979).

6.2.2 *Biologische Anhaltspunkte*

Gemeinhin werden allen höheren Tieren, insbesondere dem Menschen, mehr oder minder hohe Bewußtseinsfähigkeiten zugesprochen, welche anorganischen Entitäten abgesprochen werden. Könnte die Ausbildung eines Bewußtseins also von der Materie des Gehirns zwingend anhängig sein? Oder ist der „Stoff“, aus dem das Gehirn besteht, einfach nur eines von vielen möglichen Substraten, die solche Hirnfähigkeiten hervorbringen können?

Die Meinungen in diesem Punkt gehen weit auseinander: Von der Ansicht, daß „eine biotische Hirnmasse notwendig für Bewußtsein ist“, bis hin zur Überzeugung, daß „tote Materie mehr hervorbringen kann, als tote Materie“ (Überblick hierzu siehe Birnbacher, 1996). Differenziertere Überlegungen zu diesem Thema finden sich in Kapitel 6.3.3.1.

6.2.3 *Neurobiologische Anhaltspunkte*

Neben rein biologischen Anhaltspunkten existieren auch neurobiologische Auffälligkeiten, die während bewußter Zustände zu registrieren sind. So beschreibt z.B. John Eccles, daß Schmerzen nur dann empfunden werden, wenn Impulse von Sinnesorganen zu den höheren Ebenen der Großhirnrinde oder wenigstens bis zum Thalamus

aufsteigen⁵² (siehe Popper & Eccles, 1997). Im folgenden Abschnitt sollen solche Effekte kurz skizziert werden.

6.2.3.1 Cell assemblies

Das Konzept der *cell assemblies* wurde von Hebb (1949) entwickelt und später durch Christoph von der Malsburg modifiziert (nach Flohr, 1993). Hebb versteht darunter einen Verband von Neuronen, der durch wiederholte Stimulation spezifischer Rezeptoren zu einem geschlossenes System werden kann, und dadurch nicht mehr von kontinuierlicher Reizung abhängig ist. Die Grundlage für eine derartige aktivitätsabhängige Selbstorganisation von Neuronennetzen ist das Vorhandensein bestimmter plastischer Synapsen (siehe Kapitel 4.2.1).

Diese *cell assemblies* stellen den einfachsten Fall einer repräsentationalen Struktur dar:

„Das assembly entsteht, weil das Reizmuster koinzidierte Merkmale enthält; es ‚entdeckt‘ und repräsentiert diesen Zusammenhang“ (Flohr, 1993, S.38).

Es scheint daher einen gesetzmäßigen Zusammenhang zwischen der spezifischen Ausprägung von *cell assemblies* und phänomenalen Zuständen zu geben. Als mögliche Unterscheidungsmöglichkeit zwischen unbewußten und bewußten Zuständen sieht Flohr (1993) die Geschwindigkeit an, mit der *cell assemblies* generiert werden: *cell assemblies* unbewußter Prozesse werden nach dieser Theorie signifikant langsamer generiert als die von bewußten Prozessen.

6.2.3.2 Neuronale Muster

Der Gehirnforscher Crick (1997) sieht das Phänomen der Neuronenbindung als wesentlich für die Ausbildung von einigen Arten von Bewußtsein an. Solche Bindungen können seines Erachtens durch eine Oszillation der Aktivitätsmuster entstehen, die z.B. im olfaktorischen und visuellen System vorkommen. Diese synchronisierten Aktivitäten liegen üblicherweise im Bereich von 35 bis 75 Hertz und werden als *Gamma-Oszillationen*⁵³ bezeichnet (nach Crick, 1997).

Dewan (1976) erläutert einige Beispiele solcher gegenseitigen Anpassung von Oszillationen aus dem Bereich der Biologie. Er kommt zu dem Schluß, daß das Phänomen

⁵² Da erst Vögel und Säugetiere solche höheren Gehirnstrukturen besitzen, stellt John Eccles explizit die Frage, ob niedere Tierarten alleine wegen dieses Defizits ein Bewußtsein entwickeln können (nach Popper & Eccles, 1997).

⁵³ Stellenweise auch vereinfachend und damit fälschlich als *40Hz-Oszillationen* bezeichnet. Saum-Aldehoff (1993) bezeichnet diese spezifische Oszillation auch als *Singer'sche 40-Hertz-Welle*.

der Oszillation sich nach außen hin so zeigt, als hätten solche Systeme eine höhere Kontrollinstanz (*virtual governor*), die das ganze System reguliert. Der entscheidende Unterschied zu einer physischen Instanz ist aber, daß sich dieser *virtual governor* über das ganze System hinweg ausbreitet und nicht an einem bestimmten Platz lokalisierbar ist.

Das Ziel der hier vorgestellten neurobiologischen Forschung ist stets, neuronale Korrelate zu Bewußteinszuständen zu finden. Dieses Vorgehen bietet eine klare und wissenschaftliche Arbeitsgrundlage, jedoch muß kritisch angemerkt werden, daß dadurch wie bei allen Korrelationsstudien ein spezifisches Informationsdefizit entsteht, da unbekannt bleibt, in welcher kausalen Richtung einzelne Effekte wirken (siehe Bortz, 1989).

6.3 Theoretische Ansätze

Teilweise aufbauend auf den aus Kapitel 6.2 gewonnenen Anhaltspunkten über die Entstehung und Erfassung von bewußten Zuständen, werden im folgenden weiterführende elaborierte theoretische Ansätze vorgestellt.

6.3.1 Psychismus–Theorien

6.3.1.1 Panpsychismus

Der Panpsychismus ist eine Allbeseeltheitslehre, nach welcher alle Wesenheiten beseelt sind, ein Bewußtsein besitzen und Leben in sich tragen (nach Schischkoff, 1991). Genau betrachtet stellt sich dem Panpsychismus das Problem des Bewußtseins erst gar nicht, da es innerhalb dieses Systems ja *per definitionem* gar keine Entität ohne Bewußtsein gibt (siehe Popper & Eccles, 1997, S.52f.). Allerdings sei die Idee des Panpsychismus nur am Rande erwähnt, da sie meist nur ein Zeichen für eine relativ unreflektierte Weltanschauung darstellt.

6.3.1.2 Vitalismus

Auch der vitalistische Standpunkt, nach dem sich lebende von toter Materie lediglich durch einen von Gott eingehauchten „Lebensfunken“ unterscheidet, kann wissenschaftlich kaum überzeugen, obwohl er eine Renaissance im 20. Jahrhundert unter dem Begriff „Neuvitalismus“ erlebte. Der Neuvitalismus „verneint eine kausal-mechanische Erklärbarkeit der Lebenserscheinungen und behauptet ihre Planmäßigkeit, Zielstrebigkeit und Eigengesetzlichkeit“ (Schischkoff, 1991, S.760). Für kybernetisch denkende Biologen gilt der Vitalismus allerdings als endgültig widerlegt (siehe Dennett, 1996a).

6.3.2 Dualistische Theorien

Die philosophischen Begriffe von Leib und Seele sind ungefähr so alt wie die abendländische Philosophie selbst, jedoch tritt deren explizite Verbindung zum Leib–Seele–Problem erst mit Descartes in aller Deutlichkeit auf das philosophische Parkett (siehe Wolters, 1993)⁵⁴.

Klassische dualistische Geist–Körper–Theorien postulieren das Nebeneinander von geistigen und körperlichen Phänomenen, die als zwei verschiedene, nicht zu einer Einheit führenden Prinzipien gesehen werden. Üblicherweise wird zwischen drei Arten von dualistischen Theorien unterschieden (vgl. Zoglauer, 1998), die in der Folge besprochen werden sollen. Zusätzlich ist an die folgende Übersicht über die verschiedenartigen Theorien ein Exkurs über die Frage nach der Lokalisation von Bewußtsein angehängt.

6.3.2.1 Interaktionismus

Descartes Gedankenwelt und der damit verbundene Interaktionismus ist wohl kaum ohne den damals bestehenden Zeitgeist der mechanischen und physikalischen Entdeckungen und Errungenschaften zu verstehen. Erstmals wird konsequent ein einheitliches Verstehensmodell entwickelt, das auf keinerlei substantielle Kräfte rekurriert. Um diesem Modell zu entsprechen, werden Menschen und Maschinen gleichgesetzt: Beide gehorchen mechanischen Gesetzmäßigkeiten, wobei sich Menschen dennoch insofern von Maschinen unterscheiden als sie von Gott konstruiert wurden. Anfangs wurde dieses exklusiv auf mechanischen Prinzipien beruhende Modell auch auf den Bereich geistiger Fähigkeiten angewendet. Es scheint jedoch u.a. durch kirchlichen Druck abgeändert worden zu sein (vgl. Wolters, 1993), indem eine eigene geistige Substanz (*res cogitans*) postuliert wurde, welche die Erkenntnisleistung und das Wollen des Menschen erklären soll (vgl. Popper & Eccles, 1997). Die *res cogitans* und die körperliche Substanz (*res extensa*) sind real unterscheidbar und können wechselseitig aufeinander einwirken:

„Alles was wir klar wahrnehmen, kann von Gott so gemacht werden, wie wir es wahrnehmen [...]. Nun nehmen wir aber klar den Geist, d.h. eine denkende Substanz ohne den Körper, d.h. ohne eine ausgedehnte Substanz wahr [...]; und umgekehrt auch den Körper ohne den Geist [...]. Also kann [...] der Geist ohne den Körper sein und der Körper ohne den Geist.“

⁵⁴ Zwar beschäftigte auch Platon und Aristoteles diese Problematik, jedoch verfolgten sie dieses Problem nicht mit der von Descartes durchgeführten Stringenz (siehe Wolters, 1993, S.19).

Nun sind aber Substanzen, von denen jede ohne die andere sein kann, real verschieden⁵⁵.

Realisiert werden die Interaktionen zwischen *res cogitans* und *res extensa* durch Bewegungen. Die Annahme, daß die *res cogitans* Bewegung erzeugen kann, ist jedoch wegen des kartesischen Universalgesetzes von der Erhaltung der Gesamtenergie höchst problematisch. Daher nimmt Descartes weiter an, daß die Seele nicht neue Bewegungen initiiert, sondern lediglich die *Richtung* bereits vorhandener Bewegungen determiniert (siehe Specht, 1966). Wie schon Leibniz zeigen konnte, wird jedoch auch durch eine reine Richtungs determinierung das Energieerhaltungsgesetz verletzt (siehe Zoglauer, 1998, S.53f.).

Alleine aus diesen energetischen Gründen⁵⁶, verfolgte man in nachkartesischer Zeit die Gedanken Descartes nicht mehr in der von ihm verwendeten Reinform. Dennoch war es vor allem die kartesische Tradition die eine nachhaltige Wirkung auf die modernen Wissenschaften ausübte:

“It was Cartesianism rather than pure Descartes that survived into the late nineteenth century in both philosophy of mind and its young offspring, psychology” (Lyons, 1996, Einleitung xlv)

6.3.2.2 Parallelismus

Aus den spezifischen Problemen, die sich durch eine interaktionistische Sichtweise von Körper und Geist ergeben (siehe Leibniz, 1879, S.139ff), entwickelte Gottfried Wilhelm Leibniz die Theorie des *Parallelismus*⁵⁷. Zwar werden auch hier Geist und Körper als verschieden angenommen, sie üben hier jedoch keine Wechselwirkung aufeinander aus; vielmehr verlaufen psychische und physische Ereignisketten parallel. Als Bild hierfür wird das zweier vollkommen synchron laufender Uhren verwendet, welche stets die gleiche Uhrzeit anzeigen, obwohl sie nicht miteinander gekoppelt sind. Damit dennoch eine sinnvolle Übereinstimmung von beiden Ereignissträngen gewährleistet ist, sind alle mentalen und physikalischen Ereignisse für alle Zeiten determiniert und vorherbestimmt (siehe Popper & Eccles, 1997; Wolters, 1993; Zoglauer, 1998).

⁵⁵ aus: Descartes: *Meditationen über die Grundlagen der Philosophie*, Satz IV (zit. nach Meier-Oeser, 1997, S.276).

⁵⁶ wie Zoglauer (1998) zeigt, ist Leibniz Argument allerdings nicht unüberwindlich, da Descartes Theorie durchaus mit Hilfe von Energiefluktuationen oder spontanen quantenmechanischen Fluktuationen gestützt werden könnte.

⁵⁷ oft auch als psychophysischer Parallelismus bezeichnet (z.B. Zoglauer, 1998); in der spezifischen Ausprägung bei Leibniz auch *Prästabilierte Harmonie* genannt.

In den frühen Anfängen der empirischen Psychologie wurde das parallelistische Prinzip durch Gustav Theodor Fechner und Wilhelm Wundt in der Ausprägung der *Psychophysik* wieder aufgegriffen (siehe Dorsch, 1987; Popper & Eccles, 1997).

6.3.2.3 Epiphänomenalismus

Die dritte hier besprochene dualistische Theorie sieht den Geist lediglich als ein Epiphänomen an, d.h. als eine unwesentliche Begleiterscheinung, der keine eigene Existenz zukommt. Der Epiphänomenalismus erachtet eine Wirkung des Körpers auf den Geist für *möglich*⁵⁸, jedoch wird die Kausalrichtung von Geistigem zu Körperlichem ausgeschlossen. Aufgrund dieser unidirektionalen Kausalbeziehung nimmt der Epiphänomenalismus eine Zwischenstellung zwischen Interaktionismus und Parallelismus ein.

Der Geist wird damit als kausal impotent gesehen, der lediglich passiver Beobachter von körperlichen Zuständen und Prozessen ist. Obwohl wir unser Handeln intuitiv als von unserem Geist gesteuert sehen, ist der Wille lediglich Epiphänomen neuronaler Aktivität (siehe Zoglauer, 1998, S.76ff.). Diese Ansicht führt jedoch den Begriff des *Handelns* geradezu *ad absurdum*, da dieser wesentlich eine Bewußtseinskomponente enthält:

„Demnach wäre ein Verhalten, das dem Sich–Verhaltenden nicht intendiert (ungewollt) erscheint und zu dessen Erklärung man sich nur auf nicht–bewußtseinsfähige Regulationsebenen im Organismus beziehen kann, kein Handeln“ (Heckhausen, 1989, S.13).

Die Theorie des Epiphänomenalismus wurde maßgeblich durch den Biologen Thomas Henry Huxley entwickelt und stützt sich auf zahlreiche an Tieren durchgeführte Untersuchungen. In vielen dieser Experimente konnte gezeigt werden, daß zwar Körperbewegungen ohne mentale Ursachen zustande kommen können, daß es jedoch kein Bewußtsein ohne materielle Ursachen gibt. Auch moderne neurophysiologische Experimente von Kornhuber und Libet, welche in Kapitel 6.3.6 eigens gewürdigt werden, scheinen Indizien für eine epiphänomenalistische Theorie des Bewußtseins bereitzustellen (vgl. Zoglauer, 1998).

6.3.2.4 Exkurs: Lokalisations–Problematik

Seit jeher faszinierte den Menschen die Frage, ob es einen örtlich bestimmbaren Punkt gibt, an dem die Seele resp. das Bewußtsein des Menschen zu lokalisieren ist. Schon

⁵⁸ Popper (1997) beschreibt diese Möglichkeit folgendermaßen: „Der Epiphänomenalismus behauptet *nicht*, daß *alle* materiellen Prozesse einen psychischen Aspekt haben“ (Popper & Eccles, 1997, S.102, Herv. i. Original).

vor etwa 2500 Jahren wurde die Theorie aufgestellt, daß nur ein bestimmter Teil des Körpers, vorrangig das Gehirn, der Sitz der Seele sei⁵⁹. Auch René Descartes nimmt an, daß

„...der Geist nicht von allen Körperteilen unmittelbar beeinflußt wird, sondern nur vom Gehirn oder vielleicht sogar nur von einem ganz winzigen Teile desselben“ (zit. nach Meier-Oeser, 1997, 6. Meditation/ Satz 20).

Dieser spezielle Teil des Gehirns wird von Descartes mit der Zirbeldrüse (*conarium*) identifiziert, da diese das einzige unpaarige⁶⁰ Organ des Gehirns ist (nach Zoglauer, 1998, S.50f.). Die Zirbeldrüse soll nach Descartes Theorie nun den Austausch zwischen Körperlichem und Geistigem ermöglichen, indem ihr aus Animalgeistern hergestellte Bilder vorgehalten werden und damit Körperliches für die Seele erfahrbar wird, bzw. indem die Seele durch Neigung der Drüse auf bestimmte motorische Nervenendungen Körperliches bewirken kann (siehe Specht, 1966, S.121ff.).

Wittgenstein bekämpft die Vorstellung, daß es einen *Ort* des Denkens gebe. Falls wir nämlich annehmen, daß wir mit dem Kopf denken, so müßte man auch sagen, wir denken mit der Hand, die schreibt, und dem Mund, der spricht (siehe Wittgenstein, 1997a). Dieses antimentalistische Argument besagt, daß Denken ein nicht intrinsischer Zustand ist, der nicht im Kopfe eines Individuums lokalisierbar ist, sondern eine Tätigkeit darstellt, „die zumindest insoweit nicht bloß mental ist, als dabei externe, überindividuell gültige Symbolsysteme ins Spiel kommen“ (Krämer, 1994, S.92).

Der Medizin-Nobelpreisträger Francis Crick argumentiert dagegen neurophysiologisch: Man könne zwar niemals ein vollkommen klar umgrenztes Zentrum der Willensfreiheit und des Bewußtseins finden werden, jedoch seien besonders solche Hirnareale für den „Sitz der Seele“ prädestiniert, welche ihre Inputs aus höheren Ebenen der sensorischen Systeme empfangen und gleichzeitig ihre Outputs an ebenfalls höhere Ebenen von effektorischen Systemen weitergeben. Untersuchungen zu Hirnschädigungen zeigten, daß vor allem der Bereich um das *anteriorem Cingulum* diese notwendigen Verbindungseigenschaften aufweist (siehe Crick, 1997, S.326ff.). Gleichzeitig erachtet es Crick für vielversprechend, den Thalamus⁶¹ aufgrund seiner charak-

⁵⁹ Popper (1997) nennt Alkmaeon, Hippokrates und Platon als die Wegbereiter dieser Theorie.

⁶⁰ Da Descartes von einem einzigen Bewußtsein und einer einzigen Seele ausging, suchte er dementsprechend nach einem Organ, das in seiner Gestalt nur einzeln vorkommt.

⁶¹ auch Grush & Churchland (1996) erachten Netzwerkeffekte im Bereich des Thalamus als verantwortlich für Bewußtseinseffekte.

teristischen *Gamma-Oszillationen* als hauptsächlich für Bewußtseinsphänomene *sensu* Aufmerksamkeit verantwortlich zu machen (Crick, 1997, S.299ff.).

William H. Calvin lehnt dagegen jede feste Zuordnung von Bewußtseinsfähigkeit zu einem bestimmten Hirnbereich *a priori* ab (vgl. Rosenfield, 1996), indem er darauf hinweist, daß

„keine bestimmte Region im Cortex sehr lange ‚Zentrum des Bewußtseins‘ ist; rasch übernimmt eine andere Region diese Funktion“ (Calvin, 1998, S.210).

Dies ist auch der Grund für die in Kapitel 4.3 beschriebene hohe Flexibilität des Gehirns: Werden z.B. einzelne Neuronengruppen durch bestimmte Einwirkungen geschädigt, so gestattet die enorme Plastizität des Gehirns, daß andere Hirnareale zumindest Teile der gestellten Arbeitsanforderungen übernehmen können. Gehirnläsionen können so effektiv kompensiert werden.

Auch Leibniz stand der Suche nach einem definitiven Zentrum des Geistigen skeptisch gegenüber, allerdings aus dem Grund, daß man durch die Ortsangabe noch nichts über die zugrundeliegenden Mechanismen und Ursachen aussagt. Außerdem kritisiert er konkret die Vorgehensweise von Descartes:

„Allerdings hat Descartes der Seele engere Schranken geben wollen, indem er ihr die Zirbeldrüse als eigentlichen Sitz anwies, aber er hat gleichwohl nicht zu sagen gewagt, daß sie ausschließlich in einem Punkt dieser Drüse sich befinde; er hat damit also gar nichts gewonnen, und es ist gerade ebenso, als wenn man ihr den ganzen Körper zum Kerker oder Aufenthaltsorte anwiese“ (Leibniz, 1904, S.349).

In diesem Sinne soll und kann dieser Exkurs auch nicht entscheiden, welcher Ortspunkt oder welches Organ als der Ursprung bewußter oder geistiger Tätigkeit angesehen werden muß. Vielmehr soll dieser Abschnitt als Schaustück dienen, welche Entscheidungskriterien einzelne Philosophen und Naturwissenschaftler verwenden, um zu entscheiden, ob bestimmte Teile des menschlichen Organismus prinzipiell in der Lage wären, der Ursprungsort von Bewußtsein oder von geistigen Phänomenen allgemein zu sein.

6.3.2.5 Zusammenfassung

Wie schon an anderer Stelle besprochen, ist Leibniz der Ansicht, daß es eine scharfe Trennungslinie zwischen geistigen Phänomenen, wie etwa Bewußtsein, Wünschen, Gedanken, Emotionen und Empfindungen einerseits, und physischen Phänomenen andererseits gibt. Zur Illustration seiner Ansicht verwendet Leibniz ein Gedankenexperiment. Dazu soll man sich vorstellen, auf die Größe eines winzigen Wesens zu schrumpfen, damit man sich die inneren Vorgänge im Hirn wie eine große Maschi-

nenhalle anschauen kann. Nach Leibniz Meinung würde uns dies auch kein bißchen im Verständnis über unsere geistigen Vorgänge weiterbringen, da geistige Phänomene einen ganz anderen Teil der Realität darstellen als physische Phänomene, und somit unbeobachtbar sind (siehe Churchland, 1997).

Eine andere Möglichkeit, warum wir aus dieser direkten Beobachtung keine Erkenntnis über den Geist gewinnen würden, könnte aber auch sein, daß wir das Beobachtete nur nicht richtig deuten und klassifizieren können. So wie eine fremde Sprache zuerst lediglich unspezifisch als „Kauderwelsch“ identifiziert wird, so können wir auch die Sprache des Gehirns anfangs nicht richtig deuten. Dies kann sich aber mit fortschreitendem Wissen über diese „Sprache“ allmählich ändern. Einzelne *Zeichen* oder *Phoneme* werden dann deutlicher und mit Sinn besetzt. Solange wir nicht wissen, nach was wir überhaupt suchen müssen, und wir keinerlei Vorstellung davon haben, wie wir das Gesuchte identifizieren können, scheint es jedoch nicht angebracht, dieses Rätsel als „generell unlösbar“ einzustufen.

Das noch ungelöste Problem, was der Geist ist und wie er zustande kommt, durch eine dualistische Theorie von Leib und Seele zu einem unaufklärbaren Rätsel zu mystifizieren, scheint in Anbetracht der immensen Wissensfortschritte der Naturwissenschaften in all jenen Bereichen, die ebenfalls noch vor relativ kurzer Zeit *Mysterien* darstellten, obsolet und wissenschaftlich inakzeptabel. In diesem Sinne kann ich Paul Churchland voll und ganz beipflichten, wenn er sagt:

„Daß unser Bewußtsein der Wissenschaft momentan noch als einzigartiges und für immer unerreichbares Mysterium erscheint, könnte nur unsere eigene Unwissenheit und unsere gegenwärtige Konzeptionslosigkeit widerspiegeln und nicht irgendeine besondere, metaphysische Eigenschaft des Bewußtseins“ (Churchland, 1997, S.221, Herv. v. Verf.).

Gilbert Ryle als auch Rudolf Carnap gehen in diesem Punkt noch einen entscheidenden Schritt weiter, indem sie die Unterscheidung zwischen körperlichen und inneren mentalen Phänomenen als grundsätzlich sinnlos bezeichnen. Ihres Erachtens handelt es sich dabei lediglich um sogenannte *Scheinprobleme* (siehe Oeser & Seitelberger, 1995). Diese Probleme kommen alleine dadurch zustande, daß das mentalistische Vokabular entweder gar nicht oder nur unzureichend analysiert wird.

Auch Putnams Argumentation scheint diese Richtung zu verfolgen:

„Die verschiedenen Streitpunkte und Rätsel, aus denen sich das traditionelle Körper–Geist–Problem zusammensetzt, sind ihrem Wesen nach rein sprachlich und logisch“ (Putnam, 1994, S.146).

Seiner Meinung nach entstehen all die Streitpunkte über den Körper–Geist–Dualismus auch bei jedem beliebigen Rechensystem, das Fragen über seine eigene Struktur beantworten kann.

6.3.3 “Chauvinistische” Theorien

Die Begrifflichkeit *Chauvinistische Theorien* wird hier in Anlehnung an Dennett (1996a) verwendet, der unter *Chauvinismus der Herkunft* diejenige Kategorie von Ansichten versteht, die einen Unterschied zwischen intrinsischen Eigenschaften von Entitäten unterstellt, der lediglich auf der Herkunft der Entitäten beruht⁶².

6.3.3.1 Kohlenstoff–Chauvinismus

Während man in alten Zeiten noch glaubte, daß der menschliche Geist als eigene Substanz unzerstörbar und unverändert aus jeder körperlichen Krankheit wieder auftaucht und auch den Tod des Leibes übersteht, hat uns die medizinische Hirnforschung eines anderen belehrt (siehe Oeser & Seitelberger, 1995). Demnach sind geistige Qualitäten nämlich an einen bestimmten physikalischen Aufbau gebunden, der mit dem biologischen Tod eines Wesens nicht mehr gewährleistet ist. Die Frage ist, ob nun auch die spezifische stoffliche Zusammensetzung wesentlich für die Ausbildung geistiger Fähigkeiten ist.

In diesem Zusammenhang existieren einige Theorien, die die Existenz von echter geistiger Qualität nur auf Basis eines biotischen Systems realisierbar sehen. So konstatiert z.B. Paul Ziff: “Only living creatures can literally have feelings” (zit. nach Birnbacher, 1996, S.722) und Julian Huxley meint, daß *echte* Intentionen, Denkakte und Verständnisakte nur in biotischer Masse möglich seien. Nach Angabe von Churchland & Churchland (1990) deutet auch John Searle mehrfach an, daß seiner Meinung nach alle Ebenen, einschließlich der biochemischen, in jeder denkfähigen Maschine repräsentiert sein müssen (nach Churchland & Churchland, 1990). Tatsächlich scheint für Searle die Substanz eines möglichen künstlichen Gehirns wesentlich zu sein, zumindest wenn dieses „Gehirn“ die Fähigkeit besitzen soll, Bewußtsein zu entwickeln. Allerdings legt sich Searle nicht, wie von Churchland & Churchland (1990) behauptet⁶³,

⁶² Auch Putnam spricht von „Wasser– und Kohlenstoff–Chauvinismus“, der sich dadurch zeigt, daß postuliert wird, Kohlen– und Wasserstoff besitze *per se* (mehr) Bewußtsein (nach Oeser & Seitelberger, 1995, S.114).

⁶³ Es ist tatsächlich sehr problematisch, sich auf Beschreibungen der Ansichten von John Searle aus einer Publikation von Churchland & Churchland zu berufen, da es eine bereits sehr lange andauernde heftige Kontroverse in der Literatur zwischen diesen beiden Lagern gibt. Mit einer objektiven Darstellung des jeweiligen „Kontrahenten“ ist daher (leider) kaum zu rechnen.

auf biotische Materie als Grundstoff fest (siehe auch Oeser & Seitelberger, 1995; Wiener, 1990), sofern eine alternative Materie auch kausale Kräfte entwickeln kann:

“An ‘artificial brain’ might cause consciousness though it is made of some substance totally different from neurons, but whatever substance we use to build an artificial brain, the resulting structure must share with brains the causal power to get us over the threshold of consciousness” (Searle, 1997, S.191).

Aber auch Funktionalisten sehen keine Notwendigkeit, daß nur biotische Systeme ein Bewußtsein besitzen können. Da mentale Phänomene als abstrakte funktionale Zustände des gesamten Organismus verstanden werden, ist nur die *Funktion*, nicht aber die *Aufbauweise* eines Systems relevant, um ihm mentale Eigenschaften zuzuschreiben. Daher sind die Begriffe für mentale Zustände nicht notwendig an die gewohnte irdisch–organische Hardware gebunden (vgl. Wolters, 1993).

Birnbacher (1996) gibt zu bedenken, daß es wenig Sinn macht, die stoffliche Grundlage des Gehirns zum Gegenstand der Diskussion zu machen, falls Bewußtsein auf höheren Eigenschaften eines Netzwerks basiert:

„Die nomologische Möglichkeit eines Maschinenbewußtseins entscheidet sich an der Frage, auf welcher *Ebene* der physischen Welt die strukturellen und funktionalen Eigenschaften angesiedelt sind [...] Falls Bewußtseinsprozesse nicht so sehr von der *atomaren* Struktur neuronaler Netzwerke abhängen als vielmehr von den *systemaren* und *funktionalen* Eigenschaften dieser Netzwerke, wäre es nachgerade ein kosmischer Zufall, wenn die Emergenz von Bewußtsein an genau *diejenigen* stofflichen Elemente geknüpft wäre, aus denen die Verschaltungen im biologischen Gehirn bestehen“ (Birnbacher, 1996, S.725, Herv. i. Original).

Karl Popper dagegen argumentiert evolutionstheoretisch und gesteht „toter Materie“ zu, mehr hervorzubringen als wiederum nur tote Materie:

„[...] es sieht ganz so aus, als ob das Leben mit der chemischen Synthese von großen sich selbstreproduzierenden Molekülen entstand, und daß es sich durch natürliche Auslese weiterentwickelte [...].

Es scheint also, daß in einem materiellen Universum etwas Neues auftauchen kann. Tote Materie kann, so scheint es, mehr hervorbringen als tote Materie. Insbesondere hat sie das Bewußtsein hervorgebracht“ (Popper & Eccles, 1997, S.30f.).

Dennoch ist es vorstellbar, wenn auch unwahrscheinlich, daß die schiere Geschwindigkeit und Kompaktheit von biochemischen Prozessen, die im Gehirn ablaufen, tatsächlich in keinem anderen Medium reproduzierbar sind (siehe Dennett, 1996a).

6.3.3.2 Sprachfähigkeit = Bewußtsein

Die Idee, daß Sprachfähigkeit die Voraussetzung für Bewußtsein ist, ist eng mit der Vorstellung verknüpft, daß nur Menschen ein Bewußtsein entwickeln können (siehe Kapitel 6.3.3.3). Oft werden auch die inneren Abläufe des Bewußtseins mit einer menschenähnlichen, auf klarer Syntax und Logik (siehe McCarthy, 1996) aufbauenden Sprache, gleichgesetzt (siehe Kapitel 3.2.2). Dies zeigt sich auch im Vorgehen der klassischen KI, die mit Hilfe einer sprachähnlichen Struktur künstliche Intelligenz hervorbringen will.

So sah auch Descartes die Sprache als das untrüglichste Kriterium um zwischen Maschinen, die kein Bewußtsein haben, und Menschen, die Bewußtsein besitzen, zu unterscheiden (vgl. Calvin, 1989a):

„Keine unserer äußeren Tätigkeiten vermöchte diejenigen, die sie untersuchen, zu überzeugen, daß unser Körper nicht nur eine Maschine ist, die sich selbst bewegt, sondern daß es in ihm außerdem eine denkende Seele gibt, es seien denn unsere Wörter“ (zit. nach Specht, 1966, S.116).

Auch Silbernagl & Despopoulos (1988) erachten Sprache als einen wesentlichen Teil des menschlichen Bewußtseins:

„[Sprache] dient einerseits als zwischenmenschliches *Kommunikationsmittel* [...]. Andererseits ist Sprache auch notwendig, um Sinneseindrücke bewußt zu verarbeiten, d.h. *Begriffe* zu erarbeiten, *Konzepte* zu erarbeiten [...]. Erst mit dieser Begriffsbildung und ihrer Verbalisierung ist eine *ökonomische Speicherung* im Gedächtnis möglich“ (Silbernagl & Despopoulos, 1988, S.294).

Paul Churchland vertritt gegenüber dieser klassischen Meinung die Ansicht, daß zwar die Sprachfähigkeit von Menschen den wesentlichen Vorteil hat, daß Menschen über ein „körperloses Gedächtnis“ verfügen und daß Lernprozesse nicht mehr auf die Lebenszeit von Individuen beschränkt sind; dennoch ist Bewußtsein nicht zwingend an Sprache gebunden (siehe Churchland, 1997, S.317ff.; Churchland & Churchland, 1990).

6.3.3.3 Humanoide Exklusivität

Hauptsächlich existieren zwei Argumentationsstränge, die begründen wollen, daß nur der Mensch über Bewußtsein verfügen kann. Der eine argumentiert auf einer rein begrifflichen Ebene: Bewußtsein generell wird, aus in der Semantik des Bewußtseinsbegriffs liegenden Gründen, mit menschlichem Bewußtsein gleichgesetzt (vgl. Tetens, 1994). Auf dieser Ebene argumentiert Wittgenstein, wenn er sagt

„man könne nur vom lebenden Menschen und was ihm ähnlich ist, (sich ähnlich benimmt) sagen, es habe Empfindungen; es sähe; sei blind; sei taub; sei bei Bewußtsein, oder bewußtlos“ (Wittgenstein, 1997b, §281).

Wittgenstein will damit nicht sagen, daß sich die Hypothese einer bewußtseinsbegabten nicht-humanoiden Entität nicht denken läßt. Vielmehr gäbe es aber keinen guten Grund, eine solche Hypothese zu vertreten, da eine solche These mangels empirischer Kriterien *leer* wäre (vgl. Birnbacher, 1996, S.719ff.). Die Aussage „Eine Maschine kann doch nicht denken!“ ist daher kein Erfahrungssatz sondern eine begriffliche Selbstverständlichkeit, da wir nur vom Menschen sagen, er denke (vgl. Tetens, 1993).

Die andere Argumentationsebene könnte man laut Dennett (1996a) als „Chauvinismus der Herkunft“ bezeichnen: Dem Menschen wird dabei aufgrund seines spezifischen Aufbaus eine Exklusivität bescheinigt, die nur ihn dazu befähigt, Bewußtsein zu entwickeln. Roboter werden nach dieser Logik schlichtweg immer viel zu einfach oder vom Menschen zu unterschiedlich sein, um Bewußtsein zu besitzen.

Tatsächlich handelt es sich dabei um eine Argumentation rein *technischer* Natur. Jedoch zeigte sich im Laufe der Geschichte, daß schwierige technische Probleme, deren Lösung angeblich unmöglich war, später jedoch durchaus erfolgreich gelöst werden konnten. Zwar ist es wahr, daß das menschliche Gehirn aus Billionen von Einzelteilen besteht und somit extrem schwierig nachzubilden ist, jedoch hält Dennett (1996a) es nur für eine Frage der Zeit, wann diese Komplexität mit Hilfe künstlicher Strukturen möglich ist. In diesem Sinne meint Warwick (1998):

“...consciousness is provided a standard to be surpassed in time. Whichever way we look at it, there is only one conclusion: it is only a matter of time before there is nothing humans can do to stop machines becoming more intelligent than humans” (Warwick, 1998, S.185).

6.3.4 Materialismus/Physikalismus⁶⁴

Der alte Begriff des ontologischen Materialismus ist heute nicht mehr haltbar: Materie im klassischen Sinn hat sich mit den Fortschritten der Naturwissenschaft aufgelöst.

⁶⁴ der Begriff geht vermutlich auf Otto Neurath zurück (siehe Popper & Eccles, 1997, S.27f.).

Zwangsläufig mußte daher der Begriff *Materialismus* abgelöst werden und wird heute üblicherweise durch *Physikalismus*⁶⁵ ersetzt (nach Oeser & Seitelberger, 1995).

Die Hauptfrage, die physikalistische Theorien in Bezug auf das Bewußtsein beschäftigt, kreist um die von Julian Huxley gestellte Frage:

“How it is that anything so remarkable as a state of consciousness comes about as a result of initiating nerve tissue, is just unaccountable as the appearance of the Djin, where Aladdin rubbed his lamp in this story...” (zit. nach McGinn, 1996, S.272).

6.3.4.1 Elimination des mentalistischen Vokabulars

Ein historischer Vorläufer des eliminativen Materialismus ist der *logische Behaviorismus*, der hauptsächlich von Gilbert Ryle⁶⁶ verfolgt wurde. Für den *logischen Behaviorismus* sind mentale Zustände nichts anderes als Verhaltensdispositionen, d.h. das gesamte mentalistische Vokabular läßt sich in bedeutungsgleiche Verhaltenskonditionale übersetzen (vgl. Zoglauer, 1998). So würde z.B. ein Satz wie „Hans hat starke Zahnschmerzen“ transformiert werden können in einen Satz wie „Hans zeigt ein typisches Schmerzverhalten: Er stöhnt und hält sich einen Eisbeutel an die Wange und berichtet über große Schmerzen“. Diese nach dem Verständnis des *logischen Behaviorismus* bedeutungsgleichen Sätze sind wahrheitsäquivalent, d.h. daß unter gleichen Bedingungen beide Sätze wahr oder falsch sind. Somit ließe sich jegliches psychologisches Vokabular eliminieren und durch ein behavioristisches Vokabular substituieren. Der *logische Behaviorismus* folgt damit dem positivistischen Sinnkriterium, nach dem der Sinn eines Satzes durch seine Verifikationsbedingungen festgelegt ist. Im Gegensatz zu mentalistischen Beschreibungen ist dieses Verifikationskriterium bei Verhaltensbeschreibungen oder Aussagen über physikalische Zustände erfüllt (vgl. Tetens, 1994). Außerdem wird das mentalistische Vokabular von Seiten des *logischen Behaviorismus* verurteilt, da es unwissenschaftlich und irreführend ist (vgl. Lyons, 1996).

Gegen das Programm einer vollständigen Eliminierung von mentalistischem Vokabular spricht, daß man viele mentale Zustände auch phänomenologisch anders darstellen kann, d.h. man kann z.B. Trauer, Freude oder Schmerzen zu erheblichen Teilen simulieren oder vereiteln. Dies zeigt, daß mentale Ereignisse nicht isomorph mit entspre-

⁶⁵ vermutlich aus traditionellen Gründen wird dennoch teilweise der Begriff *Materialismus* weiterbenutzt, so etwa auch in der in Kapitel 6.3.4.2 beschriebenen physikalistischen Theorie des *eliminativen Materialismus*.

⁶⁶ allerdings muß man vorsichtig sein, Gilbert Ryle unreflektiert in die Schublade des *logischen Behavioristen in Reinform* zu stecken, da er durchaus zwischen geheuchelten und echten Gefühlsbekenntnissen unterscheidet (vgl. Popper & Eccles, 1997, S.139ff.)

chenden Verhaltensäußerungen sind. C.G. Hempel versucht, dieses Defizit dadurch zu kompensieren, daß zusätzlich zu den Verhaltensbeschreibungen auch noch physiologische und neurologische Zustände angegeben werden, um nicht auf mentalistisches Vokabular zurückgreifen zu müssen. Tatsächlich wird die zusätzliche Berücksichtigung der neuronalen und physiologischen Ebene eine genauere und treffendere Beschreibung des jeweiligen mentalen Zustandes gewährleisten. Es gilt allerdings zu bezweifeln, ob dies hinreichend dafür ist, mentale Zustände in ihrem ganzen Umfang und vor allem in ihrem Wesen adäquat abbilden zu können.

6.3.4.2 Eliminativer Materialismus

Der eliminative Materialismus verurteilt unsere normale Konzeption von psychologischen Phänomenen als eine radikal falsche Theorie. Ihr erklärtes Ziel ist, mentale Phänomene alleine durch die Neurowissenschaften erklären zu können (vgl. Churchland, 1996a), d.h. psychische Phänomene sollen sich vollkommen auf neuronale Aktivitäten reduzieren lassen. In der Klassifikation von Searle (1996) handelt es sich dabei um eine Theorie *kausaler* und *ontologischer* Reduktion (vgl. Anonymus, 1999b).

Kambartel (1993) kritisiert die Meinung, man könne Denken oder Bewußtsein vollständig auf neuronale Aktivitäten reduzieren:

„Wir denken also weder im Gehirn noch mit Hilfe des Gehirns. Und schon gar nicht denkt unser Gehirn selbst, oder bewirkt physiologisch unser Denkhandeln“ (Kambartel, 1993, S.227).

Wohl aber räumt er ein, daß es physiologische Bedingungen gibt, die das Denken ermöglichen, erleichtern oder erschweren.

Die prominentesten Vertreter des eliminativen Materialismus sind z.Zt. Paul und Patricia Churchland, die in starker Anlehnung an Francis Crick, auch mentale Phänomene wie Bewußtsein, Entscheidungsfindung und Intuition einzig und alleine auf eine physikalische Basis stellen wollen (vgl. Lyons, 1996; Zoglauer, 1998, S.119ff.). So nimmt Crick (1997) an, daß das Bewußtsein durch ein korreliertes resp. frequenzspezifisches Feuern bestimmter Neuronengruppen erklärt werden könnte.

Paul Churchland gibt zu bedenken, daß auch wenn das Unternehmen, den Geist vollständig verstehen zu können, heute noch nicht vorstellbar sei, dies dennoch kein Indiz dafür ist, daß es nicht generell möglich ist. Er will dem Rätsel des Geistes mit Hilfe eines zweiphasigen Prozesses näherkommen.

Die erste Phase besteht darin, anzuerkennen, daß

„die Materie selbst per se nie lebendig oder tot ist. Vielmehr ist eine bestimmte, komplex organisierte Materie lebendig, wenn sie bestimmte Funktionen zeigt, und sie ist tot, wenn das nicht der Fall ist“ (Churchland, 1997, S.223).

Hat man diese Ansicht erst einmal angenommen, müssen die molekularen Strukturen analysiert werden, die das Leben ausmachen. Als Hauptträger dieser Aufgabe sieht Churchland die Molekularbiologie. Sie soll „alle zentralen Mechanismen der Lebensvorgänge biochemisch entschlüsseln und zu einem Gesamtkonzept verbinden“ (Churchland, 1997, S.223).

Churchland vergleicht dieses Forschungsvorhaben mit dem mühsamen Weg der Erforschung des Erbgutes, welches erst mit der Entdeckung der DNA in den 1950er Jahren einen explosionsartigen Aufschwung erlebte. In diesem Sinne steht er dem Unternehmen, den Code, den das Gesamtsystem Hirn hervorbringt, zu entschlüsseln generell positiv gegenüber. Aufgrund der mangelnden Vorstellungskraft des Menschen sei dieses Unternehmen vielleicht zweifelhaft, jedoch gilt zu bedenken:

„Die Tatsache, daß uns ein Problem mysteriös erscheint, ist eine Aussage über uns, eine bedauerliche Folge unseres gegenwärtigen Wissens und Denkens; es sagt überhaupt nichts über das Problem selbst aus, nichts, aus dem man irgendwelche prinzipiellen Schlußfolgerungen ableiten könnte“ (Churchland, 1997, S.224).

Zusätzlich weist Churchland (1997) das Argument zurück, daß der Geist niemals wissenschaftlich erforschbar sein wird, da er nicht *objektiv*, d.h. durch „Fremde“ erschließbar ist. Für ihn scheinen solche Argumentationsweisen lediglich dafür geeignet, um *jegliches* Forschungsvorhaben *a priori* zu unterminieren:

„Die Existenz eines eigenen subjektiven Zugangs zu irgendeinem Phänomen bedeutet nicht, daß dieses Phänomen nichtphysischer Natur sein muß“ (Churchland, 1997, S.232).

Paul Churchland erläutert in seinem Buch *Die Seelenmaschine* (Churchland, 1997) ausführlich die Vorteile von neuronalen Netzen, die schon in den vorangegangenen Kapiteln diskutiert wurden (siehe Kapitel 5). Sein persönliches Hauptaugenmerk liegt auf den sogenannten *rekurrenten Netzen*, die zusätzlich zu normalen Feedforward-Netzwerken die Fähigkeit besitzen, „kohärente Verhaltenssequenzen zu erzeugen [... und] zeitlich ausgedehnte Kausalvorgänge zu erkennen“ (Churchland, 1997, S.249). Durch diese Eigenschaften kann man die menschliche Fähigkeit verstehen, „ein und dieselbe Sache auf mehrere unterschiedliche Arten wahrzunehmen, zu verstehen und zu interpretieren“ (Churchland, 1997, S.249).

Mit einer schlagkräftigen Simulation eines Pseudobewußtseins, das von einem rekurrenten Netzwerk gebildet wird, gelingt es Churchland, wesentliche Eigenschaften des Bewußtseins erfolgreich darzustellen. Auch andere Phänomene wie das Kurzzeitgedächtnis können mit Hilfe solch eines neuronalen Netzwerkes dargestellt werden, da die für ein Gedächtnis relevante Eigenschaft, daß Informationen über mehrere Prozeßzyklen bereitstehen, in einer rekurrenten Netzstruktur systemimmanent vorhanden ist. Damit ist die kontinuierliche kognitive Aktivität in einem solchen Netzwerk — vergleichbar mit unserem Bewußtsein — nicht auf den ununterbrochenen Strom externer Sinnesreize angewiesen: Das simulierte Bewußtsein generiert sich selbst!

Selbst Phänomene des Träumens werden von Churchlands Simulation beleuchtet, da sie erklären kann, warum Träume normalerweise einen so realistischen und prototypischen Charakter besitzen. Fehlen nämlich die

„üblichen Kontrollen des rekurrenten Netzwerks durch seine sensorischen Eingänge, dann bewegen sich die Vektoren offenbar hauptsächlich auf den Bahnen durch die Merkmalsräume, die schon vorher ‚ausgetreten‘ waren, nämlich denen der Prototypen (Churchland, 1997, S.262).

Churchland demonstriert eindrucksvoll viele weitere Eigenheiten des Gehirns mit Hilfe seiner neuronalen Simulation. Er nimmt mit seinem Modell eine reduktionistische Stellung ein, da er dem Netzwerk nichts weiter mitgibt als seine physikalische Struktur, bestimmte, durch Trainingsdurchgänge selbst gezogene Verbindungsstärken, und genügend Rechenkapazität, um die nötigen Verrechnungen in Echtzeit ablaufen zu können, damit menschliche Kognitionen adäquat abgebildet werden kann. Auch das Bewußtsein selbst ist nach seiner Ansicht alleine mit Hilfe dieser Mittel erklärbar, d.h. die Annahme der nötigen Existenz von weiteren Elementen wie „Lebenshauch“, „Seele“ oder „Gott“ wird überflüssig.

In Zusammenarbeit mit Patricia Churchland versucht auch Francis Crick den Akt einer Willensentscheidung auf eine rein physikalische Ebene zu transduzieren. Dazu stellt er drei Prämissen auf (siehe Crick, 1997, S.325):

1. Ein Teil des Gehirns ist damit beschäftigt, Pläne für das Handeln in der Zukunft zu machen, die natürlich nicht unbedingt ausgeführt werden. Man kann sich solcher Pläne bewußt sein, d.h. es ist wenigstens möglich, diese Pläne unmittelbar abzurufen.
2. Man ist sich nicht der „Berechnungen“ bewußt, die der betreffende Teil des Gehirns anstellt, sondern nur der „Entscheidungen“, die sich daraus ergeben.
3. Das gleiche gilt für die Entscheidung, den einen oder den anderen Plan auszuführen, d.h.: Der Inhalt der Entscheidungen, aber nicht die Berechnun-

gen, die in sie eingegangen sind, können unmittelbar abgerufen werden, auch wenn man sich des Plans bewußt ist, eine bestimmte Bewegung auszuführen.

Crick kommt daher zu dem Schluß, daß eine Maschine, für die diese drei Prämissen gilt, den Eindruck von sich selbst hat, Willensfreiheit zu besitzen (siehe Crick, 1997, S.324f.). Eine derartige Maschine kann versuchen, sich selbst mittels Introspektion zu erklären, warum sie eine bestimmte Entscheidung getroffen hat. Gelegentlich gelangt sie vielleicht sogar zu der richtigen Schlußfolgerung. Da sie von den wahren Gründen für die Entscheidung aber keinerlei bewußte Kenntnis hat, wird die Maschine jedoch lediglich konfabulieren.

Eine gern eingebrachte Argumentationsweise gegen den eliminativen Materialismus folgt einer *reductio ad absurdum* (siehe Geach, 1976); die Argumentationskette sieht dabei folgendermaßen aus (siehe Zoglauer, 1998, S.126f.):

1. Wenn der eliminative Materialismus wahr ist, dann gibt es keine mentalen Phänomene.
2. Glauben, Meinungen und Überzeugungen sind mentale Phänomene.
3. Aus Prämisse 1 und Prämisse 2 folgt, daß es, sollte der eliminative Materialismus wahr sein, keine Meinungen gibt.
4. Paul und Patricia sind überzeugte eliminative Materialisten: Sie meinen resp. sind davon überzeugt, daß der eliminative Materialismus wahr ist.

Aus Prämisse 3 und Prämisse 4 folgt, daß Paul und Patricia meinen, daß sie keine Meinung haben. Dies ist jedoch eine Kontradiktion!

Patricia Churchland hält diesem Argument jedoch entgegen, daß der aus Punkt 3 und Punkt 4 gefolgerte Widerspruch nur deshalb zustande kommt, weil sich auch eliminative Materialisten gezwungenermaßen im Rahmen der Alltagspsychologie und deren Sprachtradition bewegen müssen. Dieser Widerspruch wird sich aber sofort auflösen, wenn es einmal Alternativen zu dieser Alltagspsychologie gibt. Tatsächlich hat Churchlands Gegenargument keinerlei praktische Grundlage, da eine solche Alternative bis heute nicht existiert. Der eliminative Materialismus bleibt also widersprüchlich.

Eine völlig andere Argumentationsebene nimmt Donald Davidson ein, der die These vertritt, es gäbe für *sinnhaltige* Zusammenhänge keine Gesetze. Der eliminative Materialismus versucht aber gerade, den Sinn von Handlungen, die normalerweise mit *Intentionalerklärungen* beschrieben werden, durch *Kausalerklärungen* zu bestimmen (siehe Føllesdal, Walløe, & Elster, 1988, S.158f.). Dies ist aber nicht immer möglich, da es keine allgemeinen Gesetze darüber gibt, daß bestimmte Präferenzen und Annahmen immer zu bestimmten Handlungen führen. Es gibt Situationen, in denen das

Zusammenspiel von einer bestimmten Präferenz mit einer bestimmten Überzeugung zu einer ganz spezifischen Handlung führt. Man kann diese Präferenzen und Überzeugungen aber nicht auf physiologische Zustände reduzieren, welche ihrerseits in Kausalgesetze einfließen könnten. Als Grund dafür gibt Davidson an, daß in den Argumenten, aufgrund derer wir jemandem bestimmte Präferenzen und Annahmen zuschreiben, unvermeidlich ein normatives, also regelhaftes Element vorkommt. Um dieses normative Element in den Handlungen des Betreffenden rechtfertigen zu können, muß es sich allerdings um ein streng logisches und rational denkendes Wesen handeln, d.h. wir müssen davon ausgehen, daß zwischen ihren Präferenzen und Annahmen ein gewisser innerer Zusammenhang besteht, was jedoch praktisch nie gegeben ist.

Auch Searle (1996) meint, daß man beim Bewußtsein die *ontologische* Reduktion nicht vollziehen kann. Bewußtsein ist eine kausal emergente Eigenschaft des Verhaltens von Neuronen; mithin ist Bewußtsein auf Hirnvorgänge zurückführbar, d.h. kausal auf diese reduzierbar. Trotzdem würde selbst eine vollkommene Wissenschaft des Gehirns nicht zu einer *ontologischen* Reduktion von der Art führen, wie dies bei Phänomenen wie Wärme, Farbe oder Klang der Fall ist. Searle erklärt diesen Sachverhalt damit, daß keine Beschreibung der objektiven und physiologischen Dritte-Person-Tatsachen den subjektiven Erste-Person-Charakter adäquat abbilden könne (siehe Searle, 1996, S.137ff.). Allerdings räumt er ein, daß geistige Makro-Phänomene allesamt von Mikro-Phänomenen niedrigerer Stufe verursacht werden.

6.3.4.3 Das Autocerebroskop

Mit den heutigen Methoden wie der Elektroenzephalographie (EEG), Computertomographie, Positron-Emissions-Tomographie (PET) und Kernspintomographie lassen sich vielfältige Störungen und Prozesse im Gehirn aufzeigen. Allerdings gelingt es bis heute noch nicht, das Gehirn auf neuronaler Ebene vollständig zu identifizieren. Der Materialist Herbert Feigl meint jedoch, daß es in Zukunft möglich sein wird, ein sogenanntes *Autocerebroskop* zu konstruieren, mit dem man ein vollkommenes Bild aller Gehirnaktivitäten einer bestimmten Person erhalten kann. Während man durch Introspektion seine mentalen Zustände registriert, könnte man dann gleichzeitig mit Hilfe des *Autocerebroskops* eine vollständige Liste von psychophysischen Identifikationsbeziehungen aufstellen in Form von Typ-Identifikationen als auch von Token-Identifikationen (siehe Zoglauer, 1998, S.99ff.).

Der hehre Anspruch Feigls, solch eine Übersetzungsmöglichkeit für alle möglichen mentalen Zustände zu finden, basiert jedoch auf einer recht naiven Vorstellung von Bewußtwerdungsmechanismen. Tatsächlich muß das Unternehmen *Autocerebroskop*

schon alleine an den üblichen Grenzen introspektiver Vorgehensweisen scheitern (siehe Kapitel 6.1.2). Zusätzlich scheint die vollständige Überwachung aller neuronalen Aktivitäten zu einem bestimmten Zeitpunkt und noch mehr deren dynamische Verfolgung über bestimmte Zeitspannen hinweg ein Ding der Unmöglichkeit⁶⁷.

6.3.5 Quantenmechanische Erklärungen

In der klassischen Physik existieren, in Übereinstimmung mit unserem Alltagsverständnis, klare und streng logische Beziehungen in Form von präzise formulierten mathematischen Gleichungen. Die physikalische Wirklichkeitsvorstellung von Newton, Maxwell und Einstein gilt als etwas, das völlig unabhängig von uns existiert, d.h. egal was wir uns für eine Vorstellung von der Welt machen, sie bleibt jeweils genau dieselbe definierbare Welt. Dies bedeutet aber gleichzeitig, daß auch unsere Gehirne und unsere Körper vollständig zu dieser vollständig definierbaren Wirklichkeit gehören: Alle unsere Handlungen und Gedanken sind durch diese physikalischen Gleichungen festgelegt; unser freier und bewußter Wille ist Teil einer komplexen, aber klar formulierbaren Funktion.

Spätestens seit Heisenberg konnte jedoch gezeigt werden, daß es eindeutig meßbare Diskrepanzen zwischen der Vorhersage der klassischen Physik und dem tatsächlichen Verhalten der Welt gibt⁶⁸. Die Quantentheorie stellt nun eine Theorie dar, die fähig ist, diese Lücke zwischen theoretischer und realer Welt zu schließen. Oft werden die quantenmechanischen Effekte jedoch als unbedeutend und nicht alltagsrelevant bezeichnet; tatsächlich liegen sie jedoch durchaus vielen physikalischen Alltagsphänomenen zugrunde⁶⁹. Penrose (1991) fragt sich deshalb, ob nicht auch das Phänomen des Bewußtseins erst durch quantenmechanische Effekte erklärt werden kann (siehe Penrose, 1991, S.218ff.).

⁶⁷ Auch Tetens (1994) ist der Meinung, daß eine solche vollständige Beschreibung neuronaler Aktivität praktisch ausgeschlossen ist: Selbst die Erstellung eines solchen neuronalen Schaltplans ohne zeitliche Komponente würde nach seinen Schätzungen ca. 40.000 Jahre benötigen (siehe Tetens, 1994, S.98).

⁶⁸ Karl Popper faßt die umfassende Bedeutung der Quantenphysik treffend mit Hilfe der zwei Metaphern Uhr und Wolke zusammen (*Uhren* stehen dabei für Exaktheit und *Wolke* für Ungenauigkeit): „So hat sich also die These, daß alle physischen Systeme, einschließlich der Wolken, in Wirklichkeit Uhren seien, als falsch erwiesen. Nach der Quantenmechanik müssen wir sie durch die folgende These ersetzen: Alle physischen Systeme, einschließlich der Uhren, sind in Wirklichkeit Wolken“ (Popper & Eccles, 1997, S.59).

⁶⁹ man denke nur an „die bloße Existenz fester Körper, die Härten und physikalischen Eigenschaften von Werkstoffen, das Wesen der Chemie, die Farben von Substanzen, das Phänomen des Gefrierens und Siedens, die Verlässlichkeit des Erbguts — all diese und viele andere vertraute Eigenschaften erfordern zur Erklärung die Quantentheorie“ (Penrose, 1991, S.218).

Für seine Theorie der quantenmechanischen Erzeugung⁷⁰ von Bewußtseinsphänomenen sieht er bestimmte Zellorganellen als essentiell an. Diese *Mikrotubuli* genannten Zellorganellen sind nach seiner Ansicht deswegen so interessant, da sie bestimmte quantenmechanische Prozesse wie zum Beispiel Superstrahlung ermöglichen. Außerdem spielen Mikrotubuli eine zentrale Rolle für die Funktion von Neuronen.

Da die Änderung eines neuronalen Zustands, also ob ein einzelnes Neuron feuert oder nicht, von seinem energetischen Schwellenwert abhängt, kann man sich leicht vorstellen, daß bereits winzige Energieschwankungen das Verhalten eines einzelnen Neurons beeinflussen können. Da Neuronen jedoch stets innerhalb großer Neuronengruppen verschaltet sind, kann die Änderung einzelner Neuronen wiederum *qua* dynamischer Verstärkungseffekte den Zustand ganzer Neuronengruppen maßgeblich verändern (vgl. Zoglauer, 1998, S.54ff.). Somit liegt eine enge Beziehung zwischen quantenmechanischen Effekten und Bewußtseinsphänomen nahe.

Auch John Eccles (in Popper & Eccles, 1997) weist Quanteneffekten eine wesentliche Rolle im Gehirn zu. Nach seiner Ansicht haben solche Effekte jedoch nicht vorrangig eine bewußtseins*erzeugende*, sondern eine bewußtseins*vermittelnde* Bedeutung: Indem sie die Wahrscheinlichkeit einer Exozytose⁷¹ erhöhen, könnten sie als Vermittlungsglied zwischen *Willens-* und *Handlungsakten* fungieren (siehe auch Zoglauer, 1998, S.55).

Grush & Churchland (1996) bewerten die Idee quantenmechanischer Beeinflussung oder Erzeugung von Bewußtsein als höchst unwahrscheinlich, da Quanteneffekte auf dem Makrolevel neuronaler Vorgänge in Rauschen untergehen. Außerdem können Grush & Churchland (1996) zeigen, daß Teile der Argumentation von Penrose höchst unwahrscheinlich und spekulativ sind.

Auch Calvin (1989b) steht der Hameroff–Penrose–Theorie als Alleinerklärung von Bewußtseinsphänomenen sehr skeptisch gegenüber. Für ihn sind Quanteneffekte höchstens notwendig aber nicht hinreichend um das Phänomen *Bewußtsein* zu erklären:

⁷⁰ diese Theorie wurde von Roger Penrose und Stuart Hameroff entwickelt (*Hameroff–Penrose–Hypothese*), wird jedoch hauptsächlich mit dem Namen Penrose in Verbindung gesetzt.

⁷¹ Exozytose ist eine Emission synaptischer Vesikel (nach Zoglauer, 1998, S.55).

“Quantum mechanics is probably essential to consciousness in about the same way as crystals were once essential to radios, or spark plugs are still essential to traffic jams. Necessary, but not sufficient. Interesting in its own right, but a subject related only distantly to out mental lives — and not to be confused with the temporary levels of organization associated with higher intellectual function” (Calvin, 1989b, S.5).

An anderer Stelle gibt Calvin zu bedenken, daß quantenmechanische Überlegungen zwar interessant sind, jedoch seien Quanteneffekte eine viel zu niedrige Erklärungsebene für das Verstehen des high-level-Phänomens *Bewußtsein*:

„Zwischen der Quantenmechanik und Bewußtsein liegen vielleicht ein Dutzend dieser dauerhaften Organisationsebenen: Beispiele dafür sind chemische Bindung, Moleküle und ihre Selbstorganisation, Molekularbiologie, Genetik, Biochemie, Membranen und ihre Ionenkanäle, Synapsen und ihre Neurotransmitter, das Neuron selbst [...] und so weiter“ (Calvin, 1998, S.59f.).

Außerdem vermutet Calvin (1998), daß der quantenmechanische Ansatz im Grunde nur deswegen so populär ist, da er dem Wunsch des Menschen nachkommt, einerseits mysteriös wirkende Phänomene wie das Bewußtsein auf eine wissenschaftliche Basis zu stellen und andererseits dennoch nicht in den sonst üblichen Determinismus physikalischer Theorien zu geraten.

6.3.6 Bewußtsein als Vetoinstanz

Eine ganz neue Sichtweise auf das Phänomen Bewußtsein ergaben die gehirnphysiologischen Experimente, die von Kornhuber und Libet in den Siebziger Jahren durchgeführt wurden (siehe Penrose, 1991; Popper & Eccles, 1997). Normalerweise postuliert man, daß der Willensakt stets einer bewußten Handlung vorausgeht. Benjamin Libet konnte jedoch demgegenüber demonstrieren, daß spezifische elektrische Aktivitäten im Gehirn, die normalerweise mit Bewußtseinsphänomenen in Verbindung gebracht werden, erst später erfolgen als neuronale Impulse zum Innervieren von motorischen Komponenten. Interessanterweise haben die an diesen Versuchen beteiligten Probanden jedoch das Gefühl, daß sie bereits Bewußtsein von der ausgeführten Handlung hatten, *bevor*⁷² sie diese tatsächlich einleiteten (siehe Saum-Aldehoff, 1993). Dies bedeutet aber, daß sich das Gehirn bereits „vorbewußt entschieden“ hat, was getan werden soll, obwohl es sich bei den verlangten motorischen Ausführungen gemäß der

⁷² Saum-Aldehoff (1993) spricht davon, daß sich Bereitschaftspotentiale etwa 500 ms vor der zu beobachteten Handlung aufbauen; jedoch ist erst 200 ms vor der Handlung Bewußtsein in Form von spezifischen Hirnaktivitäten registrierbar.

Bedeutung von Handlungen explizit *nicht* um reflexartige, nicht–bewußte Verhaltensabläufe handelte.

Saum–Aldehoff (1993) interpretiert die Möglichkeiten des Bewußtsein aufgrund der von Libet und Kornhuber erhaltenen Ergebnisse als sehr eingeschränkt, da vermutlich ein automatisch initiiertes Vorgang lediglich in Form eines Vetos beeinflusst werden kann. Was wir in unserem Bewußtsein als einheitliche, wohlgeordnete Welt vorfinden, ist das Endergebnis einer langen Wahrnehmungsarbeit des Gehirns. Allerdings wird uns von diesen aufwendigen Prozeduren nichts bewußt. Was uns bewußt wird, ist lediglich eine *Interpretation* der Wirklichkeit durch das Gehirn.

6.3.7 Funktionalismus

Auf die Theorie des Funktionalismus soll an dieser Stelle nicht mehr intensiv eingegangen werden, da sie bereits im Kapitel über klassische KI–Systeme, die den Funktionalismus als Ausgangspunkt nimmt, eingehend besprochen wurde (siehe Kapitel 3.2.2).

Dennoch sei erwähnt, daß die Eigenständigkeit des Mentalen, auch wenn sie generell materiell grundorientiert ist, vom Funktionalismus ausdrücklich betont wird⁷³: Allerdings werden mentale Phänomene nur mehr als abstrakte funktionale Zustände des gesamten Organismus verstanden (vgl. Wolters, 1993). Im Gegensatz zum Physikalismus wird daher keine *physische* Identität zwischen Gehirn und Bewußtsein angenommen. Vielmehr sind innerhalb der funktionalistischen Theorie mentale Eigenschaften gerade solche, die „in Begriffen, die den physikalischen oder chemischen Aufbau des Gehirns unerwähnt lassen, definiert werden können“ (Putnam zit. nach Oeser & Seitelberger, 1995, S.112).

6.3.8 Emergenztheorie

Mit dem Satz, „ein System mit vielen einfachen Komponenten wird manchmal ein Gesamtverhalten zeigen, das organisierter erscheint als das Verhalten der individuellen Teile“ (Hillis, 1996, S.169) beginnt Daniel W. Hillis seinen Artikel *Intelligenz als emergentes Verhalten*. Übertragen auf konnektionistische Systeme mit ihren primitiven basalen Strukturen und Prozessen, könnte das bedeuten, daß sich bei einer genügend großen Anzahl von Neuronen und der damit verbundenen Komplexität ihrer Verschaltung, ein qualitativ höherwertiges Gesamtsystem einstellen könnte. Dieses Phänomen wird als *Emergenz* bezeichnet (siehe Dewan, 1976; Sperry, 1976). In der wissenschaftlichen Diskussion sind emergentische Theorien äußerst umstritten. So be-

⁷³ als Hauptvertreter des Funktionalismus gelten Hilary Putnam und Jerry Fodor (siehe Wolters, 1993).

streiten Oeser und Seitelberger (1995) jede Möglichkeit einer Bewußtseinsentwicklung durch Emergenz:

„Warum sollte [...] eine ‚Ansammlung von Lichtschaltern‘ Bewußtsein entwickeln, nur weil sie klein und zahlreich, weil sie vom Menschen klug organisiert und programmiert sind?“ (Oeser & Seitelberger, 1995, S.200).

Auch Popper kann sich eine auf emergentische Phänomene stützende Theorie von Bewußtsein, Kreativität und Intelligenz nicht vorstellen:

„Selbst die, die glauben, daß das Bewußtsein ‚bloß‘ das ursächliche Ergebnis der sich selbst organisierenden Materie ist, müßten doch spüren, daß es schwierig ist, die Neunte Symphonie oder ‚Othello‘ oder die Gravitationstheorie so zu sehen“ (Popper & Eccles, 1997, S.256).

Konrad Lorenz favorisiert anstatt des Begriffs der *Emergenz* den der *Fulguration*, der die überraschende Neuartigkeit emergenter Eigenschaften und den diskontinuierlichen Charakter der Evolution charakterisieren soll. Der Begriff *Emergenz* erweckt nach Lorenz Ansicht jedoch den Eindruck, daß dabei

„etwas Präformiertes plötzlich auftauche, wie ein luftholender Wal an der Oberfläche des Meeres, das eben noch, bei buchstäblicher Betrachtung, leer zu sein schien“ (Lorenz, 1977, S.47).

Die berechtigte Frage ist nun, wie emergentische Phänomene plötzlich auftauchen können, ohne daß sie bereits davor in irgendeiner Weise präformiert waren. Dementsprechend nimmt auch die *Theorie der Potentialitäten* (nach Popper & Eccles, 1997, S.46) an, daß alles, was im Verlauf der Evolution des Universums neu entsteht, bereits zuvor in den beteiligten physikalischen Teilchen und Strukturen bestanden hat.

Auch wenn dieser Umstand ungeklärt bleibt, existiert jedoch ohne Zweifel der Vorteil einer Fulgurations- resp. Emergenztheorie dadurch, daß man intelligente Systeme konstruieren kann, ohne sie zunächst vollständig verstehen zu müssen. Diese Chance sieht auch Smolensky:

“In these models [connectionist models] intelligent behavior emerges from the accumulation of many simple processes occurring independently” (Mayer, 1992, S.201, Erg. v. Verf.).

In diesem Sinne schlägt Daniel W. Hillis vor, die Komplexität des Gehirns abzuschätzen und mit den daraus gewonnenen Größenverhältnissen ein hirnhähnliches konnektionistisches System zu entwickeln:

„Es wäre sehr angenehm, wenn Intelligenz ein emergentes Verhalten von zufällig verbundenen Neuronen wäre, so wie Schneeflocken und Strudel emergente Verhaltensweisen sind von Wassermolekülen. Es könnte möglich sein, eine Denkmaschine zu bauen, indem einfach ein genügend großes Netz von künstlichen Neuronen zusammengehakt wird. Der Begriff ‚Emergenz‘ würde besagen, daß ein solches Netz, wenn es einmal eine kritische Masse erreicht hat, spontan zu denken anfinge.“ (Hillis, 1996, S.169).

Auch Calvin (1998) geht von einer Emergenztheorie des Bewußtseins aus, nach der erst eine bestimmte Schwelle von Komplexität überwunden werden muß, damit höhere Intelligenz oder Bewußtsein entstehen kann. Seiner Meinung nach sei die nicht ausreichende Komplexität der meisten biologischen Systeme auch der Grund dafür, daß auf der Erde relativ wenig hochintelligente Arten existieren.

Um ein intelligentes und bewußtes künstliches Wesen schaffen zu können, wäre vorausgehend natürlich noch eine der menschlichen Entwicklung parallele Lernperiode notwendig, in der die angeborenen und erworbenen Fähigkeiten trainiert werden würden. Außerdem wäre eine stete Weiterversorgung durch „sensorische Bewegungsfunktionen“ (Hillis, 1996, S.178) nötig, um die Intelligenz dauerhaft aufrechtzuerhalten.

Zoglauer (1998) weist daraufhin, daß der Emergentismus keine wissenschaftliche Theorie in dem Sinne ist, daß sie *erklären* könnte, was Bewußtsein ist. Man fragt sich jedoch mit recht, welche bisher aufgestellte Theorie, eine *wirkliche* Erklärung geistiger Phänomene liefern kann.

6.4 Entscheidungskriterien für Bewußtsein

In den vorangegangenen Kapiteln wurden verschiedene Ansatzmöglichkeiten und Theorien vorgestellt, die zu erklären suchen, wie Bewußtsein entsteht, wodurch es bedingt wird und auf welche Weise es mit physikalischen Vorstellungen vereinbar ist. Nun soll nach Entscheidungskriterien gesucht werden, nach denen man erkennen kann, ob und inwieweit ein System über Bewußtsein verfügt. Sehr hilfreich ist in diesem Zusammenhang der Artikel von Van Gulick (1996), der sich eigens der Frage widmet „Was würde als eine Erklärung von Bewußtsein zählen?“.

6.4.1 Unklare Kriterien

Wer danach fragt, ob irgendeine Entität über Bewußtsein verfügt, verhält sich in der Regel so, als ob vollkommen definiert wäre, was wir genau unter Bewußtsein verstehen. Bei näherem Hinsehen und Nachdenken zeigt sich jedoch sehr schnell, daß wir über keine klaren Kriterien und Regeln verfügen, aufgrund derer wir von Bewußtsein reden (vgl. Tetens, 1994):

„Tatsächlich verwenden wir das alltagspsychologische Vokabular viel diffuser, als wir gemeinhin annehmen. Sinn und Bedeutung des alltagspsychologischen Vokabulars sind nicht durchsichtig und liegen nicht auf der Hand. Klar ist nur, daß wir das alltagspsychologische Vokabular in jedem Fall auf Menschen anwenden und daß dabei das von ‚außen‘ beobachtbare Verhalten und Handeln von Personen ein entscheidendes Kriterium ist“ (Tetens, 1993, S.188f.).

Auch Patricia Churchland (1996c) geht davon aus, daß es keine klaren Kriterien gibt, nach denen sicher von bewußten Zuständen gesprochen werden könnte. Zwischen unbewußten und bewußten Prozessen und Zuständen, so Churchland, könne weder durch Verhaltensbeobachtungen noch durch neuronale Analyse sicher getrennt werden; vielmehr handele es sich dabei um graduelle Unterschiede.

John Searle (1997) stellt zwar ebenfalls fest, daß man über keine sicheren Merkmale für Bewußtsein verfüge. Andererseits könne man, *negativ* ausgedrückt, sagen, daß es Merkmale gibt, die Bewußtsein *nicht garantieren*. Als nicht hinreichend zur Erzeugung von Bewußtsein erachtet Searle z.B. formale Symbolmanipulationen (siehe Searle, 1993; Searle, 1994; Searle, 1996; Searle, 1997).

6.4.2 Neurophysiologie

Wir wissen, daß beim Menschen ein funktionierendes Gehirn eine notwendige Bedingung von Bewußtseinszuständen und –akten ist:

„Empirische Untersuchungen an Patienten mit Hirnläsionen so wie Experimente mit der selektiven Stimulierung bestimmter Hirnpartien haben ergeben, daß Bewußtsein von dem komplexen Zusammenspiel einer Reihe von Hirnfunktionen abhängt, die ihren Sitz in der Großhirnrinde, teils im Stammhirn [...] haben“ (Birnbacher, 1996, S.723).

Eine gängige Hypothese geht davon aus, daß erst dann Bewußtseinsphänomene auftreten, wenn in bestimmten neuronalen Verbänden eine bestimmte Schwelle der Aktivierungsrate überschritten wird. Übertragen auf künstliche neuronale Netze würde dies bedeuten, daß man zuerst ein genügend komplexes Netzwerk benötigt, das eine spezifische Oszillation von „neuronaler“ Aktivität aufweist, damit ein künstliches Bewußtsein entstehen kann (siehe Phaf & Wolters, 1997).

6.4.3 Simulationen durch Computer und Roboter

Daniel Dennett (1996a) will sich dem Problem menschlicher Intelligenz und Bewußtseinsfähigkeit im Gegensatz zu den meisten anderen Wissenschaftlern mit Hilfe eines konkret konstruktivistischen Ansatzes nähern. Dazu entwickelte er innerhalb eines

MIT⁷⁴–Projektes den menschenähnlichen Roboter *COG*, dem nach und nach typische menschliche Eigenschaften und Fähigkeiten implementiert wurden. Der Sinn des Unternehmens ist, mehr darüber zu erfahren, welches die wirklich schwierigen Probleme sind, die sich der Evolution und einzelnen lebenden Wesen stellen. Dennett (1996a) vermutet, daß durch dieses Vorgehen auch mehr Klarheit über den Sinn und Nutzen von Bewußtsein und dessen Entwicklungsbedingungen gewonnen werden kann. Die Konstrukteure von *COG* erachten es allerdings für unwahrscheinlich, daß Roboter jemals in derselben Art und Weise bewußt sind wie menschliche Wesen; jedoch müßte einem System, das sich autark verhalten kann und das über seine eigenen internen Zustände verlässlicher berichten könnte als der menschliche Programmierer selbst, durchaus eine bewußtseinsähnliche Eigenschaft zugesprochen werden.

Der weltbekannte KI–Forscher Marvin Minsky schließt sich zwar generell den Schlußfolgerungen Dennetts an, betont jedoch, daß es einen unnötigen Umweg bedeutet, einen hochkomplexen physisch–realen Roboter zu bauen: Da nämlich alle tiefergründigen Prinzipien der Organisation, die Kognitionen ermöglichen, leicht im abstrakteren Bereich der reinen Computersimulationen entdeckt werden können, sollte man sich auch auf diese weniger aufwendigere Domäne beschränken (siehe Dennett, 1996a). Sollte eine KI–Applikation sich ähnlich wie ein menschliches Referenzsystem verhalten, so müßte man dieser nach Minskys Auffassung Bewußtsein zusprechen (vgl. Kapitel 6.4.4).

Gegenüber Minskys minimalistischer Sichtweise vertreten Dreyfus & Dreyfus (1996) die Vorgehensweise von Dennett und Kollegen: Auch sie meinen, daß eine echte Verkörperung in einer realen Umgebung entscheidend für Bewußtsein ist, da man bei einer reinen Computersimulation die tieflegendsten Probleme des Designs übersehen, unterschätzen oder mißverstehen könnte (siehe auch Dennett, 1996a).

Searle (1993) schließlich hält zwar Computer und Robotersimulationen für prinzipiell hilfreich, jedoch könnten sie nicht hinreichend erklären, was Bewußtsein sei.

6.4.4 Turing–Test

Alan Turing schlug als Entscheidungsgrundlage den nach ihm benannten Turing–Test vor, der schon weiter oben kurz Erwähnung fand (siehe Kapitel 3.4). Der Vorteil dieses Tests ist, daß er sich erstens auf empirisch nachprüfbar Daten, nämlich den Dialogen der untersuchten Systeme (Mensch oder Maschine) stützt, und zweitens, daß er den Menschen als Referenz für Intelligenz und Bewußtsein benützt, welcher *per definitionem* über diese Eigenschaften verfügt (vgl. Penrose, 1991; Wiener, 1990). Dies

⁷⁴ Massachusetts Institute of Technology.

hat freilich zur Folge, daß die Untersuchung auf prototypisch menschliche Eigenschaften eingeschränkt wird⁷⁵. Wie ein Beispiel von Churchland (1997, S.273) deutlich demonstriert, können dadurch selbst Menschen, sofern sie vom alltäglichen Intelligenzverständnis abweichen, fälschlicherweise als *unintelligente* Systeme eingeschätzt werden.

Zweifelhaft bleibt auch das dem Turing-Test eigentümliche Vorgehen, alleine die *Phänomenebene* zu betrachten, und diese als Beweis für intelligentes Verhalten anzunehmen. Hochintelligente Systeme, deren Kommunikation wir nicht entschlüsseln könnten, würden deshalb auch zwangsläufig aus dem Intelligenz-Raster von Turing fallen. Für sprachbehinderte Menschen oder auch Kleinkinder trifft dies ebenfalls zu.

Als eifrigster Gegner des Turing'schen Ansatzes darf wohl John Searle gelten, der mit seinem Beispiel des chinesischen Zimmers das wohl am häufigsten zitierte Argument gegen die funktionalistische Ansicht Turings hervorbrachte:

“I do not understand Chinese, so I imagine that I am locked in a room with a lot of boxes of Chinese symbols (the database), I get small bunches of Chinese symbols passed to me (questions in Chinese), and I look up in a rule book (the program) what I am supposed to do. I perform certain operations on the symbols in accordance with the rules (that is, I carry out the steps in the program) and give back small bunches of symbols (answers to the questions) to those outside the room. I am the computer implementing a program for answering questions in Chinese, but all the same I do not understand a word of Chinese” (Searle, 1997, S.11, Klammern i. Original).

Searle kann anhand dieses Beispiels zeigen, daß die alleinige stupide Anwendung von Algorithmen nicht zu einem echten Wissen und auch keinem Bewußtsein führt⁷⁶: „Semantik ist der Syntax nicht intrinsisch“ (Searle, 1993, S.220), d.h. Syntax ist weder konstitutiv noch hinreichend für Semantik.

Summa summarum sieht Searle den Turing-Ansatz zwar als eine sehr gut operationalisierte Möglichkeit an, um maschinelle und menschliche Verhaltensweisen wissenschaftlich vergleichen zu können, jedoch kann der Turing-Test nicht sicher entscheiden, ob ein System wirklich ein Bewußtsein besitzt:

⁷⁵ Warwick (1998) stellt die Frage, wie sich denn eigentlich ein *normaler Mensch* definiert. Da man diese Frage nicht hinreichend beantworten kann, ist es auch nicht möglich, einen Computer mit einem „Prototyp von Mensch“ zu vergleichen: “Would our artificial being, like most fictional creatures of this type, be perfectly normal but perhaps rather simple? Well, why should it. Apart from us humans wanting it to be so, because then the being is not much of a problem to us. Why couldn't it instead be a psychotic killer?” (Warwick, 1998, S.114).

⁷⁶ auch Dewan (1976) meint, daß das stupide Ausführen von Anweisungen für sich gesehen allerhöchstens eine Simulation für Bewußtsein ist.

„Ontologisch gesehen sind Verhalten, funktionale Rolle und Kausalbeziehungen irrelevant für die Existenz bewußter geistiger Phänomene“ (Searle, 1996, S.86).

Penrose (1991) geht davon aus, daß sich generell jede noch so raffinierte Imitation durch genügend geschicktes Prüfen als solche enttarnen läßt⁷⁷. Falls ein künstliches System jedoch tatsächlich in Turing'scher Hinsicht ununterscheidbar von einem Menschen wäre, so würde Penrose „wenn kein Indiz dagegen spricht, annehmen, daß der Computer tatsächlich denkt, fühlt und so weiter“ (Penrose, 1991, Herv. i. Original).

6.4.5 Gibt es überhaupt eine Entscheidungsgrundlage?

Bis heute verfügen wir nicht über eine umfassende Theorie, die hinreichend erklären könnte, was Intelligenz und Bewußtsein sind und was in uns *Geist* auslöst (vgl. Sokolowski, 1996). Deshalb stellt auch das meist verwandte und zitierte Entscheidungsinstrument, der Turing-Test, eine kaum akzeptable Notlösung dar: Dieser Test operiert erstens nur auf der Verhaltensebene und zweitens greift er jeweils nur einen winzigen Ausschnitt aus dem kognitiven Repertoire des Menschen heraus⁷⁸.

Das Ziel muß jedoch sein, den *ganzen* Bereich intelligenten Verhaltens und die mit diesem verbundenen *internen Prozesse* zu berücksichtigen. Dazu müssen allerdings die Grundeigenschaften von Kognitionen weitreichend erklärt werden können. Erst dann könnte man den Schritt von der Beobachtung der *Wirkungen* geistigen Handelns hin zu den *Ursachen* dieser Prozesse vollziehen. Und nur dann haben wir die Chance, solche Entscheidungen mit gutem Gewissen treffen zu können.

⁷⁷ Stanislaw Lem (1996) widerspricht dieser Ansicht: „Wir wissen also, daß der Turing-Test von einem Computer mit entsprechender Rechenkapazität im Prinzip ‚geknackt‘ und der menschliche Gesprächspartner getäuscht werden kann“ (Lem, 1996, S.34).

⁷⁸ Die Kommunikationsmöglichkeiten der Testteilnehmer sind äußerst eingeschränkt, da sie sich über ein schriftliches Medium verständigen müssen, und das Gespräch nur über ein genau vorgegebenes, engumgrenztes Fachgebiet führen dürfen.

7 Conclusio

Seit es Computer gibt, ist die Nachbildung bewußter und intelligenter Fähigkeiten ein heiß diskutiertes Thema (vgl. Hoffmann, 1992). Gerade das Ausbilden eines Bewußtseins wird neben der Sprachfähigkeit oft als eine exklusiv dem Menschen vorbehalten Leistung angesehen, wobei diese auf Descartes zurückgehende These in der Nachfolge immer wieder stark kritisiert wurde.

Die bisherigen Anstrengungen der klassischen KI sind zwar in der Simulation von Teilbereichen menschlicher Intelligenz erfolgreich einsetzbar, doch scheint eine umfassende Simulation oder gar ein Verständnis des Geistes mit den der KI eigentümlichen Methoden nicht möglich zu sein (siehe Warwick, 1998).

Der konnektionistische Ansatz weckt realistischere Hoffnungen, da er bereits systemimmanent über wichtige Eigenschaften eines intelligenten Systems verfügt: Zu erwähnen sind Generalisierungsfähigkeit, Fehlertoleranz, Autoassoziation, Lernfähigkeit, Erkennen von Kausalvorgängen und eine Art „zeitliches Gefühl“ (siehe Churchland, 1997, S. 122ff. u. S.249f.).

Wenn man sich das erste Mal intensiv mit künstlichen neuronalen Netzen auseinandersetzt, so kann einen die Tatsache verwirren, daß solche Systeme intern nicht mehr mit Symbolen operieren. Prozesse mit subsymbolischer Manipulation kennen wir weder von traditionellen Computersystemen, noch von unserer eigenen Praxis als denkende Wesen. Bei näherem Hinsehen zeigt sich jedoch, daß genau hierin der entscheidende Vorteil gegenüber der klassischen KI liegt. Die normale Arbeitspraxis der KI besteht darin, Denkprozesse, so wie sie den Menschen in introspektiver Weise bewußt werden, auf Computern in Form von Produktionssystemen abzubilden. Tatsächlich wird dadurch aber nicht die Intelligenz oder die Fähigkeit dazu abgebildet, sondern nur die *Wirkung* derselben: Die klassische KI wird somit zum bloßen Abziehbild menschlichen Verhaltens, und nicht zu einem Adäquatum menschlicher Intelligenz!

Der Konnektionismus geht einen ganz anderen, basaleren und wirklichkeitsnäheren Weg. Er lehnt sich vom schematischen Aufbau her an natürliche Neuronennetze an

und übernimmt grundsätzliche Prozeßstrukturen von diesem Vorbild⁷⁹. Repräsentationen erfolgen auf subsymbolische Weise über das gesamte Netz verteilt. Dies führt dazu, daß sich selbsterhaltende, dynamische Systeme entwickeln können.

Noch ein weiterer wichtiger Vorteil zeigt sich durch die Verwendung konnektionistischer Systeme: Dadurch, daß „die Existenz autokonnektiver Wahrnehmungsbahnen, ihr Ursprung und ihre Funktionen ausnahmslos mit rein naturwissenschaftlichen Annahmen erklärbar sind“ (Churchland, 1997, S.235), kann man auf metaphysische Zusatzannahmen zur Erklärung des Phänomens *Geist* und *Bewußtsein* verzichten. Die klassische Descartes–Leibniz’sche Aufteilung in Geist und Physis entfällt — übrig bleiben alleine *physikalische* Strukturen.

Zwar werden neuronale Netze, bedingt durch Umbilden der Wissenschaftsgeschichte und unseres Kulturverständnisses, erst seit ca. zehn Jahren intensiv erforscht, und der Wissensstand über sie ist dementsprechend gering, jedoch scheint ihr Potential von großer Sprengkraft zu sein. Gerade solche Problembereiche, die man bisher mühsam und oft ergebnislos zu explizieren suchte, stellen lohnende Forschungsschwerpunkte für neuronale Netze dar. Zu diesen Themen gehört vornehmlich auch die Erforschung unseres Bewußtseins, welche in der jahrhundertealten Symbolverarbeitungs–Paradigma–Sackgasse steckengeblieben zu sein scheint. Hier kann die auf neuronale Netze angewendete Emergenztheorie wichtige Impulse liefern, da man auf ihrer Grundlage die Möglichkeit hat, „das emergente Substrat der Intelligenz nachzubilden, ohne alle Details verstanden zu haben“ (Hillis, 1996, S.179).

Durch die Möglichkeit, funktionsfähige intelligente Systeme zu entwickeln, ohne daß man das Gesamtsystem in seiner Gänze begreift, läßt sich schließlich auch Heideggers Ansicht über intelligente Wesen und deren unvollständiges Wissen über die Welt besser verstehen:

„Möglicherweise können neuronale Netze zeigen, daß Heidegger, der späte Wittgenstein und Rosenblatt richtig lagen, als sie zu der Annahme kamen, daß wir uns in der Welt intelligent benehmen, ohne eine Theorie dieser Welt zu haben“ (Dreyfus & Dreyfus, 1996, S.39).

Unübersehbar sind die Analogien simulierter neuronaler Netze zu biologischen Systemen: Inhärente Fehlertoleranz, Fähigkeit zur Selbstorganisation, Lernfähigkeit und massive Parallelverarbeitung sind Merkmale, die die bisher unerreichte Leistungsfähigkeit des Gehirns begründen. Obgleich wir von einer Nachbildung des Gehirns noch

⁷⁹ Dazu scheint es natürlich sinnvoll, daß man erst einmal über ein ausreichender Wissen über die komplexen Funktionen des Gehirns verfügt, sprich: Mit Hilfe des Gehirns muß das Gehirn hinreichend ent-

weit entfernt sind, scheinen die Ergebnisse der Neuroinformatik zu signalisieren, daß der Ansatz prinzipiell richtig ist, auch wenn die Modelle noch unvollständig sind:

„Trotz mancher Lücken im Verständnis bewähren sich Künstliche Neuronale Netze in der praktischen Anwendung mehr und mehr. Obwohl sie erst seit wenigen Jahren systematisch erforscht werden, sind ihre Bandbreite, ihr Einsatzspektrum und ihre Lösungsansätze vielfältig und teilweise verblüffend neuartig. Aufgabenstellungen, die bis heute mit herkömmlichen mathematischen oder algorithmischen Verfahren nicht oder nur sehr aufwendig zu bearbeiten waren, können durch die Neuroinformatik oft wesentlich eleganter und schneller gelöst werden.“ (Sperlich & Möcke, 1993).

Zusätzlich sei hervorgehoben, daß die allgemeine Konnektionismus-Forschungspraxis durch ihre syntopische Vorgehensweise eine Leitbildfunktion für andere Forschungsbereiche ausüben könnte: Die Koordination von vielen verschiedene Denkansätzen, Forschungsmethoden und Zielrichtungen, kann helfen, mehr Licht in das oftmals als *Mysterium* beschriebene Dunkel der Bewußtseinerforschung zu bringen (siehe Phaf & Wolters, 1997).

Als besonders vielversprechend erachten namhafte Forscher die integrierte Anwendung von konnektionistischen Systemen und klassischen symbolorientierten KI-Ansätzen (siehe Goschke & Koppelberg, 1993; Sperlich & Möcke, 1993). Auch das sonst sehr auf neuronale Netze fixierte Wissenschaftlerteam Churchland & Churchland gibt zu bedenken, daß rein parallel verarbeitende Systeme nicht immer ideal funktionieren, weshalb eine Verbindung mit klassischen KI-Systemen durchaus sinnvoll sei (siehe Churchland & Churchland, 1990).

Trotz der erfolgversprechenden Aussichten von konnektionistischen Systemen sollen die mit ihrem Einsatz verbundenen Probleme und Gefahren nicht unerwähnt bleiben. Da neuronale Netze für nicht exakt explizierbare Problemstellungen prädestiniert sind, werden sie in Zukunft auch verstärkt in *sensiblen* Bereichen Anwendung finden, d.h. in solchen Bereichen, die das Wohl und das Leben von Mensch, Tier und Welt nachhaltig beeinflussen können. In diesem Zusammenhang wurde bereits in Kapitel 5.2.8 ein neuronales System zur Kreditzuschreibung kritisch gewürdigt. Weitere sensible Bereiche, in denen ebenfalls neuronale Netze eingesetzt werden könnten, sind die Vergabe von Studienplätzen, die Bewertung von Arbeitsstellenbewerbern, Flugsicherung, u.v.m. Vorstellbar wäre aber auch deren Einsatz bei militär-strategischen Fragen

deckt werden! (siehe Warwick, 1998)

oder zumindest als Addendum zur Regierungsgewalt (siehe Hoffmann, 1992). Es gilt zu bedenken, inwieweit man sich als Mensch die Chance nehmen lassen will, in diesen Bereichen selbst als Experte aufzutreten. Auch Warwick (1998) warnt davor, irgendwelchen Maschinen, denen wir Intelligenz zusprechen, zuviele Entscheidungsbe-fugnisse zu gewähren. Um diesen Standpunkt zu untermauern, nennt er das Beispiel der atomaren Kriegsführung zum Ende des Zweiten Weltkriegs. Seines Erachtens hätten Menschen seitdem viel dazugelernt und eingesehen, daß ein atomarer Waffen-einsatz immer auch das Ende der *ganzen* Welt bedeuten könne. Warwick (1998) meint, daß intelligente Maschinen jedoch möglicherweise, sobald sie erst einmal akti-viert worden sind, viel konsequenter und damit zerstörerischer vorgehen würden. Die-se Argumentation erscheint jedoch in sich widersprüchlich! Denn warum sollte ma-schinelle Intelligenz nicht auch soziale, ethische und zukunftsorientierte Intelligenz mit einschließen? Oder anders gefragt: Könnte man sich nicht Intelligenz von Maschi-nen als *wirklich intelligent* im wahrsten Sinne des Wortes vorstellen, d.h. in zielfüh-render, zukunftsorientierter und moralisch überlegener Weise? Schließlich ist unsere menschliche Intelligenz aufgrund einer relativ langsam verlaufenden Evolution in ab-sehbarer Zeit kaum mehr signifikant zu steigern. Mit dem rasanten technologischen Fortschritt könnte jedoch auch ein starker Anstieg von kognitiven Leistungen künstli-cher Systeme einhergehen: “it is only a matter of time before there is nothing humans can do to stop machines becoming more intelligent than humans” (Warwick, 1998, S.185). Diese Perspektive kann, aber muß nicht unbedingt erschrecken, da sie letzt-endlich auch dem Menschen, der heute unter starkem evolutionärem Druck steht, zu-gute kommen kann, indem er diese Techniken sinnvoll und nutzbringend anwendet (vgl. Steels, 1996).

Die zahlreichen, sich teilweise wechselseitig widersprechenden Theorien und Überle-gungen, die in der vorliegenden Arbeit vorgestellt wurden, zeigen deutlich, daß man sich noch immer recht unsicher ist, ob konnektionistische Systeme bzw. künstliche Systeme allgemein eine Art von *Bewußtsein* entwickeln können. Es scheint so, als würde sich dieser Bereich tatsächlich einer wissenschaftlich fundierten Überprüfung entziehen, da es sich letztendlich um ein genuin philosophisches Thema handelt (vgl. Hoffmann, 1992). In diesem Sinne resümiert auch Holm Tetens:

„Wir sind es, die den Maschinen mentale Zustände zuschreiben oder auch nicht“ (Tetens, 1993, S.187).

Die Frage ist nur, wann wir beginnen, diese Zuschreibung mit gutem Gewissen und voller Überzeugung durchzuführen.

8 Literaturverzeichnis

- Anonymus. (1953). *Der Große Brockhaus* (16., völlig neubearbeitete Aufl.) (Vol. 2). Wiesbaden: F.A.Brockhaus.
- Anonymus. (1999a). Consciousness. In E. Britannica (Hrsg.), *Encyclopaedia Britannica CD 99 Multimedia Edition*. Chicago: Encyclopaedia Britannica.
- Anonymus. (1999b). Reductionism. In E. Britannica (Hrsg.), *Encyclopaedia Britannica CD 99 Multimedia Edition*. Chicago: Encyclopaedia Britannica.
- Anonymus. (1999c). Thorndike, Edward Lee. In E. Britannica (Hrsg.), *Encyclopaedia Britannica CD 99 Multimedia Edition*. Chicago: Encyclopaedia Britannica.
- Baddeley, A. (1990). *Human Memory. Theory and Practice*. London: Lawrence Erlbaum.
- Barr, T. (1991). Netze im Aufwind. Künstliche neuronale Netze - Stand der Forschung und praktischer Einsatz. *c't*(4), 78ff.
- Bechtel, W. (1993a). Das Ende der Verbindung zwischen dem mentalen Bereich und der Sprache. Eine konnektionistische Perspektive. In A. Elepfandt & G. Wolters (Hrsg.), *Denkmaschinen? Interdisziplinäre Perspektiven zum Thema Geist und Gehirn*. Konstanz: Universitätsverlag Konstanz.
- Bechtel, W. (1993b). Multiple Ebenen der Analyse in der Kognitionswissenschaft. In H. Hildebrandt & E. Scheerer (Hrsg.), *Interdisziplinäre Perspektiven der Kognitionsforschung*. Frankfurt am Main: Peter Lang.
- Beckermann, A. (1994). Der Computer - ein Modell des Geistes? In S. Krämer (Hrsg.), *Geist - Gehirn - künstliche Intelligenz* (S. 71-87). Berlin: Walter de Gruyter.
- Beneke, T. & Schwippert, W. (1995). Erkennungsdienst. Trends in der Bildverarbeitung. *c't*(3), 109ff.
- Bertuch, M. (1992). Magische Augen. Bildererkennung mit neuronalen Netzen ist anwendungsreif. *c't*(4), 34ff.

- Bharati, S. P. (1996). Körper, Geist und Bewußtsein im indischen Veda. In C. Maar, E. Pöppel & T. Christaller (Hrsg.), *Die Technik auf dem Weg zur Seele. Forschungen an der Schnittstelle Gehirn/Computer*. Reinbek: Rowohlt.
- Bieri, P. (1996). Was macht Bewußtsein zu einem Rätsel? In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3.,ergänzte Aufl.). Paderborn: Schöningh.
- Birbaumer, N. & Schmidt, R. F. (1991). *Biologische Psychologie* (2. Aufl.). Berlin: Springer.
- Birkelbach, J. (1995). Kurswechsel. Mit dem PC an die Börsen der Welt. *c't*(4), 232ff.
- Birnbacher, D. (1996). Künstliches Bewußtsein. In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3.,ergänzte Aufl.). Paderborn: Schöningh.
- Bortz, J. (1989). *Statistik für Sozialwissenschaftler* (3.,neu bearbeitete Aufl.). Berlin: Springer.
- Brunak, S. & Lautrup, B. (1993). *Neuronale Netze: Die nächste Computer-Revolution* (Heim, Malte; Übers.). München: Hanser.
- Calvin, W. H. (1989a). *The Cerebral Symphony*. o.O.: Bantam.
- Calvin, W. H. (1989b). Competing for Consciousness: A Darwinian Mechanism at an Appropriate Level of Explanation. *Journal of Consciousness Studies*, 5(4), 389-404.
- Calvin, W. H. (1994). The Emergence of Intelligence. *Scientific American*, 271(4), 100-107.
- Calvin, W. H. (1998). *Wie das Gehirn denkt. Die Evolution der Intelligenz* (Niehaus-Osterloh, Monika; Übers.). Heidelberg: Spektrum.
- Calvin, W. H. & Ojemann, G. A. (1994). *Conversations with Neil's Brain*. Reading/MA: Addison-Wesley.
- Christaller, T. (1996). Kognitive Robotik. In C. Maar, E. Pöppel & T. Christaller (Hrsg.), *Die Technik auf dem Weg zur Seele. Forschungen an der Schnittstelle Gehirn/Computer*. Reinbek: Rowohlt.
- Churchland, P. (1996a). Eliminative Materialism and the Propositional Attitudes. In W. Lyons (Hrsg.), *Modern Philosophy of Mind*. London: Everyman.
- Churchland, P. M. (1997). *Die Seelenmaschine* (Numberger, Markus; Übers.). Heidelberg: Spektrum.
- Churchland, P. M. & Churchland, P. S. (1990). Ist eine denkende Maschine möglich? *Spektrum der Wissenschaft*, 3, 47-54.

- Churchland, P. S. (1996b). Die Neurobiologie des Bewußtseins. Was können wir von ihr lernen? In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3.,ergänzte Aufl.). Paderborn: Schöningh.
- Churchland, P. S. (1996c). Vernunft braucht Gefühle. In C. Maar, E. Pöppel & T. Christaller (Hrsg.), *Die Technik auf dem Weg zur Seele. Forschungen an der Schnittstelle Gehirn/Computer*. Reinbek: Rowohlt.
- Cowan, J. D. & Sharp, D. H. (1996). Neuronennetze und Künstliche Intelligenz. In S. R. Graubard (Hrsg.), *Probleme der Künstlichen Intelligenz. Eine Grundlagendiskussion* (S. 85-120). Wien: Springer.
- Crick, F. (1997). *Was die Seele wirklich ist. Die naturwissenschaftliche Erforschung des Bewußtseins* (Gavagi, Harvey P.; Übers.). Reinbek: Rowohlt.
- Dennett, D. (1994). *Philosophie des menschlichen Bewußtseins*. Hamburg: Hofmann und Campe.
- Dennett, D. (1996a). COG: Schritte in Richtung auf Bewußtsein in Robotern. In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3.,ergänzte Aufl.). Paderborn: Schöningh.
- Dennett, D. C. (1996b). Bewußtsein hat mehr mit Ruhm als mit Fernsehen zu tun. In C. Maar, E. Pöppel & T. Christaller (Hrsg.), *Die Technik auf dem Weg zur Seele. Forschungen an der Schnittstelle Gehirn/Computer*. Reinbek: Rowohlt.
- Dennett, D. C. (1996c). Wenn Philosophen Künstlicher Intelligenz begegnen. In S. R. Graubard (Hrsg.), *Probleme der Künstlichen Intelligenz. Eine Grundlagendiskussion* (S. 269-282). Wien: Springer.
- Dewan, E. M. (1976). Consciousness as an Emergent Causal Agent in the Context of the Control System Theory. In G. G. Globus, G. Maxwell & I. Savodnik (Hrsg.), *Consciousness and the Brain. A Scientific and Philosophical Inquiry*. New York: Plenum.
- Dewan, E. M., Eccles, J. C., Globus, G. G., Gunderson, K., Knapp, P. H., Maxwell, G., Pribram, K. H., Savage, C. W., Savodnik, I., Scriven, M., Sperry, R. W., Weimer, W. B. & Wimsatt, W. C. (1976). The Role of Scientific Results in the Theories of Mind and Brain: A Conversation among Philosophers and Scientists. In G. G. Globus, G. Maxwell & I. Savodnik (Hrsg.), *Consciousness and the Brain. A Scientific and Philosophical Inquiry*. New York: Plenum.
- Dörner, D. (1994). Über die Mechanisierbarkeit der Gefühle. In S. Krämer (Hrsg.), *Geist - Gehirn - künstliche Intelligenz* (S. 131-164). Berlin: Walter de Gruyter.
- Dorsch, F. (Hrsg.). (1987). *Psychologisches Wörterbuch* (11.,erg. Aufl.). Bern: Hans Huber.
- Dreyfus, H. L. & Dreyfus, S. E. (1996). Einen Geist bauen gegen ein Gehirn modellieren. In S. R. Graubard (Hrsg.), *Probleme der Künstlichen Intelligenz. Eine Grundlagendiskussion* (S. 15-44). Wien: Springer.

- Ebeling, A. (1995). Licht ins Dunkel. Neurotechnologie entwickelt Prothesen fürs Gehirn. *c't*(8), 96ff.
- Eckmiller, R. (1994). Neuroinformatik. In S. Krämer (Hrsg.), *Geist - Gehirn - künstliche Intelligenz* (S. 223-233). Berlin: Walter de Gruyter.
- Eysenck, M. W. & Keane, M. T. (1995). *Cognitive Psychology. A Student's Handbook*. Hove/East Sussex: Psychology Press.
- Flohr, H. (1993). Hirnprozesse und phänomenales Bewußtsein. In H. Hildebrandt & E. Scheerer (Hrsg.), *Interdisziplinäre Perspektiven der Kognitionsforschung*. Frankfurt am Main: Peter Lang.
- Floyd, C. (1994). Künstliche Intelligenz - Verantwortungsvolles Handeln. In S. Krämer (Hrsg.), *Geist - Gehirn - künstliche Intelligenz* (S. 256-278). Berlin: Walter de Gruyter.
- Foerst, A. (1995). Maschine Mensch. Das Menschenbild der künstlichen Intelligenz. *c't*(1), 92ff.
- Føllesdal, D., Walløe, L. & Elster, J. (1988). *Rationale Argumentation*. Berlin: Walter de Gruyter.
- Fremerey, F. (1999). Supercomputer für jedermann. Rekonfigurierbare Hardware statt herkömmlicher Prozessoren. *c't*(15), 20.
- Frey, H. G. & Schach, J. (1996). Modern geknotet. Neuro-Netze: Strukturen, Typen, Anwendungen. *c't*(2), 256ff.
- Geach, P. T. (1976). *Reason and Argument*. Guilford: Billing.
- Görz, G. (Hrsg.). (1995). *Einführung in die künstliche Intelligenz* (2. Aufl.). Bonn: Addison-Wesley.
- Goschke, T. & Koppelberg, D. (1993). Konnektionistische Repräsentation, semantische Kompositionalität und die Kontextabhängigkeit von Konzepten. In H. Hildebrandt & E. Scheerer (Hrsg.), *Interdisziplinäre Perspektiven der Kognitionsforschung*. Frankfurt am Main: Peter Lang.
- Graubard, S. R. (Hrsg.). (1996). *Probleme der Künstlichen Intelligenz. Eine Grundlegendiskussion* (Vol. IX). Wien: Springer.
- Grush, R. & Churchland, P. S. (1996). Lücken im Penrose-Parkett. In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3., ergänzte Aufl.). Paderborn: Schöningh.
- Hamilton, P. (1996). Kunst-Augen. Neuroprothetik: Hoffnung für Blinde. *c't*(9), 116ff.
- Hamilton, P. (1997). Neues Sehen. Forscher entwickeln elektronisches Auge mit biologienahen Neuronetzen. *c't*(10), 138ff.

- Hastedt, H. (1993). Anstatt eines Vorwortes. Bemerkungen zur Interdisziplinarität. In P. Scheffe, H. Hastedt, Y. Dittrich & G. Keil (Hrsg.), *Informatik und Philosophie*. Mannheim: Bibliographisches Institut.
- Hatfield, G. (1993). Repräsentation und kognitive Regeln in konnektionistischen Systemen. In H. Hildebrandt & E. Scheerer (Hrsg.), *Interdisziplinäre Perspektiven der Kognitionsforschung*. Frankfurt am Main: Peter Lang.
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
- Heckhausen, H. (1989). *Motivation und Handeln* (2. Aufl.). Berlin: Springer.
- Hillis, W. D. (1996). Intelligenz als emergentes Verhalten oder: Songs aus dem Paradies. In S. R. Graubard (Hrsg.), *Probleme der Künstlichen Intelligenz. Eine Grundlegendiskussion* (S. 183-202). Wien: Springer.
- Hinton, G. E. (1992). Wie neuronale Netze aus der Erfahrung lernen. *Spektrum der Wissenschaft*(11), 134-143.
- Hoffmann, N. (1992). Neuronale Netze. *Gestalt Theory*, 14(1), 43-59.
- Hofstadter, D. R. (1979). *Gödel, Escher, Bach: ein Endloses Geflochtenes Band* (Wolff-Windegg, P., Feuersee, H.; Übers.). Stuttgart: Klett-Cotta.
- Husserl, E. (1965). *Philosophie als strenge Wissenschaft*. Frankfurt am Main: Klostermann.
- Kambartel, F. (1993). Kann es gehirnphysiologische Ursachen unseres Handelns geben? In A. Elefant & G. Wolters (Hrsg.), *Denkmaschinen? Interdisziplinäre Perspektiven zum Thema Geist und Gehirn*. Konstanz: Universitätsverlag Konstanz.
- Keil, G. (1993). Die zwei Teilthesen des Funktionalismus. In P. Scheffe, H. Hastedt, Y. Dittrich & G. Keil (Hrsg.), *Informatik und Philosophie*. Mannheim: Bibliographisches Institut.
- Kirchner, W. (1990). Fehlerkorrektur im Rückwärtsgang. Neuronales Backpropagation-Netz zum Selbermachen. *c't*(11), 248ff.
- Kirk, R. (1996). Wie ist Bewußtsein möglich? In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3., ergänzte Aufl.). Paderborn: Schöningh.
- Kluge, F. (1989). *Etymologisches Wörterbuch der deutschen Sprache* (22. Aufl.). Berlin: Walter de Gruyter.
- Kosslyn, S. M. & Pomerantz, J. R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 9, 52-76.
- Kotulak, R. (1996). *Inside the Brain. Revolutionary Discoveries of how the Mind works*. Kansas City: Andrews McMeel.

- Krämer, S. (1994). Geist ohne Bewußtsein? Über einen Wandel in den Theorien vom Geist. In S. Krämer (Hrsg.), *Geist - Gehirn - künstliche Intelligenz* (S. 88-112). Berlin: Walter de Gruyter.
- Kühn, M. (1991). Gewagte Geschäfte. Neuronale Netze an der Börse. *c't*(12), 62ff.
- La Mettrie, J. O. d. (1875). *Der Mensch eine Maschine* (Ritter, Adolf; Übers.). Berlin: Erich Koschny.
- Leibniz, G. W. (1879). *Die Theodicee* (Von Kirchmann, J. H.; Übers.). Leipzig: Dürr.
- Leibniz, G. W. (1904). *Neue Abhandlungen über den menschlichen Verstand* (Schaarschmidt, Carl; Übers.) (2. Aufl.). Berlin: Dürr.
- Lem, S. (1996). Unsaubere Schnittstelle Mensch/Maschine. In C. Maar, E. Pöppel & T. Christaller (Hrsg.), *Die Technik auf dem Weg zur Seele. Forschungen an der Schnittstelle Gehirn/Computer*. Reinbek: Rowohlt.
- Lorenz, K. (1973). Bewußtsein. In H. Krings, H. M. Baumgartner & C. Wild (Hrsg.), *Handbuch philosophischer Grundbegriffe* (Bd. 1-6). München: Kösel.
- Lorenz, K. (1977). *Die Rückseite des Spiegels*. München: dtv.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford: Oxford University Press.
- Lyons, W. (1996). *Modern Philosophy of Mind*. London: Everyman.
- Mallot, H. A. (1995). Neuronale Netze. In G. Görz (Hrsg.), *Einführung in die künstliche Intelligenz* (2. Aufl.). Bonn: Addison-Wesley.
- Malsburg, C. v. d. (1996). Die Barriere zwischen Gehirn und Computer. In C. Maar, E. Pöppel & T. Christaller (Hrsg.), *Die Technik auf dem Weg zur Seele. Forschungen an der Schnittstelle Gehirn/Computer*. Reinbek: Rowohlt.
- Manhart, K. (1991). Menschliche Maschinen. Die lange Geschichte der KI. *c't*(3), 32ff.
- Matlin, M. W. (1989). *Cognition* (2. Aufl.). Fort Worth: Holt, Rinehart and Winston.
- Mayer, R. E. (1992). *Thinking, problem solving, cognition* (2. Aufl.). New York: Freeman.
- McCarthy, J. (1996). Mathematische Logik in Künstlicher Intelligenz. In S. R. Graubard (Hrsg.), *Probleme der Künstlichen Intelligenz. Eine Grundlagendiskussion* (S. 283-296). Wien: Springer.
- McGinn, C. (1996). Can We Solve the Mind-Body Problem? In W. Lyons (Hrsg.), *Modern Philosophy of Mind*. London: Everyman.
- Meier-Oeser, S. (Hrsg.). (1997). *René Descartes*. München: Diederichs.

- Metzinger, T. (1996a). Begriffliche Grundlagen. Einleitung. In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3.,ergänzte Aufl.). Paderborn: Schöningh.
- Metzinger, T. (1996b). Das Problem des Bewußtseins. In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3.,ergänzte Aufl.). Paderborn: Schöningh.
- Metzinger, T. (1996c). Künstliches Bewußtsein. Einleitung. In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3.,ergänzte Aufl.). Paderborn: Schöningh.
- Metzinger, T. (1996d). Neurowissenschaften und Philosophie des Bewußtseins. Einleitung. In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3.,ergänzte Aufl.). Paderborn: Schöningh.
- Moravec, H. (1996). Körper, Roboter und Geist. In C. Maar, E. Pöppel & T. Christaller (Hrsg.), *Die Technik auf dem Weg zur Seele. Forschungen an der Schnittstelle Gehirn/Computer*. Reinbek: Rowohlt.
- Oeser, E. & Seitelberger, F. (1995). *Gehirn, Bewußtsein und Erkenntnis* (2.,überarb. und erw. Aufl.). Darmstadt: Wiss. Buchgesellschaft.
- Opwis, K. & Plötzner, R. (1996). *Kognitive Psychologie mit dem Computer*. Heidelberg: Spektrum.
- Papert, S. (1996). Verstehen von Differenzen. In S. R. Graubard (Hrsg.), *Probleme der Künstlichen Intelligenz. Eine Grundlagendiskussion* (S. 1-14). Wien: Springer.
- Penrose. (1991). *Computerdenken* (Springer, Michael; Übers.). Heidelberg: Spektrum.
- Phaf, R. H. & Wolters, G. (1997). A Constructivist and Connectionist View on Conscious and Nonconscious Processes. *Philosophical Psychology*, 10(3), 287-305.
- Pöppel, E. (1996). Radikale Syntopie an der Schnittstelle von Gehirn und Computer. In C. Maar, E. Pöppel & T. Christaller (Hrsg.), *Die Technik auf dem Weg zur Seele. Forschungen an der Schnittstelle Gehirn/Computer*. Reinbek: Rowohlt.
- Popper, K. R. & Eccles, J. C. (1997). *Das Ich und sein Gehirn* (6. Aufl.). Heidelberg: Springer.
- Preilowski, B. (1993). Geist und Gehirn: Bewußtsein und Gehirnfunktionen aus der Sicht der Neuropsychologie. In A. Elefant & G. Wolters (Hrsg.), *Denkmaschinen? Interdisziplinäre Perspektiven zum Thema Geist und Gehirn*. Konstanz: Universitätsverlag Konstanz.
- Putnam, H. (1994). Geist und Maschine. In W. C. Zimmerli & S. Wolf (Hrsg.), *Künstliche Intelligenz. Philosophische Probleme*. Stuttgart: Reclam.
- Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, 88, 16-45.

- Pylyshyn, Z. W. (1989). *Computation and Cognition. Toward a Foundation for Cognitive Science* (5. Aufl.). Cambridge/MA: Bradford.
- Reeke, G. N. J. & Edelman, G. M. (1996). Wirkliche Gehirne und Künstliche Intelligenz. In S. R. Graubard (Hrsg.), *Probleme der Künstlichen Intelligenz. Eine Grundlagendiskussion* (S. 139-168). Wien: Springer.
- Reiter, M. (1996a). Nervengesteuerte Prothese. *c't*(6), 51.
- Reiter, M. (1996b). Teraflop in Deutschland. *c't*(6), 51ff.
- Roediger, H. L. & McDermott, K. B. (1993). Implicit memory in normal human subjects. In H. Spinnler & F. Boller (Hrsg.), *Handbook of Neural Psychology* (Vol. 8) (S. 63-131). Amsterdam: Elsevier.
- Rosenfield, I. (1996). Kein Erkennen ohne Gedächtnis. In C. Maar, E. Pöppel & T. Christaller (Hrsg.), *Die Technik auf dem Weg zur Seele. Forschungen an der Schnittstelle Gehirn/Computer*. Reinbek: Rowohlt.
- Rumelhart, D. E., McClelland, J. L. & Group, P. R. (1986). *Parallel Distributed Processing. Explorations in the microstructure of cognition* (Vol. 1&2). Cambridge/MA: MIT Press.
- Sarnow, K. (1990a). Matrix-Assoziationen. OCR als neuronales Netz auf dem Atari ST. *c't*(4), 218ff.
- Sarnow, K. (1990b). Simulierte Simulation. Backpropagation - Simulation von neuronalen Netzen. *c't*(10), 164ff.
- Sarnow, K. (1996). Nachgelegt. OCR: OmniPage 6.0 und EasyReader Elite 2.0. *c't*(1), 72ff.
- Saum-Aldehoff, T. (1993). Wie das Gehirn die Welt konstruiert. *Psychologie Heute*, 1, 58-63.
- Schefe, P. (1993). Informatik und Philosophie. Eine Einführung. In P. Schefe, H. Hastedt, Y. Dittrich & G. Keil (Hrsg.), *Informatik und Philosophie*. Mannheim: Bibliographisches Institut.
- Schischkoff, G. (Hrsg.). (1991). *Philosophisches Wörterbuch* (22. Aufl.). Stuttgart: Kröner.
- Schulemann, G. (1999). Biologischer Computer aus Blutegel-Neuronen. *c't*(15), 28.
- Schwartz, J. T. (1996). Der neue Konnektionismus: Über die Entwicklung von Beziehungen zwischen Neurowissenschaft und Künstlicher Intelligenz. In S. R. Graubard (Hrsg.), *Probleme der Künstlichen Intelligenz. Eine Grundlagendiskussion* (S. 121-138). Wien: Springer.

- Searle, J. R. (1993). Ist das Gehirn ein Digitalcomputer? In P. Scheffe, H. Hastedt, Y. Dittrich & G. Keil (Hrsg.), *Informatik und Philosophie*. Mannheim: Bibliographisches Institut.
- Searle, J. R. (1994). Geist, Gehirn, Programm. In W. C. Zimmerli & S. Wolf (Hrsg.), *Künstliche Intelligenz. Philosophische Probleme*. Stuttgart: Reclam.
- Searle, J. R. (1996). *Die Wiederentdeckung des Geistes* (Gavagai, Harvey P.; Übers.). München: Artemis.
- Searle, J. R. (1997). *The Mystery of Consciousness*. London: Granta.
- Siebers, A. B. J. & Weigert, M. M. (Hrsg.). (1998). *Börsen-Lexikon* (2., verbesserte Aufl.). München: R. Oldenbourg.
- Siekman, J. H. (1994). Künstliche Intelligenz. In S. Krämer (Hrsg.), *Geist - Gehirn - künstliche Intelligenz* (S. 203-222). Berlin: Walter de Gruyter.
- Silbernagl, S. & Despopoulos, A. (1988). *Taschenatlas der Physiologie* (3., überarb. und erw. Aufl.). Stuttgart: Thieme.
- Smolensky, P. (1989). Connectionist Modeling: Neural Computation/Mental Connections. In L. Nadel, L. A. Cooper, P. Culicover & R. M. Harnish (Hrsg.), *Neural Connections, Mental Connections*. London: MIT Press.
- Sokolowski, R. (1996). Natürliche und Künstliche Intelligenz. In S. R. Graubard (Hrsg.), *Probleme der Künstlichen Intelligenz. Eine Grundlegendiskussion* (S. 45-64). Wien: Springer.
- Specht, R. (1966). *Descartes* (7. Aufl.). Reinbek: Rowohlt.
- Sperlich, T. & Möcke, F. (1993). Der achte Tag. Computerwissenschaftler versuchen sich an künstlichem Leben. *c't*(8), 52ff.
- Sperry, R. W. (1976). Mental Phenomena as Causal Determinants in Brain Function. In G. G. Globus, G. Maxwell & I. Savodnik (Hrsg.), *Consciousness and the Brain. A Scientific and Philosophical Inquiry*. New York: Plenum.
- Steels, L. (1996). Homo cyber-sapiens oder Robo hominidus intelligens: Maschinen erwachen zu künstlichem Leben. In C. Maar, E. Pöppel & T. Christaller (Hrsg.), *Die Technik auf dem Weg zur Seele. Forschungen an der Schnittstelle Gehirn/Computer*. Reinbek: Rowohlt.
- Tetens, H. (1993). Informatik und die Philosophie des Geistes. In P. Scheffe, H. Hastedt, Y. Dittrich & G. Keil (Hrsg.), *Informatik und Philosophie*. Mannheim: Bibliographisches Institut.
- Tetens, H. (1994). *Geist, Gehirn, Maschine. Philosophische Versuche über ihren Zusammenhang*. Stuttgart: Reclam.
- Tholey, P. (1992). Computer. *Gestalt Theory*, 14(1), 60-67.

- Thompson, R. F. (1990). *Das Gehirn: Von der Nervenzelle zu Verhaltenssteuerung* (Behnke, Merlet; Übers.). Heidelberg: Spektrum.
- Van Gulick, R. (1996). Was würde als eine Erklärung von Bewußtsein zählen? In T. Metzinger (Hrsg.), *Bewußtsein: Beiträge aus der Gegenwartsphilosophie* (3., ergänzte Aufl.). Paderborn: Schöningh.
- Waltz, D. L. (1996). Die Perspektiven für den Bau wirklich intelligenter Maschinen. In S. R. Graubard (Hrsg.), *Probleme der Künstlichen Intelligenz. Eine Grundlegendiskussion* (S. 15-44). Wien: Springer.
- Warwick, K. (1998). *In the Mind of the Machine. The Breakthrough in Artificial Intelligence*. o.O.: Arrow.
- Weischedel, W. (Hrsg.). (1983a). *Kant, Immanuel: Kritik der reinen Vernunft. Teil 1* (5. Aufl.) (Vol. 3). Darmstadt: Wiss. Buchgesellschaft.
- Weischedel, W. (Hrsg.). (1983b). *Kant, Immanuel: Schriften zur Metaphysik und Logik* (5. Aufl.) (Vol. 5). Darmstadt: Wiss. Buchgesellschaft.
- Werntges, H. & Eckmiller, R. (1988). Neuronale Computer. Grundlagen, Stand der Forschung und erste Ergebnisse. *c't*(10), 70.
- Weyl, H. (1966). *Philosophie der Mathematik und Naturwissenschaft* (3. Aufl.). München: Oldenbourg.
- Wiener, O. (1990). *Probleme der Künstlichen Intelligenz*. Berlin: Merve.
- Wimsatt, W. C. (1976). Reductionism, Levels of Organization, and the Mind-Body Problem. In G. G. Globus, G. Maxwell & I. Savodnik (Hrsg.), *Consciousness and the Brain. A Scientific and Philosophical Inquiry*. New York: Plenum.
- Wittgenstein, L. (1997a). *Das Blaue Buch* (11. Aufl.) (Vol. 5). Frankfurt am Main: Suhrkamp.
- Wittgenstein, L. (1997b). *Tractatus logico-philosophicus* (11. Aufl.) (Vol. 1). Frankfurt am Main: Suhrkamp.
- Wolters, G. (1993). Geist und Maschine: Historische Bemerkungen zu einem nicht ganz alten philosophischen Problem. In A. Elepfandt & G. Wolters (Hrsg.), *Denkmaschinen? Interdisziplinäre Perspektiven zum Thema Geist und Gehirn*. Konstanz: Universitätsverlag Konstanz.
- Wundt, W. (1906). *Einleitung in die Philosophie* (4. Aufl.). Leipzig: Wilhelm Engelmann.
- Zimbardo, P. G. (1992). *Psychologie*. Berlin: Springer.
- Zimmerli, W. C. & Wolf, S. (Hrsg.). (1994). *Künstliche Intelligenz. Philosophische Probleme*. Stuttgart: Reclam.

Zobel, C. (1993). Teraflops in Sichtweite. Aktuelle Meilensteine in der Parallelrechner-Technologie. *c't*(8), 68ff.

Zoglauer, T. (1998). *Geist und Gehirn: Das Leib-Seele-Problem in der aktuellen Diskussion*. Göttingen: UTB.

Hiermit erkläre ich, daß ich die vorliegende Magisterabschlußarbeit selbständig verfaßt und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Magisterabschlußarbeit wurde bisher keinem anderen Prüfungsamt in gleicher oder vergleichbarer Form vorgelegt. Sie wurde bisher auch nicht veröffentlicht.

Berlin, im August 1999

Claus-Christian Carbon