
Oldenbourg

Handbuch der Künstlichen Intelligenz

herausgegeben von
Günther Görz,
Claus-Rainer Rollinger
und Josef Schneeberger

4., korrigierte Auflage

Oldenbourg Verlag München Wien

Die ersten beiden Auflagen sind unter dem Titel „Einführung in die künstliche Intelligenz“ erschienen bei Addison-Wesley (Deutschland) GmbH

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Rosenheimer Straße 145, D-81671 München
Telefon: (089) 45051-0
www.oldenbourg-verlag.de

Lektorat: Christian Kornherr
Herstellung: Rainer Hartl
Umschlagkonzeption: Kraxenberger Kommunikationshaus, München
Gedruckt auf säure- und chlorfreiem Papier
Gesamtherstellung: Druckhaus „Thomas Müntzer“ GmbH, Bad Langensalza

ISBN 3-486-27212-8

Vorwort

Das vorliegende „Handbuch der Künstlichen Intelligenz“ ist die um wesentliche Teile erweiterte und überarbeitete dritte Auflage der „Einführung in die Künstliche Intelligenz“[1].

Mit diesem Handbuch wird eine repräsentative Übersicht über die wissenschaftliche Disziplin der „Künstlichen Intelligenz“ (KI) vorgelegt, deren Autoren ausschließlich dem deutschsprachigen Raum entstammen. Nachdem auch die zweite Auflage ausverkauft war, stellte sich die Frage, welche Anforderungen an eine Neuauflage angesichts der Fortschritte in der Wissenschaft und der Veränderungen im wissenschaftsorganisatorischen, technischen und wirtschaftlichen Umfeld zu richten seien. Es lag nahe, daß es nicht nur um eine Aktualisierung der vorhandenen Beiträge gehen konnte, sondern vielmehr um eine in mehrfacher Hinsicht thematisch ergänzte und abgerundete Neufassung. Dies in Angriff zu nehmen war nur möglich durch die Bereitschaft der Kollegen Claus-Rainer Rollinger und Josef Schneeberger zur Mitarbeit im Herausgeberkreis, so daß die Verantwortung nunmehr auf drei Schultern ruht. Auch die Diskussion in der Fachgruppe 1 „Künstliche Intelligenz“ der Gesellschaft für Informatik (GI) e.V., die nach wie vor die Herausgabe dieses Werks mitträgt, stützte die Weiterentwicklung der Konzeption, die sich auch im neuen Titel ausdrückt.

Nach dem Übergang des deutschsprachigen Lehrbuchprogramms an den Oldenbourg-Verlag, München, hat letzterer nicht nur unverzüglich Gespräche mit den Herausgebern zur Planung einer Neuauflage aufgenommen, sondern die gesamten Vorbereitungsarbeiten konstruktiv begleitet, wofür ihm ein herzlicher Dank gebührt.

Das Vorwort zur ersten Auflage charaktierisierte das Werk folgendermaßen:

> „Seine Herausgeber und Autoren haben sich das Ziel gesetzt, damit eine weithin als schmerzlich empfundene Lücke auf dem Lehrbuchsektor zu schließen. (...) Daß ein solches Lehrbuch ein dringendes Desiderat ist, ist unbestritten, denn zum einen ist die KI inzwischen an vielen unserer Universitäten – zumeist als Teilgebiet der Informatik – vertreten, andererseits decken die zum Teil hervorragenden Lehrbücher aus dem angelsächsischen Sprachraum das Gebiet nicht in allen Aspekten umfassend ab. Angesichts des raschen Fortschritts der Forschung lassen manche dieser Werke in ihrer Aktualität Wünsche offen. Zudem hat die KI in Deutschland und Europa durchaus eigenständige Sichtweisen und Ansätze entwickelt, die auch den vorliegenden Band geprägt haben.“

Dieses Buch verdankt ursprünglich sein Entstehen einer Serie von Frühjahrsschulen zum Thema Künstliche Intelligenz (KIFS), die seit 1982 jährlich – von einer Ausnahme abgesehen – bis 1996 von der Fachgruppe 1 „Künstliche Intelligenz“ der Gesellschaft

[1] ursprünglich erschienen im Addison-Wesley Verlag, Bonn.

für Informatik (GI) e.V. durchgeführt werden. Ziel der Frühjahrsschulen war es, in der Form von Kursen, die jeweils in etwa einer zweistündigen Vorlesung entsprachen, eine breit angelegte moderne Einführung in das Fach sowie einen Überblick über aktuelle Forschungsgebiete zu bieten. Zu einigen der Frühjahrsschulen wurden Tagungsbände vorgelegt, die die dort gehaltenenen Kurse dokumentieren[2]. Durch den Ausbau der KI an den deutschen Universitäten und Fachhochschulen hatte die Nachfrage nach der zentralen Bildungsaufgabe, die die KIFS erfüllt hatte, spürbar nachgelassen. Ihre Nachfolge hat seitdem ein von mehreren wissenschaftlichen Gesellschaften am selben Ort durchgeführtes „Interdisziplinäres Kolleg Kognitionswissenschaft" angetreten.

Die Erfahrung mit den ersten beiden Auflagen zeigte, daß viele der Kapitel in der Lehre eingesetzt wurden, sei es als Hauptreferenz oder als ergänzende Lektüre, selten aber das Werk in seiner Gesamtheit als Lehrbuch. Dies lag zum einen an seinem Umfang, zum anderen aber auch an der Vielfalt der eher einführenden und eher vertiefenden Kapitel, die es doch von der Geschlossenheit eines klassischen Lehrbuchs unterscheidet, das von einem einzigen Autor oder einem kleinen Autorenteam verfaßt wurde. Zudem wurde vielfach angeregt, die thematische Breite des Werks zu vergrößern. Die Herausgeber hoffen, mit der Überarbeitung und der Weiterentwicklung in Richtung eines „Handbuchs" dem veränderten Anforderungsprofil weitestgehend gerecht zu werden.

Für alle Beiträge dieses Werks war maßgeblich, daß sie eine straffe und qualitativ hochstehende Darstellung des jeweiligen Themengebiets geben und Hinweise auf notwendige und sinnvolle Vertiefungsmöglichkeiten, u.a. auch in der Form einer Auswahlbibliographie, bieten. Es ist offensichtlich, daß bei einem angestrebten Umfang von ca. 50 Seiten pro Beitrag nicht *alles*, was thematisch wichtig ist, abgehandelt werden kann.

Wir danken ganz herzlich allen Autorinnen und Autoren, daß sie so engagiert an dieser nicht einfachen Aufgabe mitgewirkt haben. Neben den hohen qualitativen Anforderungen dieses ambitionierten Buchprojekts stellten die vielen notwendigen inhaltlichen Absprachen bis hin zu einer einheitlichen Layoutgestaltung und die – im großen und ganzen vorbildlich eingehaltene – terminliche Disziplin eine nicht alltägliche Herausforderung dar. Zur Qualitätssicherung wurde eine Referierungsprozedur durchgeführt: Jeder Beitrag wurde von jeweils mindestens zwei anderen Buchmitarbeitern gelesen und die kritischen Anmerkungen wurden den Autoren zur endgültigen Überarbeitung zugesandt.

Auf der Internetseite dieses Buches unter www.oldenbourg-verlag.de (Titelsuche" beim Oldenbourg Wissenschaftsverlag) haben Sie die Möglichkeit, zusätzliche Materialien zu diesem Buch kostenlos herunterzuladen. Dort sind auch alle im Buch erwähnten URLs aufgelistet und werden ggf. aktualisiert.

Günther Görz, Claus-Rainer Rollinger, Josef Schneeberger

[2]Erschienen in der Reihe Informatik-Fachberichte (IFB) im Springer-Verlag, Berlin: Teisendorf 1982: IFB Nr. 59, Dassel 1984: IFB Nr. 93, Dassel 1985: IFB Nr. 159, Günne 1987: IFB Nr. 202, Günne 1989: IFB Nr. 203

Vorwort zur 4. Auflage

Dass ein anspruchsvolles Handbuch auf so große Akzeptanz stößt, so dass nach zwei Jahren bereits eine Neuauflage notwendig wird, ist für die Herausgeber eine besondere Freude. Als wir vor einem guten halben Jahr vom Verlag informiert wurden, dass die dritte Auflage des Handbuchs noch vor dem Ende des Jahres 2002 ausverkauft sein wird, war unmittelbar klar, dass innerhalb weniger Monate eine grundsätzliche Revision nicht in Angriff genommen werden kann. Hierfür ist ein Planungshorizont von weit mehr als einem Jahr erforderlich, insbesondere dann, wenn — wie schon bei der dritten Auflage — auch ein Begutachtungsverfahren durchgeführt werden soll. Allerdings bot sich die Chance, anstelle eines bloßen Nachdrucks kurzfristig realisierbare Verbesserungen vorzunehmen. Entsprechend wurden an den Kapiteln 9, 12, 23 und 24 von den Verfassern nach Durchsicht kleinere Veränderungen vorgenommen. Die in der Zeitschrift „Künstliche Intelligenz", Nr. 4, Oktober 2002, veröffentlichte ausführliche Rezension von Mirjam Minor enthält eine Reihe wertvoller Anregungen und zweifelsohne berechtigter Verbesserungsvorschläge, die unbedingt in die für das nächste Jahr projektierte Planung der nächsten Auflage eingehen werden.

Günther Görz, Claus-Rainer Rollinger, Josef Schneeberger

Inhaltsverzeichnis

Kapitel 1

Einleitung

Günther Görz und Ipke Wachsmuth

„Künstliche Intelligenz" (KI) ist eine wissenschaftliche Disziplin, die das Ziel verfolgt, menschliche Wahrnehmungs- und Verstandesleistungen zu operationalisieren und durch Artefakte, kunstvoll gestaltete technische – insbesondere informationsverarbeitende – Systeme verfügbar zu machen[1]. Unter den zahlreichen Definitionen der Disziplin, die von ihren Fachvertretern angegeben wurden, sei beispielhaft die von [Winston 1992] formulierte genannt, die die Bestimmung des Gegenstands der KI in folgender Weise präzisiert:

> „Künstliche Intelligenz ist die Untersuchung von Berechnungsverfahren, die es ermöglichen, wahrzunehmen, zu schlußfolgern und zu handeln."

Diese Aufgabenstellung impliziert den grundsätzlich interdisziplinären Charakter der KI: Obwohl durch ihre Genese zumeist in der Informatik als Teilgebiet verankert und damit ihre ingenieurwissenschaftliche Komponente betonend, ist KI-Forschung gleichwohl nur in enger Zusammenarbeit mit Philosophie, Psychologie, Linguistik und den Neurowissenschaften möglich, die für ihre kognitionswissenschaftliche Komponente grundlegend sind. So unterscheidet sich die KI von der klassischen Informatik wegen der Betonung von Wahrnehmung, Schlußfolgern und Handeln, und sie unterscheidet sich von der Psychologie wegen der Betonung des Aspekts der Berechnung. Ein Kernpunkt dabei ist die These, daß das „Räsonnieren" konstitutiv für höhere Intelligenzfunktionen ist. Schlußfolgerndes Denken – in einem sehr allgemeinen Sinn – involviert interne Prozesse, die es einem Individuum ermöglichen, darüber nachzudenken, was die beste Weise zu handeln ist, bevor tatsächlich gehandelt wird. Entscheidend dabei ist der Rückgriff auf Wissen über die Welt und über alternative Möglichkeiten des Handelns in der Welt. Die Bezeichnung „Künstliche Intelligenz" ist historisch zu verstehen: Zunächst im Englischen als „Artificial Intelligence" geprägt, ist sie als wörtliche übersetzung nicht sinngemäß und gibt Anlaß zu dem Mißverständnis, sie würde eine Definition von „Intelligenz" liefern oder hätte gar einen operationalisierbaren Intelligenzbegriff insgesamt zu entwickeln. Da die KI eine junge Disziplin ist, zeichnet sich ihre Grundlagendiskussion zudem durch eine metaphernreiche und aufgrund ihres Gegenstands auch stark anthropomorphe Sprache aus.

[1] Mit Kant wollen wir unter „Verstand" das Vermögen der Regeln verstehen – im Unterschied zur Vernunft als Vermögen der Prinzipien.

Dennoch bleibt uns eine grundsätzlichere Auseinandersetzung mit dem Begriff der *Intelligenz* nicht erspart. „Intelligenz ist die allgemeine Fähigkeit eines Individuums, sein Denken bewußt auf neue Forderungen einzustellen; sie ist allgemeine geistige Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens." Diese noch recht unpräzise Bestimmung durch den Psychologen William Stern aus dem Jahre 1912 hat eine Vielzahl von Versuchen nach sich gezogen, eine zusammenhängende Intelligenztheorie zu erstellen, deren keiner dem komplexen Sachverhalt auch nur annähernd gerecht werden konnte [Irrgang, Klawitter 1990]. Heute besteht weitgehend Konsens darüber, daß Intelligenz zu verstehen ist als Erkenntnisvermögen, als Urteilsfähigkeit, als das Erfassen von Möglichkeiten, aber auch als das Vermögen, Zusammenhänge zu begreifen und Einsichten zu gewinnen.

Sicherlich wird Intelligenz in besonderer Weise deutlich bei der Fähigkeit, Probleme zu lösen. Die Art, die Effizienz und die Geschwindigkeit, mit der sich der Mensch bei der Problemlösung an die Umwelt anpaßt (Adaptation) oder die Umwelt an sich angleicht (Assimilation), ist ein wichtiges Merkmal von Intelligenz. Dabei äußert sich Intelligenz durchaus nicht nur in abstrakten gedanklichen Leistungen wie logischem Denken, Rechnen oder Gedächtnis und insbesondere in der Fähigkeit zur Reflexion, sondern wird ebenso offenkundig beim Umgang mit Wörtern und Sprachregeln oder beim Erkennen von Gegenständen und Situationsverläufen. Neben der konvergenten Fähigkeit, eine Vielzahl von Informationen zu kombinieren, um dadurch Lösungen zu finden, spielt bei der Problemlösung aber auch die *Kreativität* eine wichtige Rolle, insbesondere auch das Vermögen, außerhalb der aktuellen Informationen liegende Lösungsmöglichkeiten einzubeziehen. Andererseits ist gerade die Fähigkeit zur Begrenzung der Suche nach Lösungen bei hartnäckigen Problemen eine typische Leistung der Intelligenz.

Und all dies, so müssen wir an dieser Stelle fragen, soll Gegenstand einer künstlichen Intelligenz sein? Kurz gesagt: nein, denn schon wenn wir Intelligenz beurteilen oder gar messen wollen, bedarf es einer Operationalisierung, wodurch wir einen übergang vom personalen Handeln zum schematischen, nicht-personalen Operieren vollziehen. Das, was operationalisierbar ist, läßt sich grundsätzlich auch mit formalen Systemen darstellen und damit auf einem Computer berechnen. Vieles aber, was das menschliche Denken kennzeichnet und was wir mit intentionalen Termini wie Kreativität oder Bewußtsein benennen, entzieht sich weitgehend einer Operationalisierung. Dies wird jedoch angezweifelt von Vertretern der sog. „starken KI-These", die besagt, daß Bewußtseinsprozesse *nichts anderes* als Berechnungsprozesse sind, die also Intelligenz und Kognition auf bloße Informationsverarbeitung reduziert. Ein solcher Nachweis konnte aber bisher nicht erbracht werden – die Behauptung, es sei *im Prinzip* der Fall, kann den Nachweis nicht ersetzen. Hingegen wird kaum bestritten, daß Intelligenz *auch* Informationsverarbeitung ist – dies entspricht der „schwachen KI-These".

So, wie wir Intelligenz erst im sozialen Handlungszusammenhang zuschreiben, ja sie sich eigentlich erst darin konstituiert, können wir dann allerdings auch davon sprechen, daß es – in einem eingeschränkten Sinn – Intelligenz in der Mensch-Maschine-Interaktion, in der Wechselwirkung gibt, als „Intelligenz für uns". Es besteht gar keine Notwendigkeit, einem technischen System, das uns als Medium bei Problemlösungen unterstützt, Intelligenz *per se* zuzuschreiben – die Intelligenz manifestiert sich in der Interaktion.

Eine generative Theorie der Intelligenz

Zentrales wissenschaftliches Ziel der KI ist es, zu bestimmen, welche Annahmen über die Repräsentation und Verarbeitung von Wissen und den Aufbau von Systemen die verschiedenen Aspekte der Intelligenz erklären können [Winston 1992]. Der vorherrschende Gedanke in den verschiedenen theoretischen Fassungen des Intelligenzbegriffs in der KI ist, daß Intelligenz aus der Interaktion vieler einfacher Prozesse „im Konzert" emergiert und daß Prozeßmodelle intelligenten Verhaltens mit Hilfe des Computers im Detail untersucht werden können. Die Organisation des Zusammenwirkens verschiedener Softwarekomponenten, die bestimmte Teilaufgaben versehen, die Systemarchitektur, ist also ein wichtiges Thema. Sie legt die Voraussetzungen für die Entstehung von Synergieeffekten: das Zusammenspiel vieler – oft relativ einfacher – Komponenten kann komplexes Verhalten bewirken.

„Künstliche Intelligenz" ist ein synthetischer Begriff, der – vermöge seines suggestiven Potentials – viele Mißverständnisse und falsche Erwartungen verursacht hat. Sein Ursprung läßt sich auf das Jahr 1956 zurückverfolgen, ein Jahr, das in vielerlei Hinsicht bedeutsam war. Zum Beispiel erschien in diesem Jahr das Buch „Automata Studies" mit einer Reihe heute berühmter Artikel im Gebiet der Kybernetik [Shannon, McCarthy 1956]. Ebenfalls in diesem Jahr erhielten Bardeen, Shockley und Brattain den Nobelpreis für die Erfindung des Transistors. Noam Chomsky war im Begriff, seinen berühmten Artikel über syntaktische Strukturen zu veröffentlichen, der den Weg für eine theoretische Betrachtung der Sprache eröffnete [Chomsky 1957].

Die Bezeichnung „Artificial Intelligence" wurde von John McCarthy als Thema einer Konferenz geprägt, die im Sommer 1956 am Dartmouth College stattfand und an der eine Reihe renommierter Wissenschaftler teilnahmen (u.a. Marvin Minsky, Nathaniel Rochester, Claude Shannon, Allan Newell, Herbert Simon). Dieses Treffen wird allgemein als Gründungsereignis der Künstlichen Intelligenz gewertet. Im Förderungsantrag an die Rockefeller-Stiftung wurde ausgeführt (s. [McCorduck 1979], S. 93):

> „Wir schlagen eine zweimonatige Untersuchung der Künstlichen Intelligenz durch zehn Personen vor, die während des Sommers 1956 am Dartmouth College in Hanover, New Hampshire, durchgeführt werden soll. Die Untersuchung soll auf Grund der Annahme vorgehen, daß jeder Aspekt des Lernens oder jeder anderen Eigenschaft der Intelligenz im Prinzip so genau beschrieben werden kann, daß er mit einer Maschine simuliert werden kann."

Es geht, so McCarthy später, um die „Untersuchung der Struktur der Information und der Struktur von Problemlösungsprozessen, unabhängig von Anwendungen und unabhängig von ihrer Realisierung". Und Newell: „Eine wesentliche Bedingung für intelligentes Handeln hinreichender Allgemeinheit ist die Fähigkeit zur Erzeugung und Manipulation von Symbolstrukturen. Zur Realisierung symbolischer Strukturen sind sowohl die Instanz eines diskreten kombinatorischen Systems (lexikalische und syntaktische Aspekte), als auch die Zugriffsmöglichkeiten zu beliebigen zugeordneten Daten und Prozessen (Aspekte der Bezeichnung, Referenz und Bedeutung) erforderlich."

Als Instrument der Forschung sollte der Universalrechner dienen, wie Minsky begründete: „... weil Theorien von mentalen Prozessen zu komplex geworden waren und sich zu schnell entwickelt hatten, als daß sie durch gewöhnliche Maschinerie realisiert

werden konnten. Einige der Prozesse, die wir untersuchen wollen, nehmen substantielle Änderungen in ihrer eigenen Organisation vor. Die Flexibilität von Computerprogrammen erlaubt Experimente, die nahezu unmöglich in 'analogen mechanischen Vorrichtungen' wären".

Im September 1956 fand am Massachusetts Institute eine zweite wichtige Konferenz statt, das „Symposium on Information Theory". So, wie die KI ihren Ursprung auf die Dartmouth Conference zurückführt, kann dieses Symposium als Grundsteinlegung der Kognitionswissenschaft gelten (vgl. [Gardner 1985], Kap. 14). Unter den Teilnehmern beider Konferenzen waren Allen Newell und Herbert Simon. Zusammen mit John Shaw hatten sie gerade die Arbeiten an ihrem „Logic Theorist" abgeschlossen, einem Programm, das mathematische Sätze aus Whiteheads und Russells „Principia Mathematica" beweisen konnte. Dieses Programm verkörperte schon, was später der Informationsverarbeitungs-Ansatz des Modellierens genannt wurde. Der Grundgedanke dieses Ansatzes ist, daß Theorien des bewußten menschlichen Handelns auf der Basis von Informationsverarbeitungs-Systemen formuliert werden, also Systemen, die aus Speichern, Prozessoren und Steuerstrukturen bestehen und auf Datenstrukturen arbeiten. Seine zentrale Annahme besteht darin, daß im Hinblick auf intelligentes Verhalten der Mensch als ein solches System verstanden werden kann.

Ein Charakteristikum für das methodische Vorgehen der KI als akademische Disziplin ist, menschliche Intelligenz dadurch zu verstehen und zu erklären, daß es gelingt, Effekte der Intelligenz – nämlich intelligentes Verhalten – zu produzieren. Fortschritte werden aufgrund lauffähiger Systeme angestrebt: Synthese vor Analyse. Es ist nicht das Ziel, intelligente Systeme zu konstruieren, *nachdem* ein Verständnis menschlicher Intelligenz erlangt wurde, sondern menschliche Intelligenz *durch* die Konstruktion solcher Systeme verstehen zu lernen.

Die Entwicklung der KI

In der Folge der genannten Tagungen wurden an verschiedenen universitären und außeruniversitären Einrichtungen einschlägige Forschungsprojekte ins Leben gerufen. Die Prognosen waren zunächst optimistisch, ja geradezu enthusiastisch: Die Künstliche Intelligenz sollte wesentliche Probleme der Psychologie, Linguistik, Mathematik, Ingenieurwissenschaften und des Managements lösen. Fehlschläge blieben nicht aus: So erwies sich das Projekt der automatischen Sprachübersetzung, dessen Lösung man in greifbarer Nähe sah, als enorm unterschätzte Aufgabe. Erst in den neunziger Jahren wurde es – allerdings mit größerer Bescheidenheit – wieder in Angriff genommen.

In den zahlreichen Versuchen, die Aufgaben der Disziplin zu formulieren, werden zwei Aspekte deutlich, in denen kognitionswissenschaftliche und ingenieurwissenschaftliche Zielsetzungen zum Tragen kommen:

- *Kognitive Modellierung,* d.h. Simulation kognitiver Prozesse durch Informationsverarbeitungsmodelle;
- Konstruktion *„intelligenter" Systeme*, die bestimmte menschliche Wahrnehmungs- und Verstandesleistungen maschinell verfügbar machen.

In der Entwicklung der Künstlichen Intelligenz kann man vier Phasen unterscheiden: Die Gründungsphase Ende der fünfziger und Anfang der sechziger Jahre, gekennzeichnet durch erste Ansätze zur symbolischen, nicht-numerischen Informationsverarbeitung, beschäftigte sich mit der Lösung einfacher Puzzles, dem Beweisen von Sätzen der Logik und Geometrie, symbolischen mathematischen Operationen, wie unbestimmter Integration, und Spielen wie Dame und Schach. Das Gewicht lag darauf, die grundsätzliche technische Machbarkeit zu zeigen. In dieser ersten Phase – oft durch die Bezeichnung „Power-Based Approach" charakterisiert, erwartete man sehr viel von allgemeinen Problemlösungsverfahren, deren begrenzte Tragweite allerdings bald erkennbar wurde.

Die zweite Entwicklungsphase der KI ist gekennzeichnet durch die Einrichtung von Forschungsgruppen an führenden amerikanischen Universitäten, die begannen, zentrale Fragestellungen der Künstlichen Intelligenz systematisch zu bearbeiten, z.B. Sprachverarbeitung, automatisches Problemlösen und visuelle Szenenanalyse. In dieser Phase begann die massive Förderung durch die „Advanced Research Projects Agency" (ARPA) des amerikanischen Verteidigungsministeriums.

In den siebziger Jahren begann eine dritte Phase in der Entwicklung der KI, in der u.a. der Entwurf integrierter Robotersysteme und „expertenhaft problemlösender Systeme" im Mittelpunkt stand. Letztere machten Gebrauch von umfangreichen codierten Wissensbeständen über bestimmte Gebiete, zunächst in Anwendungen wie symbolische Integration oder Massenspektroskopie. Im Gegensatz zum „Power-Based Approach" trat die Verwendung formalisierten Problemlösungswissens und spezieller Verarbeitungstechniken in den Vordergrund, was durch die Bezeichnung „Knowledge-Based Approach" charakterisiert wird. Durch diese Schwerpunktsetzung wurden große Fortschritte bei Techniken der Wissensrepräsentation und in der Systemarchitektur, besonders im Hinblick auf Kontrollmechanismen, erzielt. Im weiteren Verlauf wurde erhebliches Gewicht auf komplexe Anwendungen gelegt: Erkennung kontinuierlich gesprochener Sprache, Analyse- und Synthese in der Chemie, medizinische Diagnostik und Therapie, Prospektion in der Mineralogie, Konfiguration und Fehleranalyse technischer Systeme. Zu dieser Zeit waren auch in Europa, vor allem in Großbritannien und Deutschland, KI-Forschungsgruppen an verschiedenen Universitäten entstanden und Förderprogramme installiert.

Der Eintritt in die vierte Entwicklungsphase der KI erfolgte um 1980, die vor allem durch eine umfassende Mathematisierung des Gebiets, eine Präzisierung des Konzepts der Wissensverarbeitung und das Aufgreifen neuer Themen wie Situiertheit, Verteilte KI und Neuronale Netzwerke gekennzeichnet ist. Verbunden damit ist seit etwa 1990 ein sehr deutlicher Trend zu integrierten Ansätzen zu beobachten und eine entsprechende Erweiterung der Begriffe „Wissensverarbeitung" und „intelligentes System" auf die neuen Themen. Anwendungen und Anwendungsperspektiven (in vielen Fällen mit dem Boom des Internet verbunden) beeinflussen die aktuellen Forschungsarbeiten in einem sehr hohen Maß. Zudem wurde der Bedarf deutlich, heterogene Wissensquellen in übergreifenden Anwendungen zusammenzuführen und vorhandene Wissensbestände kurzfristig auf konkrete Einsatzzwecke zuschneiden zu können. Dies führte zu Arbeiten, die sich auf die Wiederverwendung von Wissensbasen und das sog. Wissensmanagement beziehen. Ob dies als eine fünfte Entwicklungsphase der KI zu werten ist, bleibt abzuwarten.

Symbolische Repräsentation – die Wissensebene

In allen Entwicklungsphasen der KI wurde mit jeweils verschiedenen Ansätzen das Ziel verfolgt, Prinzipien der Informationsverarbeitung zu erforschen und zwar dadurch, daß

1. strikte *Formalisierungen* versucht und
2. exemplarische Realisierungen durch *Implementation* vorgenommen werden.

Dabei galt und gilt auch heute noch zentrale Aufmerksamkeit der Repräsentation und Verarbeitung von Symbolen als wichtige Basis interner Prozesse, von denen man annimmt, daß sie rationales Denken konstituieren. In der Arbeit an ihrem „Logic Theorist" hatten Simon und Newell erste Eindrücke von den Möglichkeiten des Computers zur Verarbeitung nichtnumerischer Symbole erlangt. Symbole wurden dabei als bezeichnende Objekte verstanden, die den Zugriff auf Bedeutungen – Benennungen und Beschreibungen – ermöglichen. Die symbolische Ebene, repräsentiert in den frühen Arbeiten von Newell, Shaw und Simon [Newell et al. 1958] wie auch 1956 von Bruner, Goodnow und Austin (vgl. [Bruner et al. 1956]), ermöglicht die Betrachtung von Plänen, Prozeduren und Strategien; sie stützt sich ebenfalls auf Vorstellungen regelgeleiteter generativer Systeme [Chomsky 1957].

Dabei ist der wichtigste Aspekt, daß sich geistige Fähigkeiten des Menschen auf der symbolischen Ebene unabhängig von der Betrachtung neuronaler Architekturen und Prozesse untersuchen lassen.[2] Gegenstand der symbolischen KI sind folglich nicht das Gehirn und Prozesse des Abrufs von Gedächtnisbesitz, sondern vielmehr die Bedeutung, die sich einem Prozeß vermöge symbolischer Beschreibungen zuordnen läßt. Unbestreitbar hatten die Arbeiten von Newell und Simon in der Präzisierung des Informationsverarbeitungs-Paradigmas einen entscheidenden forschungsorientierenden Einfluß, der zur Ausformung der „symbolischen KI" führte. Die These, Intelligenzphänomene allein auf der Basis von Symbolverarbeitung untersuchen zu können, ist mittlerweile durch den Einfluß der Kognitions- und der Neurowissenschaften relativiert worden; es zeigte sich, daß zur Erklärung bestimmter Phänomene – insbesondere der Wahrnehmung – der Einbezug der physikalischen Basis, auf der Intelligenz realisiert ist, zu weiteren Erkenntnissen führt. Allerdings ist damit der Ansatz der symbolischen KI nicht obsolet, sondern eher sinnvoll ergänzt und zum Teil integriert worden.

Ein zentrales Paradigma der symbolischen KI wurde mit der Beschreibung des *„intelligenten Agenten"* („general intelligent agent") [Newell, Simon 1972] formuliert. Auf einer abstrakten Ebene betrachten die Autoren den Gedächtnisbesitz des Individuums und seine Fähigkeit, beim Handeln in der Welt darauf aufzubauen, als funktionale Qualität, die sie mit *Wissen* bezeichnen. Der intelligente Agent verfügt über Sensoren, zur Wahrnehmung von Information aus seiner Umgebung, und über Aktuatoren, mit denen er die äußere Welt beeinflussen kann. Spezifisch für diese Auffassung ist, daß der Agent zu einem internen „Probehandeln" fähig ist: Bevor er in der Welt handelt und sie dadurch möglicherweise irreversibel verändert, manipuliert er eine interne Repräsentation der Außenwelt, um den Effekt alternativer, ihm zur Verfügung stehender Methoden abzuwägen. Diese sind ihm in einem internen Methodenspeicher verfügbar, und ihre Exploration wird durch ebenfalls intern verfügbares Weltwissen geleitet.

[2] Diese These ist allerdings nicht unumstritten; vor allem von Forschern auf dem Gebiet der neuronalen Netze wurde, sie zu relativieren.

Die Fragen, mit denen sich vor allem Newell in den frühen achtziger Jahren befaßte, waren die folgenden [Newell 1981]:

- Wie kann Wissen charakterisiert werden?
- Wie steht eine solche Charakterisierung in Beziehung zur Repräsentation?
- Was genau zeichnet ein System aus, wenn es über „Wissen" verfügt?

Die Hypothese einer *Wissensebene* („Knowledge Level Hypothesis") wurde von Newell in seinem Hauptvortrag auf der ersten National Conference on Artificial Intelligence in Stanford 1980 (s. [Newell 1981]) unterbreitet. In ihr wird eine besondere Systemebene postuliert, über die Ebene der Programmsymbole (und die Ebenen von Registertransfer, logischem und elektronischem Schaltkreis und physikalischem Gerät) hinausgehend, die durch Wissen als das Medium charakterisiert ist. Repräsentationen existieren auf der Symbolebene als Datenstrukturen und Prozesse, die einen Wissensbestand auf der Wissensebene realisieren. Die Verbindung zwischen Wissen und intelligentem Verhalten wird durch das *Rationalitätsprinzip* beschrieben, welches besagt: Wenn ein Agent Wissen darüber hat, daß eine seiner möglichen Aktionen zu einem seiner Ziele beiträgt, dann wird der Agent diese Aktion wählen. In dieser Perspektive spielt Wissen die Rolle der Spezifikation dessen, wozu eine Symbolstruktur in der Lage sein soll. Wichtiger noch wird mit dieser Konzeption Wissen als eine *Kompetenz* betrachtet – als ein Potential, Aktionen zu generieren (zu handeln) – und mithin als eine abstrakte Qualität, die an eine symbolische Repräsentation gebunden sein muß, um einsatzfähig zu sein. Newell und Simon postulieren, daß ein dafür geeignetes physikalisches Symbolsystem zur Ausstattung eines jeden intelligenten Agenten gehört [Newell, Simon 1972; Newell 1980].

Wissensrepräsentation und -modellierung

Eine zentrale Feststellung in Newells oben genanntem Ansatz besagt, daß Logik ein fundamentales Werkzeug für Analysen auf der Wissensebene ist und daß Implementationen von Logikformalismen als Repräsentationsmittel für Wissen dienen können. Der Wissensebenen-Ansatz in der KI ist damit ein Versuch der Mathematisierung bestimmter Aspekte der Intelligenz – unabhängig von Betrachtungen ihrer Realisierung auf Symbolebene; dies betrifft vor allem die Aspekte des rationalen Handelns und des logischen Schlußfolgerns beim Problemlösen. Dementsprechend werden Logikformalismen vielfach in der KI benutzt, um eine explizite Menge von Überzeugungen (für wahr gehaltene Aussagen, engl. „Beliefs") eines rationalen Agenten zu beschreiben. Eine solche Menge von Überzeugungen, ausgedrückt in einer Repräsentationssprache, wird typischerweise mit dem Terminus *Wissensbasis* bezeichnet.

Diese logikorientierte Auffassung der Wissensebene hat zur Klärung zahlreicher Debatten, die bis Ende der siebziger Jahre um den Begriff der internen Repräsentation geführt wurden, beigetragen [Brachman 1979]. Zum Beispiel wurden unterschiedliche Ansätze der Darstellung von Wissen – wie semantische Netzwerke oder Frame-Strukturen – als notationelle Varianten herausgestellt, soweit es Ausdrucks- und Schlußfähigkeit anbelangt [Charniak, McDermott 1985]. Die Prominenz, die diese alternativen Notationen nach wie vor in vielen Anwendungsfeldern haben, leitet sich aus ihrer „Objekt-Zentriertheit" ab, die eine Bequemlichkeit der Beschreibung von Wissensbeständen durch

ausgezeichnete Konzepte bietet, und das gesamte Gebiet der objektorientierten Programmierung ist in Verbindung damit großgeworden. Formalismen für die Wissensrepräsentation sind mittlerweile sehr weitgehend und grundsätzlich untersucht worden. Zentrale Gesichtspunkte sind hier u.a. die Ausdrucksfähigkeit und die Komplexität von Repräsentationen, aber auch ihre prädikatenlogische Rekonstruktion bzw. Spezifikation.

Ein standardisiertes Vorgehen bei der Wissens- und Domänenmodellierung hat sich als ausgesprochen schwierig erwiesen. Die Erfahrung hat gezeigt, daß beim Entwurf eines wissensbasierten Systems vor allem hinsichtlich der Wissensbasis eine komplexe kreative Design-Leistung gefordert ist, die mit beinahe jedem neuen System und für jeden weiteren Gegenstandsbereich neu erbracht werden muß. Deshalb stellt sich angesichts des wachsenden Umfangs projektierter wissensbasierter Systeme immer stärker die Frage nach einer Wiederverwendbarkeit schon existierender Wissensbasen bzw. nach einer Aggregation großer Wissensbasen aus bibliotheksmäßig gesammelten oder inkrementell entwickelten Teilen. Vorstöße in dieser Richtung sind verschiedene Ansätze des „Knowledge Sharing" (vgl. [Neches et al. 1991]) oder der Modularisierung wissensbasierter Systeme (s. z.B. [Meyer-Fujara et al. 1994]). Die Entwicklung allgemeiner Vorgehensweisen zum Entwurf wissensbasierter Systeme orientiert sich dabei an der von Newell proklamierten Wissensebene mit ihrer Abgrenzung von der Ebene der symbolischen Verarbeitung (vgl. z.B. die KADS-Methode [Schreiber 1999]). Hiermit verbindet sich der Anspruch, die Wissensinhalte und ihre Funktion für einen Systemzweck ins Zentrum der Modellierungstätigkeit zu stellen und zu abstrahieren von der Form der symbolischen Darstellung des Wissens und den symbolverarbeitenden Prozeduren, die die Funktionalität eines Systems hervorbringen.

Vorangetrieben wurde diese Entwicklung vornehmlich im Kontext des Entwurfs von Expertensystemen. Typischerweise ist der Gegenstandsbereich hier ein eng umrissenes Spezialgebiet, in dem hohes Potential an spezifischer Problemlösefähigkeit in einem weitgehend vorab festgelegten Verwendungsrahmen verlangt ist. Bei der Entwicklung von Systemen, die zur semantischen Verarbeitung von natürlicher Sprache fähig sind, geht es dagegen zentral um die Identifikation und Modellierung intersubjektivierbarer Bestände an Welt- oder Hintergrundwissen. Die Modellierung von Alltagswissen, d.h. von allgemeinen Kenntnissen und Fertigkeiten, erhält einen wesentlich höheren Stellenwert und muß umfangreicheren Begriffsuniversen Rechnung tragen. Bereits bei Expertensystemen war nun aber eine bittere Erfahrung, daß maschinell verfügbare Expertise genau dort ihre Grenzen hat, wo Alltagswissen und Alltagserfahrung entscheidend zum Tragen kommen. Menschliches Problemlösen zeichnet sich dadurch aus, daß das dabei verwendete Wissen zumeist vage und unvollständig ist. Die Qualität menschlicher Experten zeigt sich gerade darin, daß und wie sie unerwartete Effekte und Ausnahmesituationen aufgrund ihrer Berufserfahrung bewältigen können, daß sie aus Erfahrung *lernen*, ihr Wissen also ständig erweitern, und daß sie aus allgemeinem Wissen nicht nur nach festen Schlußregeln, sondern auch durch Analogie und mit Intuition Folgerungen gewinnen.

Richten sich die systematischen Ansätze im Bereich Expertensysteme vornehmlich auf Strukturen und Typentaxonomien von Problemlösungsaufgaben („Problemlöseontologien"), so stellt die systematische Untersuchung formal repräsentierbarer kognitiver Modelle menschlicher Weltwahrnehmung, wie sie etwa in den Projekten CYC [Lenat, Guha 1990] und LILOG (s. hierzu [Klose et al. 1992]) angegangen wurde, eher noch größere Anforderungen. Hier geht es nicht nur darum, generische Problemlöseaufgaben zu betrachten,

sondern auch darum, das Fakten- und Relationengefüge diverser Domänen wie auch der Strukturen von Wissensmodellen des Menschen zu erschließen – z.B. durch gestaffelte generische und bereichsbezogene formale „Ontologien" –, um den Entwurf wissensbasierter Systeme bei der Wissensrepräsentation zu systematisieren. Derartige Ansätze sind allerdings in jüngerer Zeit durch den Versuch, heterogene Informationsbestände im Internet semantisch zu erschließen, erheblich beflügelt worden.

Verteilung und Situiertheit

Eine vom Modell des „General Intelligent Agent" abweichende Perspektive wird in Minskys „Society of Mind" [Minsky 1986] eingenommen. Dieses Paradigma, welches intelligentes Verhalten in der verteilten Tätigkeit vieler kleiner und noch kleinerer Systeme („Agenten") begründet sieht, beeinflußt eine immer noch wachsende Zahl von Forschern in der KI und führt offensichtlich zu einer andersartigen Vorstellung von Intelligenz. Auf der technischen Seite haben andererseits die Versuche, immer größer und komplexer werdende wissensbasierte Systeme zu entwickeln, Nachteile zentralisierter „Single- Agent"-Architekturen enthüllt und die Konzipierung einer „Verteilten Künstlichen Intelligenz" (VKI) beflügelt [Adler et al. 1992] [Müller 1993]. Sog. Multi-Agenten-Systeme stellen den Aspekt der aufgabenbezogenen Kooperation im Wettbewerb unabhängiger (autonomer) Teilsysteme (Agenten) heraus, bei welchen kein Agent eine globale Sicht des gesamten Problemlöseprozesses innehat, also keine zentrale Systemsteuerung vorliegt.

Unter „Agenten" werden heute vielfach hardware- oder auch software-basierte Systeme („Software-Agenten") verstanden, die als mehr oder weniger unabhängige Einheiten innerhalb größerer Systeme agieren. Solche Systeme werden bereits in vielen Disziplinen betrachtet, nicht nur in der KI, sondern als Modellierungsmittel z.B. auch in der Biologie, den Wirtschafts- und den Sozialwissenschaften (Stichwort: „Sozionik"). Der Einsatz von Agenten-Techniken interessiert uns in der KI besonders im Hinblick auf Systeme, die in einer dynamischen, sich verändernden Umgebung eingesetzt werden und in größerem Umfang Anteile von Lösungen eigenständig erarbeiten können. Ähnlich wie der allgemeinere Begriff „Objekt" befindet sich der Begriff des „Agenten" noch stark in der Diskussion; in der gegenwärtigen Literatur läßt sich deswegen kaum eine allgemein akzeptierte Definition finden. Noch am ehesten ist damit ein Gesamtsystem bezeichnet, das Fähigkeiten der Wahrnehmung, Handlung und Kommunikation miteinander verbindet und, bezogen auf eine zu erfüllende Aufgabe, situationsangemessen ein- und umsetzen kann. Dabei zeichnen sich unterschiedlich stark gefaßte Agentenbegriffe ab [Wooldridge & Jennings 1995]. In einem schwachen Sinne ist ein Agent ein System mit Eigenschaften der Autonomie (selbstgesteuertes Handeln ohne direkte Außenkontrolle), sozialen Fähigkeiten (Kommunikation und Kooperation mit anderen Agenten), Reaktivität (Verhalten in Erwiderung äußerer Stimuli) und Proaktivität (zielorientiertes Verhalten und Initiative-Übernahme). In der KI werden zumeist stärkere Annahmen gemacht; hier kann ein Agent zusätzlich über „mentalistische" Eigenschaften verfügen, die mit Begriffen wie Wissen, Überzeugung, Intention, Verpflichtung und zuweilen auch Emotion charakterisiert werden.

Beim Entwurf von Multi-Agenten-Systemen ist neben der Realisierung von Fähigkeiten der einzelnen Agenten die Art der Teilnahme an einem Kooperationsverfahren zur Aufgabenverteilung mit anderen Agenten (ggfs. auch dem beteiligten Benutzer) und der Zugriff

auf Kommunikationskanäle zu regeln [Steiner et al. 1992]. Kooperationsprotokolle, die in der VKI betrachtet werden, sind z.B. Master-Slave oder Contract-Net (Vertragsverhandlung). Zur Durchführung von Kooperation erfolgt in der Regel ein Nachrichtenaustausch in einer geeigneten Kommunikationssprache; dazu wird beispielsweise die aus dem oben erwähnten Ansatz des „Knowledge Sharing" übernommene Knowledge Query and Manipulation Language (KQML) eingesetzt [Finin et al. 1997]. In Anlehnung an die Sprechakttheorie spezifiziert KQML verschiedene sog. Performative, mit denen Nachrichtentypen (wie Aussage, Frage, Antwort) mit übermittelt werden können. Für die einzelnen Agenten wird je nach Erfordernis ein offenes Fähigkeitsspektrum betrachtet (vgl. bereits bei [Müller & Siekmann 1991]), das von sensorgetriebenen, reaktiven bis zu schlußfolgernden Fähigkeiten reicht. Auch wenn in der Regel kein Agent Überblick über das gesamte zu lösende Problem hat, können „höhere" Agenten Wissen über andere Agenten und ihre Fähigkeiten haben oder erlangen.

Eines der kritischsten Probleme in bisherigen Intelligenzmodellen der KI wie auch in vielen technischen Anwendungen liegt allerdings darin, daß das benötigte Weltwissen kaum jemals vollständig verfügbar bzw. modellierbar ist. Dies beruht auf der kontextuellen Variabilität und der Vielzahl von Situationen, mit denen ein intelligenter Agent konfrontiert sein wird. Deshalb geht die Forschungsrichtung der „situierten KI" von der Erkenntnis aus, daß die Handlungsfähigkeit eines intelligenten Agenten entscheidend von seiner Verankerung in der aktuellen Situation abhängt [Brooks 1991]. Situiertheit bezieht sich auf die Fähigkeit eines intelligenten Systems, die aktuelle Situation – durch Wahrnehmung seiner Umgebung oder durch Kommunikation mit kooperierenden Partnern – in weitestgehendem Maße als Informationsquelle auszunutzen, um auch Situationen bewältigen zu können, für die kein komplettes Weltmodell vorliegt (vgl. [Lobin 1993]). Diese Ansätze haben entscheidenden Einfluß auf die Entwicklung einer kognitiven Robotik gehabt, aber auch auf Techniken für Software-Agenten, die zur Erfüllung ihrer Aufgaben durch Situierung in der digitalen Umwelt zusätzliche Informationen beschaffen können.

Raschen Aufschwung nimmt derzeit die kognitive Robotik mit der Untersuchung von stationären und mobilen Systemen, die mit Sensoren ihre dynamische Umgebung während der Ausführung von Aufgaben wahrnehmen können. Die Szenarien reichen von forschungsorientierten Ansätzen wie kommunikationsfähigen Montagerobotern und RoboCup-Fußball zu ersten Anwendungen wie teilautonomen Rollstühlen und Robotern für die Kanalisations-Inspektion.

Anwendungsfelder für Software-Agenten sind heute bereits in zahlreichen Bereichen zu finden, u.a. bei verteilten Systemen/Netzwerken und dem Internet (Stichworte: Informations- und Internet-Agenten). Im Bereich der Mensch-Maschine-Interaktion werden sog. Interface-Agenten vielfach diskutiert; sie sind typischerweise als „Personal Assistants" ausgelegt und können durch Einbettung in die Aufgabenumgebung Wissen über die Tätigkeiten, Gewohnheiten und Präferenzen ihrer Benutzer einsetzen, um an deren Stelle Handlungen auszuführen [Laurel 1990]. Um das Problem der Wissensakquisition durch explizite Programmierung zu umgehen, sind Lernverfahren für Einzel- wie auch Multi-Agenten-Systeme entwickelt worden (z.B. [Maes & Kozierok 1993] bzw. [Lenzmann & Wachsmuth 1997]).

Neuronale Netzwerke

Bereits 1949, als die ersten Digitalrechner ihren Siegeszug angetreten hatten, wurde von D.O. Hebb die Grundlage für ein Verarbeitungsmodell formuliert, das eher in der Tradition des Analogrechnens steht [Hebb 1949]. Er postulierte, daß eine Menge von (formalen) Neuronen dadurch lernen könnte, daß bei gleichzeitiger Reizung zweier Neuronen die Stärke ihrer Verbindung vergrößert würde. F. Rosenblatt griff diese Idee auf und arbeitete sie zu einer Alternative zum Konzept der KI in symbolverarbeitenden Maschinen aus:

> „Viele der Modelle, die diskutiert wurden, beschäftigen sich mit der Frage, welche logische Struktur ein System besitzen muß, um eine Eigenschaft X darzustellen... Ein alternativer Weg, auf diese Frage zu schauen, ist folgender: Was für ein System kann die Eigenschaft X (im Sinne einer Evolution) hervorbringen? Ich glaube, wir können in einer Zahl von interessanten Fällen zeigen, daß die zweite Frage gelöst werden kann, ohne die Antwort zur ersten zu kennen." [Rosenblatt 1962]

1956, im selben Jahr, als Newell und Simons Programm einfache Puzzles lösen und Sätze der Aussagenlogik beweisen konnte, war Rosenblatt bereits in der Lage, ein künstliches neuronales Netzwerk, das Perceptron, lernen zu lassen, gewisse Arten ähnlicher Muster zu klassifizieren und unähnliche auszusondern. Er sah darin eine gewisse Überlegenheit seines Ansatzes und stellte fest:

> „Als Konzept, so scheint es, hat das Perceptron ohne Zweifel Durchführbarkeit und Prinzip nichtmenschlicher Systeme begründet, die menschliche kognitive Funktionen darstellen können... Die Zukunft der Informationsverarbeitungssysteme, die mit statistischen eher als logischen Prinzipien arbeiten, scheint deutlich erkennbar." [Rosenblatt 1962]

Zunächst jedoch gewann der symbolische Ansatz in der KI die Oberhand, was nicht zuletzt darin begründet war, daß Rosenblatts Perceptron gewisse einfache logische Aufgaben nicht lösen konnte – eine Beschränkung, die aber ohne weiteres überwunden werden kann. So erfuhr Rosenblatt in den letzten fünfzehn Jahren eine Rehabilitation, und das Arbeitsgebiet der Neuronalen Netze bzw. des „Konnektionismus" hat sich rapide zu einem umfangreichen Teilgebiet der KI entwickelt. Der theoretische Informatiker B. Mahr hat diese Konzeption treffend charakterisiert, so daß wir hier auf seine Darstellung zurückgreifen [Mahr 1989]:

„Für die Erzeugung künstlicher Intelligenz wird ein Maschinenmodell zugrundegelegt, das Arbeitsweise und Struktur des Neuronengeflechts im Gehirn imitiert. Den Neuronenkernen mit ihren Dendriten und deren Verknüpfung über Synapsen entsprechen 'processor'-Knoten, die über Verbindungen miteinander gekoppelt sind. ... Die Idee des Lernens durch die Stärkung der Verbindung, die auch schon Rosenblatts Perceptron zugrundelag, findet sich hier in der Fähigkeit wieder, daß die Gewichte der Verbindungen sich ändern können und daß so nicht nur das Pattern der Verbindungen wechselt, sondern auch das Verhalten des gesamten Systems. ... Das 'Wissen', das in einem System steckt, erscheint dann als Pattern der Verbindungsgewichte. ... Künstliche neuronale Netze geben als Computerarchitektur die Manipulation bedeutungstragender Symbole auf ... Sie

stellen 'Wissen' ... nicht als aus einzelnen Wissensbestandteilen zusammengesetztes Ganzes dar."

Als Vorteile künstlicher neuronaler Netze gelten ihre Eigenschaften der verteilten Repräsentation, der Darstellung und Verarbeitung von Unschärfe, der hochgradig parallelen und verteilten Aktion und die daraus resultierende Geschwindigkeit und hohe Fehlertoleranz. Dennoch gilt für beide Ansätze, den subsymbolischen und den symbolischen, daß wohl keiner von ihnen alleine *die* Methodik der KI ausmachen kann. Vielmehr wird die Zukunft in einer Synthese beider liegen, in hybriden Systemen, in die jeder Ansatz seine besonderen Stärken einbringen kann. Überdies wird mit den Modellierungsansätzen der neuronalen Netze auch ein wichtiges Bindeglied zu den Neurowissenschaften und damit eine Erweiterung des Erkenntnisfortschritts verfügbar.

Teilbereiche der KI

Mit den folgenden Kapiteln wird versucht, die etablierten Grundlagen- und Anwendungsbereiche der KI so weit wie möglich abzudecken. Zur Frage einer systematischen Anordnung der verschiedenen Teildisziplinen der KI gibt es durchaus unterschiedliche Auffassungen. Daher wurde lediglich eine grobe Einteilung in zwei große Gruppen vorgenommen, wobei der erste Teil die grundlegenden Theorien und Methoden und der zweite die darauf aufbauenden und weiterführenden Theorien, Methoden und Anwendungen umfaßt.

- Die Betrachtung der *Kognition* als Informationsverarbeitung liefert Grundlagen für eine Fülle von Methoden der KI, die sich schon immer dadurch auszeichnete, nicht nur technische Lösungen zu erarbeiten, sondern diese zur Informationsverarbeitung in Organismen und besonders beim Menschen in Bezug zu setzen.
- *Künstliche Neuronale Netze* betrachten Verarbeitungsmodelle, die sich durch Lernfähigkeit, Darstellung und Verarbeitung von Unschärfe, hochgradig parallele Aktion und Fehlertoleranz auszeichnen.
- *Heuristische Suchverfahren*, die dem Zweck dienen, in hochkomplexen Suchräumen schnellere Lösungswege zu finden, sind in fast allen Teilgebieten der KI von großer Bedeutung.
- Die *Wissensrepräsentation* befaßt sich mit der Darstellung von Objekten, Ereignissen und Verläufen und von Performanz- und Meta-Wissen durch formale, i.a. logikbasierte Systeme.
- Kalküle für *automatisches Beweisen* werden u.a. auf die Herstellung und Überprüfung mathematischer Beweise sowie die Analyse (Verifikation) und Synthese von Programmen mit deduktiven Methoden angewandt.
- Um aus normalerweise unvollständigem Wissen dennoch Fakten ableiten zu können, die für Entscheidungen, Handlungen und Pläne erforderlich sind, werden z.B. Regeln mit Ausnahmen verwendet. *Nichtmonotones Schließen* behandelt allgemein den Umgang mit Verfahren, die fehlendes Wissen ergänzen.
- Zahlreiche Aufgaben der KI können durch Systeme von *Constraints* modelliert und gelöst werden, so daß man hier mit Recht von einer Querschnittsmethodik der KI sprechen kann.

- Gerade in alltäglichen Situationen wie auch in der Praxis technischer Anwendungen stoßen die Idealisierungen einer strikten logischen Formalisierung an Grenzen. Hier können neue Methoden zum Umgang mit *unsicherem und vagem Wissen* weiterhelfen.
- Ein wichtiger Grund für eine eigene Darstellung des *Wissens über Raum und Zeit* liegt darin, daß es eine ausgezeichnete Rolle in sehr vielen Anwendungen der KI spielt.
- *Fallbasiertes Schließen, modellbasierte Systeme und qualitative Modellierung* gehören zum zentralen Methodeninventar wissensbasierter Systeme, deren hauptsächliche Einsatzgebiete in der Lösung komplexer *Planungs-, Konfigurations-, und Diagnoseprobleme* liegen.
- Gerade auf dem Gebiet der *Planung* wurden in den letzten Jahren bahnbrechende Fortschritte erzielt, die zu effizienten Algorithmen für komplexe Planungsaufgaben und die dynamische Planrevision führten.
- Verfahren des *maschinellen Lernens* sind die Grundlage von Programmsystemen, die aus „Erfahrung" lernen, also neues Tatsachen- und Regelwissen gewinnen oder Priorisierungen adaptieren können. Sie sind u.a. auch für die Entdeckung zweckbestimmt relevanter Beziehungen in großen Datenmengen („Data Mining") von großer Bedeutung.
- Unter *„Knowledge Engineering"* werden alle Tätigkeiten zusammengefaßt, die zur Erfassung, Verwaltung und Verarbeitung großer praxisrelevanter Wissensbestände dienen.
- Verfahren zur *Verarbeitung der natürlichen Sprache* richten sich darauf, Einsicht in den „Mechanismus" der Sprache – ihren Aufbau, ihre Verarbeitung und ihre Verwendung – zu gewinnen und diese für die Mensch-Maschine-Interaktion nutzbar zu machen.
- Beim *Bildverstehen* geht es um Aufgaben der Wahrnehmung, um Merkmale aus optischen Daten zu gewinnen und daraus Interpretationen von stehenden und bewegten *Bildern* zu erzeugen.
- Die *kognitive Robotik* befaßt sich mit der Konstruktion von Robotern als autonome intelligente Systeme. Ihre besondere Herausforderung liegt in der Synthese vielfältiger Techniken, von der Sensorik über die Dateninterpretation bis hin zu Inferenz, Aktionsplanung und -ausführung in künstlichen und natürlichen Umwelten.
- *Software-Agenten* sind rein softwarebasierte autonome intelligente Systeme, die insbesondere im Umfeld des Internet eine Vielzahl von Aufgaben lösen – als einfachstes Beispiel sei hier nur das gezielte Sammeln und Filtern von Daten genannt. Komplexe Aufgaben werden typischerweise durch die Kooperation mehrerer Agenten bearbeitet.

Aus der Grundlagenforschung ging eine Reihe zunächst eher prototypischer Anwendungssysteme hervor, viele ihrer Ergebnisse sind aber heute bereits Teil in der Praxis genutzter Anwendungen geworden. Die Vorteile der *KI-Technologie* sind im wesentlichen von zweierlei Art: Zum einen eröffnet sie neue Anwendungen wie z.B. im maschinellen Sprach- oder Bildverstehen, in der Robotik und mit Expertensystemen. Zum anderen aber ermöglicht sie auch bessere Lösungen für alte Anwendungen; hierzu gehören vor allem die maschinelle Unterstützung von Planen, Entscheiden und Klassifizieren sowie

die Verwaltung, Erschließung und Auswertung großer Wissensbestände und schließlich die Simulation und die Steuerung technischer Anlagen.

Ab etwa 1990 schien sich zunächst im Gebiet „Künstliche Intelligenz" ein Paradigmenwechsel abzuzeichnen – von einer globalen Betrachtung intelligenten Verhaltens hin zu einer Sicht von einfacheren interagierenden Systemen mit unterschiedlichen Repräsentationen – oder auch gar keiner Repräsentation –, vertreten durch die Arbeiten zu Multi-Agenten-Systemen, Verteilter KI und Neuronalen Netzwerken zeigen. Diese Ansätze bringen eine Erweiterung auf die Untersuchung „situierter" Systeme ein, welche durch Sensoren und Aktuatoren in ständigem Austausch mit ihrer Umgebung stehen, um etwa auch während einer Problemlösung Situationsdaten aufzunehmen und auszuwerten. Doch wurden auch herkömmliche wissensbasierte Sichtweisen weiterentwickelt, freilich nicht mehr vorwiegend in der Form des autonomen „Expertensystems", so daß die heutige Situation eher durch das Eindringen wissensbasierter Methoden in vielfältige anspruchsvolle Anwendungssysteme gekennzeichnet ist. Manche Forderungen nach einer feinkörnigen und umfassenden formalen Modellierung komplexer Anwendungsbereiche ließen sich nicht so einfach einlösen, wie man zunächst geglaubt hatte; vor allem aus Komplexitätsgründen müssen immer wieder Vereinfachungen vorgenommen und Kompromisse zwischen formaler Ausdruckskraft und praktischer Beherrschbarkeit geschlossen werden. So bietet sich heute insgesamt ein Bild der KI, das durch eine Koexistenz unterschiedlicher Herangehensweisen, methodischer Ansätze und Lösungswege gekennzeichnet ist und damit aber auch eine beachtenswerte Bereicherung erfahren hat.

Bislang hat kein einzelner Ansatz eine Perspektive geboten, mit der sich alle Aspekte intelligenten Verhaltens reproduzieren oder erklären ließen, wie es auf der Dartmouth-Konferenz als Programm formuliert wurde. Vor zehn Jahren hat der „Scientific American" [3] Minsky mit dem treffenden Satz zitiert: „The mind is a tractor-trailor, rolling on many wheels, but AI workers keep designing unicycles." Erscheinen nach wie vor noch viele Fragen als grundsätzlich ungelöst, so gibt es doch Evidenz dafür, daß mittlerweile mehr als ein „Rad" untersucht wird und daß gerade die Integration verschiedener Ansätze weitergehende Perspektiven für die Grundlagenforschung und Anwendungsentwicklung eröffnet.

Danksagung. Die Autoren danken Clemens Beckstein für eine Reihe hilfreicher Hinweise.

Literaturverzeichnis

[Adler et al. 1992] Adler, M., Durfee, E., Huhns, M., Punch, W., Simoudis, E.: *AAAI Workshop on Cooperation Among Heterogeneous Intelligent Agents.* AI Magazine 13 (2), 1992, 39–42.

[Brachman 1979] Brachman, R.J.: *On the Epistemological Status of Semantic Networks.* In: Findler, N.V. (Ed.): *Associative Networks: Representation and Use of Knowledge by Computers.* New York: Academic Press, 1979, 3–50.

[Brooks 1991] Brooks, R.A:. *Intelligence without reason.* Proceedings IJCAI-91, Sydney, 1991, 569–595.

[Bruner et al. 1956] Bruner, J.S., Goodnow, J.J., Austin, G.A.: *A Study of Thinking.* New York: Wiley, 1979.

[3] Scientific American, Nov. 1993, Profiles: „Marvin L. Minsky – The Mastermind of Artificial Intelligence", S. 14–15

[Charniak, McDermott 1985] Charniak, E., McDermott, D.: *Introduction to Artificial Intelligence.* Reading, MA: Addison-Wesley, 1985.

[Chomsky 1957] Chomsky, N.: *Syntactic Structures.* The Hague: Mouton, 1957.

[Feigenbaum, Feldman 1963] Feigenbaum, E.A., Feldman, J.: *Computers and Thought.* New York: McGraw-Hill, 1963.

[Finin et al. 1997] Finin, T., Labrou, Y., Mayfield, J.: *KQML as an Agent Communication Language.* In: J. Bradshaw (Ed.), *Software Agents.* Cambridge, MA: MIT Press, 1997.

[Gardner 1985] Gardner, H.: *The Mind's New Science — A History of the Cognitive Revolution.* New York: Basic Books, 1985.

[Hebb 1949] Hebb, D.O.: *The Organization of Behavior.* New York: Wiley, 1949.

[Irrgang, Klawitter 1990] Irrgang, B., Klawitter, J.: *Künstliche Intelligenz - Technologischer Traum oder gesellschaftliches Trauma?* In: Irrgang, B., Klawitter, J. (Hg.): *Künstliche Intelligenz.* Edition Universitas, Stuttgart: Hirzel, 1990, 7–54.

[Klose et al. 1992] Klose, G., Lange, E., Pirlein, Th. (Hg.): *Ontologie und Axiomatik von LILOG.* Berlin: Springer (IFB 307), 1992.

[Laurel 1990] Laurel, B.: *Interface Agents: Metaphors with Character.* In: B. Laurel (Ed.): *The Art of Human-Computer Interface Design.* Reading, MA: Addison-Wesley, 1990.

[Lenat, Guha 1990] Lenat, D.B., Guha, R.V.: *Building Large Knowledge-Based Systems — Representation and Inference in the Cyc Project.* Reading, MA: Addison-Wesley, 1990.

[Lenzmann & Wachsmuth 1997] Lenzmann, B., Wachsmuth, I.: *Contract-Net-Based Learning in a User-Adaptive Interface Agency.* In G. Weiss (Ed.): *Distributed Artificial Intelligence Meets Machine Learning - Learning in Multi-Agent Environments.* Berlin u.a.: Springer (LNAI 1221), 1997, 202–222.

[Lobin 1993] Lobin, H.: *Situiertheit.* KI, 1993, Nr. 1, (Rubrik KI-Lexikon), 63

[Maes & Kozierok 1993] Maes, P., Kozierok, R.: *Learning Interface Agents.* Proceedings of the 11th National Conference on Artificial Intelligence (AAAI-93). Menlo Park: AAAI Press/The MIT Press, 1993, 459–465.

[McCorduck 1979] McCorduck, P.: *Machines Who Think.* San Francisco: Freeman, 1979.

[Mahr 1989] Mahr, B.: *Chaos-Connection. Einwände eines Informatikers.* Kursbuch 98, 1989, 83–99.

[Meyer-Fujara et al. 1994] Meyer-Fujara, J., Heller, B., Schlegelmilch, S., Wachsmuth, I.: *Knowledge-level modularization of a complex knowledge base.* In: Nebel, B., Dreschler-Fischer, L. (Eds.): *KI-94: Advances in Artificial Intelligence.* Berlin: Springer (LNAI), 1994, 214–225.

[Minsky 1986] Minsky, M.L.: *The Society of Mind.* New York: Simon & Schuster, 1986.

[Müller 1993] Müller, J. (Hg.): *Verteilte Künstliche Intelligenz - Methoden und Anwendungen.* Mannheim: BI Wissenschaftsverlag, 1993.

[Müller & Siekmann 1991] Müller, J., Siekmann, J.: *Structured Social Agents.* In: W. Brauer, D. Hernández (Hg.): *Verteilte Künstliche Intelligenz und kooperatives Arbeiten.* 4. Internationaler GI-Kongreß Wissensbasierte Systeme. Heidelberg: Springer, 1991.

[Neches et al. 1991] Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., Swartout, W.: *Enabling Technology for Knowledge Sharing.* AI Magazine 12(3), 1991, 37–56.

[Newell, Simon 1956] Newell, A., Simon, H.A.: *The Logic Theory Machine.*
IRE Transactions on Information Theory, September 1956. Abgedruckt in: [Feigenbaum, Feldman 1963].

[Newell et al. 1958] Newell, A., Shaw, J.C., Simon, H.A.: *Chess playing programs and the problem of complexity.* IBM Journal of Research and Development 2(4), 1958. Abgedruckt in: [Feigenbaum, Feldman 1963].

[Newell, Simon 1972] Newell, A., Simon, H.A.: *Human Problem Solving.* Englewood Cliffs, N.J.: Prentice-Hall, 1972.

[Newell 1980] Newell, A.: *Physical Symbol Systems.* Cognitive Science 4, 1980, 135–183

[Newell 1981] Newell, A.: *The Knowledge Level.* AI Magazine 2(2), 1981, 1–20. Wiederveröffentlicht in Artificial Intelligence 18(1), 1–20.

[Rosenblatt 1962] Rosenblatt, F.: *Strategic Approaches to the Study of Brain Models.* In: v.Foerster, H. (Ed.): *Principles of Self-Organization.* Elmsford, N.Y.: Pergamon Press, 1962, 387.

[Schreiber 1999] Schreiber, A.Th.: *Knowledge Engineering and Management: The CommonKADS Methodology.* Cambridge, MA: MIT Press, 1999

[Shannon, McCarthy 1956] Shannon, C.E., McCarthy, J.: *Automata Studies.* Annals of Mathematics Studies No. 34. Princeton, NJ: Princeton University Press, 1956.

[Steiner et al. 1992] Steiner, D., Haugeneder, H., Kolb, M., Bomarius, F., Burt, A.: *Mensch-Maschine-Kooperation.* KI Nr. 1, 1992, 59–63.

[Winston 1992] Winston, P.H.: *Artificial Intelligence.* (3rd edition) Reading, MA: Addison-Wesley, 1992.

[Wooldridge & Jennings 1995] Wooldridge, M., Jennings, N.R.: *Agent Theories, Architectures, and Languages: A Survey.* In: M. Wooldridge, N.R. Jennings (Eds.): *Intelligent Agents: Theories, Architectures, and Languages.* Berlin: Springer (LNAI 890), 1995, 1–21.

Teil I

Grundlagen

Kapitel 2

Kognition

Gerhard Strube, Christopher Habel, Lars Konieczny und Barbara Hemforth[1]

Als Teilbereich der Informatik ist die KI damit befaßt, hochkomplexe Softwaresysteme zu konstruieren, von denen die meisten direkt mit menschlichen Benutzern interagieren und kommunizieren. Außerdem sind KI-Systeme mehr und mehr Bestandteil von Anwendungen, durch die individuelle und soziale Abläufe innerhalb von Organisationen direkt beeinflußt und oft absichtsvoll gestaltet werden. Es folgt daraus, daß Informatiker über die menschliche „Umgebung" der von ihnen entwickelten Systeme ebenso gut informiert sein sollten wie über deren Software-Umgebung.

Speziell für die KI relevant ist außerdem die Überlegung, daß das Ziel der Entwicklung „intelligenter" Systeme sich am Kriterium vorhandener natürlicher Intelligenz – meist der menschlichen – messen lassen muß. Zwar müssen KI-Lösungen in der Regel (Ausnahme: Sprache) nicht das menschliche Vorbild imitieren, aber es kann nützlich sein zu wissen, wie bestimmte Leistungen von Menschen und Tieren erbracht werden.

2.1 Kognitive Systeme

Kognitive Systeme sind dadurch charakterisiert,

- daß sie in eine Umgebung (zu der in der Regel auch andere kognitive Systeme gehören) handelnd eingebunden sind und mit ihr auch informationell im Austausch stehen,
- daß sie ihr Handeln flexibel und umgebungsadaptiv steuern, und zwar dadurch, daß sie systemrelevante Aspekte der Umwelt repräsentieren, und
- daß ihre Informationsverarbeitung durch Lernfähigkeit und Antizipation ausgezeichnet ist.

Typischerweise sind kognitive Systeme entweder Organismen (biologische kognitive Systeme) oder technische Systeme (Roboter, Agenten). Es können aber auch Gruppen solcher Systeme (z.B. gemischte Mensch-Maschine-Verbünde) als ein (komplexes) kognitives System analysiert werden.

[1] Autoren: Habel (2.3 und 2.6), Konieczny & Hemforth (2.7), Strube (übrige Abschnitte).

Kognitive Prozesse werden als solche der Informationsverarbeitung verstanden, also als Berechnungsvorgänge. Dies ist die Grundlage, von der aus biologische und technische Systeme gleichermaßen hinsichtlich ihrer kognitiven Funktionen betrachtet werden können.

2.1.1 Kognitive Systeme und ihre Umwelt

Ist ein kognitives System denkbar ohne eine Umwelt, auf die es sich bezieht? Es wäre zumindest eine im strikten Sinn sinnlose Maschine. Denn die Daten eines solchen Systems sollen ja etwas (im allgemeinen etwas außerhalb des Systems Liegendes) *repräsentieren*; ohne derartige Bezugnahme kann das System sich nicht „auf etwas beziehen". Solche Bezugnahme, in der Philosophie „Intentionalität" genannt, erscheint aber als charakteristisch für menschliche Kognition; wir denken und sprechen über etwas, glauben oder wollen etwas, usw. [Searle, 1983]. Unabhängig davon, inwieweit unsere Wahrnehmung realistisch ist (oder bloß Konstruktion unserer Kognition) entwickeln sich Begriffe in handelnder Interaktion mit der Umwelt, erfahren Wörter ihre Bedeutung unter Bezugnahme auf die Welt „draußen". Dies gilt sowohl für Symbolverarbeitung – die Symbole stehen *für etwas* –, als auch für konnektionistische Ansätze, denn auch „subsymbolische" Einheiten sind nicht bedeutungslos. Ein kognitives System bedarf also einer Semantik, und diese konstituiert sich im Bezug auf äußere und innere Gegebenheiten (reale und mögliche Welten, Situationen etc.). Im übrigen gilt auch für prozedurale Ansätze der Semantik, daß die Systemhandlungen wiederum auf Externes (Kommunikationspartner, Veränderungen der Umwelt) bezogen sind.[2]

Durch den Umweltbezug kognitiver Systeme kommt es zu Rückkopplungen mit der Umwelt. Die vom System aufgenommene Information dient zur Handlungssteuerung des Systems, und die Wahrnehmung dadurch bewirkter Veränderungen wiederum zur Kontrolle der Systemtätigkeit. Am deutlichsten wird dies am Dialog zwischen zwei Systemen, deren jedes „Umwelt" für das andere ist, wie bei der Mensch-Computer-Interaktion.

2.1.2 Kognition im Dienste der Handlungskontrolle

Zwar sind kognitive Systeme durch ihre kognitiven Funktionen gekennzeichnet, aber darüber darf nicht vergessen werden, daß diese Funktionen eingebettet sind in weitere, nicht kognitive Regulationssysteme. Zu solchen „primitiven" Regulationsmechanismen zählen bei Organismen physiologische Regelkreise und Reflexe, bei technischen Systemen elementare *behaviors* (im Sinne von [Brooks, 1991]).

Für die nicht-kognitive Regulation ist kennzeichnend, daß diese Mechanismen nicht lernfähig sind, sondern höchstens Habituation (z.B. durch physiologische Ermüdung) stattfindet. Bei Organismen sind solche starren Reaktionsweisen auf spezifische Reizklassen artspezifisch vorgebildet und nicht modifizierbar (z.B. die Reaktion einer Amöbe auf

[2]Semantik umfaßt mehrere Aspekte, die alle in diesem Kapitel wichtig sind: 1. die Semantik von Symbolen oder Aussagen stellt eine Interpretation dar, 2. Symbole referieren auf etwas (z.B. in der Welt); dies nennt man extensionale Semantik, 3. Symbole bzw. die ihnen entsprechenden Begriffe stehen in bestimmten Relationen zueinander; dies nennt man intensionale Semantik. Auskunft über diese und weitere Begriffsschattierungen gibt das „Wörterbuch der Kognitionswissenschaft" [Strube 1996b].

taktile Reizung eines Pseudofüßchens). Es besteht also eine direkte, nicht modifizierbare Verknüpfung zwischen bestimmten Reizklassen und bestimmten Reaktionsweisen.

Kognitive Funktionen brechen diesen starren Zusammenhang auf: Kognition interveniert zwischen Reiz(wahrnehmung) und Verhalten: neue Reaktionsweisen werden gelernt, neue Reiz-Reaktions-Bezüge hergestellt, Reizklassen generalisiert, Verhaltensschemata modifiziert oder neu ausgebildet, und schließlich kann an die Stelle von Reiz-Reaktions-Verknüpfungen eine von den aktuellen situativen Gegebenheiten unabhängige Abschätzung der Handlungsmöglichkeiten und ihrer jeweiligen Konsequenzen treten, welche zu planen erlaubt und in eine Entscheidung für eine Handlungsalternative mündet. Kognition trägt also ganz wesentlich zu einer Bereicherung und Verbesserung der Handlungskontrolle bei. Aus einer evolutionstheoretischen Perspektive ist Kognition aufgrund dieser überlebensdienlichen Funktionen entstanden.

Wichtig ist, daß in allen natürlichen (und auch in vielen technischen) kognitiven Systemen kognitive Funktionen nicht einfach an die Stelle evolutionär älterer, nicht-kognitiver Regulation getreten sind, sondern daß sie zusätzlich zu solchen funktionieren – zuweilen nicht ohne wechselseitige Störung, so daß z.B. ein ausgelöster Reflex eine geplante Bewegung verhindert. Entwürfe einer kognitiven Architektur (s.u.) berücksichtigen dies aber bisher noch kaum.

2.1.2.1 Ein Beispiel: Kontrolle der Nahrungsaufnahme

Die Nahrungsaufnahme von Säugetieren wird durch einen physiologischen Regelkreis kontrolliert. Bei Ratten ließ sich im Laborversuch demonstrieren, wie gut das funktioniert: teilte man vier Wochen alte Ratten in drei Gruppen auf, die (i) weiterhin fressen durften, soviel sie wollten, (ii) auf Magerkost gesetzt wurden, oder (iii) gemästet wurden, so war nach zwei Wochen das durchschnittliche Körpergewicht deutlich unterschiedlich. Danach ließ man alle Ratten wieder nach ihrem Belieben fressen, und binnen dreier weiterer Wochen hatten alle drei Gruppen wieder gleiches Durchschnittsgewicht (Keesey et. al. zit. n. [Kandel *et.al.*, 1995]). Doch bei uns Menschen klappt das offenbar nicht so gut.

Ein wesentlicher Grund ist, daß das primäre Hungermotiv durch zahlreiche erlernte, sog. sekundäre Motive überlagert wird, die fördernd oder hemmend auf die Nahrungsaufnahme wirken: Aussehen und Geruch von Speisen; sozialer Druck, noch etwas mehr zu nehmen; sozialer Druck, schlank zu erscheinen, usw. Solche kognitiv vermittelten Motive können die physiologische Regelung so stark überlagern, daß es zu massiven Eßstörungen kommt.

2.1.2.2 Interaktion von kognitiver und nicht-kognitiver Kontrolle

Die meisten Körperfunktionen sind nicht so „offen" für kognitiv vermittelte Beeinflussung wie die Nahrungsaufnahme: Man kann zwar willentlich verhungern, aber nicht willentlich so lange die Luft anhalten, bis man erstickt. Anderes, wie der Blutdruck, läßt sich normalerweise überhaupt nicht willentlich beeinflussen. Mit der Technik des Bio-Feedback hingegen (der Blutdruck wird fortlaufend gemessen und für die Probanden auf einem Monitor als Farbbalken dargestellt) läßt sich eine kognitive Kontrolle des Blutdrucks erlernen. Daraus wird ersichtlich, daß es für kognitive Kontrolle einer bewußt wahrnehmbaren

und mental repräsentierten Größe bedarf: eine weitere Veranschaulichung, daß kognitive Prozesse nur auf mentalen Repräsentationen operieren können. (Ein Rahmenmodell für die Interaktion unterschiedlicher Kontrollsysteme findet sich bei [Strube, 1998], zur Psychologie intentionaler Handlungskontrolle vgl. [Goschke, 1996]).

2.1.3 Kognition und Intelligenz

Wie verhält Kognition sich zu dem, was wir mit „Intelligenz" bezeichnen? In der Psychologie (und auch im Alltag) versteht man unter „Intelligenz" eine allgemeine geistige Leistungsfähigkeit, die bei Menschen in unterschiedlichem Maße ausgeprägt ist: Es gibt intelligentere und weniger intelligente Personen.

Im Unterschied dazu versteht man in der KI unter „Intelligenz" so etwas wie die artspezifische menschliche geistige Leistungsfähigkeit: Erst wenn KI-Systeme leisten können, was ein dreijähriges Kind oder ein (wenig „intelligenter") Erwachsener vermögen, ist das ursprüngliche Leitziel der KI erreicht. Allerdings ist auch richtig, daß gerade die frühe KI sich für Menschen besonders schwierige Aufgaben (etwa das Schachspielen) vorgenommen hat.

In der Kognitionsforschung ist wenig von Intelligenz die Rede. Das hat mit der alltäglichen Abnutzung des Intelligenzbegriffs zu tun, aber auch damit, daß das Hauptinteresse nicht dem Vergleich kognitiver Leistungen gilt, sondern der Aufklärung, aufgrund welcher Repräsentationen und Prozesse sie zustande kommen.

2.1.4 Kompetenz und Performanz

Eine kognitionswissenschaftliche Theorie ist in der Regel als implementiertes oder implementierbares Modell formuliert (s. Abschnitt 2.8). Hierbei lassen sich zwei Zielrichtungen unterscheiden, nämlich

- zu klären, auf welcher Grundlage eine kognitive Leistung überhaupt zustande kommen kann (Kompetenzmodell), oder
- zu klären, angesichts welcher strukturellen und funktionalen Gegebenheiten (z.B. Beschränkungen) eine solche Leistung konkret bei bestimmten Systemen (z.B. bei Individuen) realisiert wird (Performanzmodell).

Die Unterscheidung Kompetenz–Performanz geht auf Chomsky [1965] zurück, der linguistische Kompetenztheorien charakterisiert als „bezogen auf einen idealen Sprecher-Hörer in einer völlig homogenen Sprachgemeinschaft, der seine Sprache perfekt beherrscht und völlig unabhängig ist von beschränkter Gedächtniskapazität, Ablenkungen der Aufmerksamkeit, wechselnden Interessen und Fehlern..."

Demgegenüber sind die meisten kognitionswissenschaftlichen Modelle Performanzmodelle, die zwar von zufälligen Fehlern und interindividuellen Unterschieden abstrahieren, aber strukturelle Beschränkungen der Informationsverarbeitung (z.B. knappe Ressourcen) einbeziehen.

2.2 Menschliche Informationsverarbeitung

Wenn wir der grundlegenden These der Kognitionswissenschaft folgen, wonach alle geistigen Prozesse als solche der Informationsverarbeitung anzusehen sind, dann stellt sich als erstes die Frage: Wie können wir uns die „Architektur" der menschlichen Kognition vorstellen? An der Vielzahl konkreter Vorschläge wird auch deutlich, daß es nicht um eine Computer*metapher* geht, sondern um die ernst gemeinte Annahme, daß kognitive Prozesse Berechnungsvorgänge sind.

2.2.1 Kognitionspsychologische Ansätze

Das traditionsreichste Modell in der KI wie der kognitiven Psychologie ist das der aufeinanderfolgenden Verarbeitungsstufen. Neisser [1976] charakterisierte sie bissig als *„processing, more processing, still more processing..."* Eine solche Kaskaden-Architektur ist nicht mit serieller (vs. paralleler) Verarbeitung zu verwechseln. Die frühen Stadien der Verarbeitung visueller Reize zeigen, daß eine kaskadierte Architektur sehr wohl mit hochparalleler Verarbeitung gekoppelt sein kann. Dieses Modell, das in seiner einflußreichsten Fassung [Atkinson und Shiffrin, 1968] in strikter Analogie zu herkömmlichen Computerarchitekturen gestaltet wurde, dominierte die kognitionspsychologische Forschung bis in die achtziger Jahre. Sein Hauptkennzeichen ist eine Art CPU, das Kurzzeitgedächtnis, dessen angenommene Kapazität von ungefähr 7 Einheiten (wobei jede Einheit ein bekanntes Konzept oder Faktum ist) recht gut empirischen Ergebnissen [Miller, 1956] entspricht. Eine Weiterentwicklung stellt das Strukturmodell des Arbeitsgedächtnisses von [Baddeley, 1986] dar (vgl. 2.4.2).

2.2.2 Kognitionswissenschaftliche Ansätze

Alle sogenannten „kognitiven Architekturen" gehen letztlich auf das an der CMU bereits in den sechziger Jahren in der Gruppe um Newell und Simon entwickelte Programm GPS (General Problem Solver, ein Programm, das Mittel-Ziel-Analyse als Suchheuristik verwendete) und dessen Reimplementierung als Produktionensystem zurück [Newell und Simon, 1972].

Produktionensysteme bestehen aus (i) einem (beliebig großen) Arbeitsgedächtnis, das als Input wie als Output für die in (ii) einem Langzeitgedächtnis gespeicherten Produktionsregeln dient. Dies sind Wenn-dann-Regeln, deren Bedingungsteil mit dem aktuellen Inhalt des Arbeitsgedächtnisses verglichen wird. Bei Übereinstimmung kann die Regel ausgeführt werden („feuern") und so den Inhalt des Arbeitsgedächtnisses verändern. (iii) Ein Regelinterpreter sorgt für die Ausführung und Konfliktlösung (z.B. für den relativ häufigen Fall, daß mehrere Regeln feuern könnten). Zur Realisierung der Mittel-Ziel-Analyse wurde ein *goal stack* als besonderer Bestandteil des Arbeitsgedächtnisses eingeführt.

Aus dieser Grundarchitektur haben Newell und Mitarbeiter das System SOAR entwickelt [Laird *et.al.*, 1987; Newell, 1990]. Dieses System zeichnet sich gegenüber seiner Urform vor allem durch Lernfähigkeit aus (zur Zielerreichung nacheinander ausgeführte Operationen werden bei Erfolg zu einer komplexen Produktion compiliert); auch ist der

Regelspeicher den Zielen entsprechend partitioniert. Zahlreiche anspruchsvolle Anwendungen sind inzwischen in SOAR implementiert worden, z.B. eine Luftkampf-Simulation, die zur Pilotenausbildung bei der USAF verwendet wird [Tambe *et al.*, 1995].

Ebenfalls an der CMU wurde das Produktionssystem ACT [Anderson, 1976] samt seinen Nachfolgern ACT* (Adaptive Control of Thought: [Anderson, 1983]) und ACT-R (Atomic Components of Thought: [Anderson und Lebiere, 1998]) entwickelt. In seiner gegenwärtigen Form zeichnet sich diese Architektur dadurch aus, daß sie außer dem „prozeduralen Gedächtnis" (d.h. dem Regelspeicher) ein deklaratives Gedächtnis besitzt, nämlich ein konnektionistisches Netzwerk, dessen Knoten Konzepte sind, auf die per Aktivationsausbreitung zugegriffen werden kann. ACT-R hat also mehrere Repräsentationsformalismen und ist somit eine hybride Architektur. Gegenwärtig wird verstärkt versucht, Komponenten für Wahrnehmung und Motorik zu integrieren, um der Situiertheit menschlicher Kognition Rechnung tragen zu können.

Weitere ähnliche Architekturen sind EPIC [Meyer und Kieras, 1997] und CAPS [Just und Carpenter, 1992], das explizit die knappe Kapazität des menschlichen Arbeitsgedächtnisses berücksichtigt. Überhaupt ist kritisch an den gängigen „kognitiven Architekturen" zu sehen, daß sie zu wenige Constraints aufweisen und im Verdacht stehen, Turing-äquivalent zu sein (also alles zuzulassen); in ACT-R hat man deshalb bereits versucht, drastische Einschränkungen zu definieren.

2.2.3 KI-Architekturen

Grundsätzlich stellen alle Berechnungsmodelle – also alle KI-Architekturen samt den hier nicht näher betrachteten konnektionistischen Netzen (vgl. dazu Kap. 3) – mögliche Architekturen für kognitive Modelle dar.

Überblickshalber seien kurz erwähnt: (1) die Kaskadenarchitektur. Hier wird die serielle Verarbeitung ergänzt um die Möglichkeit einer Rückkopplung jeweils zwischen benachbarten Verarbeitungsschritten. (2) die hierarchische Architektur. Dies ist wohl das in der KI verbreitetste Verarbeitungsmodell, bei dem die Kontrolle der gesamten Verarbeitung durch ein besonderes Modul („Supervisor") gewährleistet wird. (3) Produktionssysteme (s.o.) ergänzen die direkte Kontrolle, die durch einen Interpreter realisiert wird, durch eine zentrale Datenstruktur („Arbeitsgedächtnis"). (4) die Blackboard-Architektur [Erman *et.al.*, 1980]. Hier kommunizieren mehrere autonome Subsysteme durch eine gemeinsam zugängliche Datenstruktur, die sogenannte Blackboard. (5) im Multi-Agenten-Modell [Hewitt, 1977] wird auch die Kommunikation dezentralisiert, indem autonome Subsysteme (Agenten, „actors") direkt Nachrichten austauschen; diesem Paradigma entspricht die Softwaretechnik des objektorientierten Programmierens. Hierher gehören auch quasi-konnektionistische Modelle wie das von Maes [1994]. (6) Subsumptionsarchitektur [Brooks, 1991]. Diese basiert auf selbständigen Funktionseinheiten, über die weitere, diese aktivierende oder hemmende Einheiten, geschichtet werden. (7) Agenten-Architekturen, insbesondere solche für soziale Agenten (z.B. INTERRAP: [Müller, 1996]).

2.2.4 Modularität von Geist und Gehirn

Heute hat sich allgemein die Ansicht durchgesetzt, daß die funktionelle Architektur des menschlichen Zentralnervensystems sehr komplex ist. Bereits Arbib [1972] bezeichnete den Cortex als einen *somatotopic layered computer,* womit er die in (meist sechs) Zellschichten organisierte und topologisch dem Körperschema bzw. dem Sehfeld entsprechende Struktur der visuellen Verarbeitung wie der Bewegungssteuerung meinte.

Obwohl die funktionelle Architektur des Gehirns nur bedingt der sichtbaren neuroanatomischen Struktur entspricht, lassen sich für eine große Zahl von Funktionsbereichen grobe Lokalisierungen im Gehirn angeben, insbesondere seitdem die bildgebenden Verfahren (PET; Kernspintomografie, fMRI) in die Hirnforschung Einzug gehalten haben. Aber auch neuropsychologische Untersuchungen an hirngeschädigten Personen [Wallesch *et.al.*, 1996] geben Hinweise auf die Lokalisierung von Funktionen, beispielsweise darauf, daß semantische Kategorien an unterschiedlichen Orten (und dies auch in unterschiedlicher Weise bei verschiedenen Personen) gespeichert sind [Rapp und Caramazza, 1995].

Insbesondere das Phänomen der Hemisphärendominanz ist intensiv erforscht worden. Bei fast allen Menschen sind die sprachbezogenen Funktionen in einer Großhirnhälfte lokalisiert (bei den meisten die linke). Insgesamt aber sind viele Ergebnisse nicht eindeutig und mit methodischen Problemen behaftet, obwohl auch hier die bildgebenden Verfahren helfen (Überblick bei [Gazzaniga und Hutsler, 1999]).

Daß viele kognitive Prozesse in spezifischen Bereichen des Gehirns ablaufen, ist Konsens. Allerdings hat Fodor [1983] mit seiner These, daß auch die sprachbezogenen Funktionen ein von anderen Funktionsbereichen weitgehend isoliertes und funktionell autonomes Subsystem, ein „Modul", darstellten, eine hitzige Diskussion entfacht [Garfield, 1987]. Für viele Funktionen, beispielsweise die Bewegungssteuerung, sind inzwischen die einzelnen Funktionseinheiten bekannt. Der gegenwärtige Forschungsstand läßt sich dahingehend zusammenfassen, daß die funktionelle Organisation des Gehirns zahlreiche neuroanatomisch lokalisierbare und funktionell spezialisierte Einheiten erkennen läßt, die parallel zu anderen kognitiven Prozessen arbeiten, und in deren Tätigkeit uns subjektive Einsicht (per Introspektion) versagt bleibt [Kolb und Wishaw, 1990; Karmiloff-Smith, 1999].

Von daher erscheinen die eher monolithisch organisierten „kognitiven Architekturen" (s.o.) weniger plausibel als relativ selbständige, kleinere Module. Zudem darf, wie oben erwähnt, nicht vergessen werden, daß kognitive Funktionen beim Menschen mit physiologischen Regulationsmechanismen ebenso koexistieren wie mit motivationalen und emotionalen Subsystemen.

2.2.5 Lernfähigkeit

Die Adaptivität natürlicher kognitiver Systeme, ihre Lernfähigkeit, ist vielleicht seine hervorstechendste Eigenschaft und zugleich diejenige, in der sie sich von den heutigen technischen Systemen am deutlichsten unterscheiden. Denn trotz zahlreicher Algorithmen des maschinellen Lernens (vgl. Kap. X) und der für konnektionistische Netze entwickelten Lernverfahren [Hinton, 1989; Hassoun, 1995] erreicht keine in der KI angewandte Methode auch nur annähernd die Lernleistungen des menschlichen kognitiven Systems. Ein Beispiel dafür, das bis heute die Forscher fasziniert, ist der Erwerb der Muttersprache,

wofür nicht nur von Chomsky [1965], sondern auch von ganz anderer Seite [Prince und Smolensky, 1997] besondere artspezifische Constraints angenommen werden, weil rein induktive Lernverfahren diese Leistung nicht erklären können [Pinker, 1984]. Demnach wäre der Spracherwerb nicht nur auf allgemeines Lernen, sondern auch auf eine angeborene sprachspezifische Lernbereitschaft zurückzuführen (vgl. hierzu auch das sehr lesenswerte Buch von Pinker [1994]). Allgemein kann man davon ausgehen, daß kognitive Fertigkeiten immer in Interaktion genetisch angelegter Lernbereitschaften mit Lerngelegenheiten der Umwelt erworben werden.

Dabei bezeichnet „Lernen" Vielfältiges: vom Speichern explizit vermittelten deklarativen Wissens über Erwerb und Übung prozeduralen Wissens und speziell von Bewegungsmustern bis hin zur Ausbildung umgebungsadäquater habitueller Verhaltensmuster – letzterem gilt ein gut Teil unserer Erziehung, die Tierdressur und die Aufmerksamkeit der behavioristischen Psychologie (ca. 1915-1960), deren Lerntheorien fast ausschließlich der Auslösung, Unterdrückung und Modifikation bereits im Verhaltensrepertoire befindlicher Verhaltensmuster gewidmet sind, nicht dem Erwerb neuen Verhaltens.
Hierher gehört insbesondere die zuerst von Pawlow beschriebene Technik des klassischen Konditionierens, bei dem ein ursprünglich neutraler Reiz (z.B. ein Klingelton) durch wiederholte zeitliche Nähe zu einem biologischen Auslösereiz (z.B. Futter beim berühmten Hundeversuch Pawlows) selbst verhaltensauslösende Qualität erlangt. Von ähnlich grundlegender Bedeutung ist die von Skinner entwickelte Technik der Verhaltensmodifikation durch „Verstärkung" (Bekräftigung, reinforcement), auch als operantes Konditionieren bekannt. In diesen Forschungstraditionen sind auch die Bedingungen von Generalisierung und Reizdiskriminierung ausgiebig untersucht worden.

In neuerer Zeit sind sowohl Konditionierungsverfahren als kognitive Prozesse reanalysiert [Rescorla und Wagner, 1972], als auch in einen größeren ethologischen und kognitionswissenschaftlichen Kontext gestellt worden [Gallistel, 1993]. Auch im KI-Kontext sind Modelle assoziativen Lernens entstanden (z.B. Classifier-Systeme), die zum Teil strikt an psychologischen Resultaten orientiert sind und für die Robotik verwendet werden [Balkenius, 1995]. Die neuronale Basis scheint nach neuesten Befunden in der Deblockierung von NMDA-Rezeptoren in Synapsen nach beiderseitiger Aktivierung zu bestehen [Tang *et al.*, 1999] – eine späte Bestätigung der Vermutungen von Hebb [1949].

Das Korrelat zu diesen Arten des Lernens, die in Reiz-Reaktionsverbindungen (oder Reaktions-Antizipations-Verbindungen) resultieren und daher als assoziativ zu kennzeichnen sind, ist die (motorische und auch kognitive) *Übung*, durch die Sequenzen von Bewegungen oder mentalen Inferenzschritten sozusagen zu Paketen gebündelt werden, so daß motorische Programme entstehen (die vermutlich im Kleinhirn gespeichert sind) oder komplexe Regeln. Solche Prozesse werden in SOAR als *chunking*, in ACT* als *knowledge compilation* bezeichnet.

Daneben gibt es noch zwei wichtige Lernformen. Die erste ist die Imitation von Handlungen, die zweite direkte Instruktion. Für Säuglinge haben Meltzoff und Moore [1983] nachgewiesen, daß sie schon binnen einer Stunde nach der Geburt Mimik (Mund öffnen, Zunge herausstrecken) imitieren können. Diese angeborene Imitationsfähigkeit bietet die Basis ganz wesentlicher Lernvorgänge im sozialen und motorischen Bereich. Die Basis dieser Fähigkeit ist noch weitgehend ungeklärt, obwohl bei Affen Neurone gefunden wurden, die sowohl beim Beobachten wie beim Ausführen einer bestimmten motorischen

Handlung aktiv waren [Rizzolatti *et.al.*, 1996].

Wissenserwerb durch Lesen oder Hören ist die verbreitetste schulische Form des Lernens. Untersucht wurde hier vor allem der Wissenserwerb aus Texten und Beispielen [Kintsch, 1997; Schmalhofer, 1999]. Dabei geht es um Fakten und Beschreibungen von Handlungen, die dann (oft mühsam) in Handeln umgesetzt werden können, aber durch Übung prozeduralisiert werden können (*knowledge compilation*, s.o.). In neueren pädagogischen Ansätzen wird versucht, diese Art des Lernens mit anderen zu kombinieren (z.B. mit explorativem Lernen, wo in einer Lernumwelt Antizipationen von Handlungsfolgen gelernt werden können).

2.2.6 Charakteristika menschlicher Informationsverarbeitung

Menschliche Kognition hat einige charakteristische Züge, deren Modellierung auf Computern zwar möglich ist, die aber einem einfach-klassischen Symbolverarbeitungsansatz Mühe machen.

2.2.6.1 Fließende Grenzen und graduelle Unterschiede

Wir denken selten in streng getrennten Kategorien, sondern bilden in der Regel fließende Übergänge. Dies gilt für Begriffe ebenso wie für fast alle Arten von Urteilen. Vergleichsurteile des „mehr" oder „weniger" sind leichter zu treffen als absolute Klassifizierungen. Und selbst dort gibt es subtile Abstufungen. Beispielsweise lassen sich Urteile, welche Zahl eines Paares natürlicher Zahlen größer ist, umso schneller fällen, je größer die Differenz dieser Zahlen ist (symbolischer Distanzeffekt, [Moyer und Landauer, 1967]).

Daß die Zugehörigkeit zu einer Kategorie als graduell abgestufte Typikalität wahrgenommen wird, hat zur Entstehung der Prototypentheorie) von Begriffen [Rosch, 1973; Rosch, 1975] Anlaß gegeben. Typikalitätseffekte treten aber auch bei künstlichen Kategorien auf (z.B. ist 16 eine besonders „typische" gerade Zahl) und sogar bei „Ad-hoc-Kategorien" wie „Dinge, die man auf dem Flohmarkt verkaufen kann" [Barsalou, 1983]. Hierzu Näheres in Abschnitt 2.3.1.

2.2.6.2 Kontextbezug und Repräsentativität

Wir scheinen Sachverhalte nie ohne den Kontext wahrzunehmen, in den sie eingebettet sind, und können sie auch kaum aus diesem Kontext lösen. Daß selbst irrelevante Kontexte wirksam sind, ist u.a. von Godden und Baddeley [1975] demonstriert worden. Sie ließen englische Marinetaucher Wortlisten lernen, und zwar einige unter Wasser, andere am Strand. Die Listen konnten dann besser erinnert werden, wenn sie in der gleichen Umgebung abgefragt wurden, in der sie auch gelernt worden waren.

Im günstigsten Fall erleichtert ein passender Kontext die Erkennung von Objekten [Biederman *et.al.*, 1982] oder die Lösung sonst schwieriger Probleme (z.B. [Wason und Johnson-Laird, 1972]). Andererseits kann die Orientierung an typischen Kontexten zu eklatanten Verstößen gegen die Logik führen, wie Johnson-Laird & Byrne [1991] anhand deduktiver Schlüsse, oder Kahneman und Tversky [1973; 1982] am Beispiel von Wahrscheinlichkeitsurteilen nachgewiesen haben. Hier eine ihrer Aufgaben:

Vorgelegt wird folgende Personenbeschreibung: „Linda ist 31 Jahre alt, sie lebt allein, ist freimütig (*outspoken*), sehr begabt und Magister der Philosophie. Während ihres Studiums hat sie sich intensiv mit Fragen sozialer Gerechtigkeit und Diskriminierung befaßt und auch an Anti-Atom-Demonstrationen teilgenommen." Danach ist die Wahrscheinlichkeit des Zutreffens einer Reihe von Aussagen über Linda zu beurteilen, darunter die beiden folgenden: (a) „Linda ist Kassiererin bei einer Bank" und (a∧b) „Linda ist Kassiererin bei einer Bank und in der Frauenbewegung aktiv". Obwohl aus rein logischen Gründen die Wahrscheinlichkeit für (a∧b) höchstens so groß sein kann wie die für (a) allein, waren über 80% der Versuchspersonen von Kahneman und Tversky [1982] der Ansicht, (a∧b) sei wahrscheinlicher. Selbst Studierende mit nachgewiesenen Statistik-Kenntnissen blieben von diesem Fehlschluß nicht verschont. Kahneman und Tversky erklären dieses Ergebnis damit, daß die Repräsentativität der Aussage im Kontext beurteilt wird.

2.2.6.3 Beschränkte Rationalität

Die von Kahneman und Tversky beobachteten Phänomene werden verständlich unter der Annahme, daß wir Menschen oft nicht alle uns verfügbare Information in die Verarbeitung einbeziehen, sondern uns auf Heuristiken (wie Repräsentativität oder leichte Verfügbarkeit beim Erinnern) oder auf stark vereinfachte „mentale Algorithmen" [Gigerenzer und Goldstein, 1996] stützen. Dieses Prinzip geht letztlich auf ein schon von Simon [1969] formuliertes Prinzip zurück, demzufolge es unter den Bedingungen der Evolution ausreicht, ein „genügend gutes" (*satisficing*) Verfahren zur Verfügung zu haben.

Daraus ergibt sich eine Lösung für das Paradox, daß die meisten der für die KI interessanten Probleme beweisbar NP-hart sind: Die Algorithmen menschlicher oder tierischer Kognition lösen gar nicht das gesamte Problem, sondern eine leichter bewältigbare Unterklasse. Solange die (in der Natur oft tödlichen) Folgen nicht den Bestand der Art gefährden, sind diese Mechanismen „gut genug".

Dennoch ist es sinnvoll, Menschen (und Tieren, vgl. [McFarland und Bösser, 1996]) Rationalität zuzuschreiben. Rationalität meint, daß Systeme gemäß ihrer jeweils verfolgten Zielsetzung handeln und den besten Weg zur Erreichung ihres Zieles suchen. Dies schließt nicht aus, daß unter realistischen Ressourcenbeschränkungen (z.B. Zeitdruck) suboptimale, aber dennoch brauchbare Lösungen akzeptiert werden. Eben dies wird „beschränkte Rationalität" (*bounded rationality*) genannt.

2.3 Denken und Problemlösen

Unter der Perspektive des Informationsverarbeitungsparadigmas kann Denken als mentale Tätigkeit über internen Repräsentationen aufgefaßt werden; Prozesse des Denkens und Problemlösens basieren auf internen Repräsentationen der externen Welt [Smith, 1995].

2.3.1 Begriffe und Kategorien

„Wozu brauchen wir Begriffe?" Mit dieser Frage beginnt J. Hoffmann [1986] seine Darstellung psychologischer Untersuchungen des menschlichen Wissens. Die charakteristische

Grundsituation bei der menschlichen Kognition und Perzeption, die von manchen geradezu als paradox empfunden wird, ist, daß die Aufgaben, denen wir uns beim Wahrnehmen, Denken und Problemlösen gegenüber sehen, im überwiegenden Teil der Fälle neu sind: Wir sehen ständig neue Objekte, Dinge und Personen in unbekannten Umgebungen, wir hören neue Sätze über neue Sachverhalte, wir haben neue Probleme zu lösen, und trotz der überwältigenden Anzahl neuer Gegebenheiten können wir in den meisten Fällen die genannten Aufgaben ohne erkennbaren zusätzlichen Aufwand bewältigen.

Der Grund hierfür liegt darin, daß die genannten Aufgaben eben nicht vollständig neu sind, sondern in einer engen, normalerweise systematischen Beziehung zu alten, d.h. vertrauten und bekannten, Gegebenheiten stehen. Oder anders ausgedrückt: Die Lösung kognitiver und perzeptiver Aufgaben basiert auf früheren Erfahrungen, auf unserem Vorwissen. (vgl. Abschnitte 2.4 und 2.6.)
Nun ist es nicht damit getan, Vorwissen zu besitzen; die entscheidende Leistung besteht darin, auf das Vorwissen, d.h. auf die Erinnerung, strukturiert so zugreifen zu können, daß die für die Lösung kognitiver Aufgaben benötigten Wissensentitäten auch wirklich zur Verfügung stehen. Der Schlüssel zu diesem Problem liegt darin, daß Individuen, über die wir nachdenken oder sprechen bzw. die wir wahrnehmen, Klassen zugeordnet werden können. Anders ausgedrückt:

- Begriffe schaffen Ordnung in unserem Denken dadurch, daß sie Klassen von Objekten mit gemeinsamen Eigenschaften bereitstellen.

Mit *Begriff*, bzw. – angelsächsischen Tradition der Kognitionswissenschaft folgend – mit *Konzept*, beziehen wir uns hierbei auf kognitive Entitäten, d.h. interne Objekte unserer Kognition. Sie korrespondieren zu Klassen von Entitäten der realen, externen Welt, die wir im weiteren als *Kategorien* bezeichnen. Diese sehr allgemeine Charakterisierung von Begriff ist mit eigentlich allen Sichtweisen auf die Frage „Was sind und was leisten Begriffe?" verträglich. (Eine ausführliche Darstellung verschiedener Sichtweisen findet sich bei Smith & Medin [1981] und Smith [1995]. Dort wird u.a. auch die Frage nach der internen Struktur / Organisation von Konzepten diskutiert; vgl. auch Abschnitt 2.6)

Die klassische Sichtweise geht davon aus, daß alle Instanzen eines Konzepts gemeinsame Eigenschaften besitzen. Daß diese Annahme problematisch ist, hat schon Wittgenstein [1953] in den „Philosophischen Untersuchungen" am Beispiel des Begriffs „Spiel" demonstriert. Die Angabe eines Satzes von notwendigen und hinreichenden Eigenschaften ist bisher für „interessante Begriffe" des Alltagslebens, wie etwa „Spiel", nicht gelungen.

Konzepte haben ihre Bedeutung primär dadurch, daß sie Schlüsse ermöglichen: So wird z.B. ein Individuum (Instanz) über einen Begriff – bzw. die zugehörigen Klassifikationsschlüsse – klassifiziert, d.h. einer Kategorie zugeordnet. Begriffe können also verwendet werden, um neue Objekte zu alten in Beziehung zu setzen. Wenn eine Instanz zu einer Kategorie gehört, so gehen wir davon aus, daß sie über die Eigenschaften verfügt, die Individuen dieser Kategorie besitzen. Derartige Schlüsse, in der KI auch als Vererbungsschlüsse bezeichnet, seien sie strikt oder default-mäßig, sind für die menschliche Kognition von hohem Stellenwert. Um diese Funktionen zu erläutern, seien zwei Beispiele angeführt:

- Angenommen, wir sehen einen Gegenstand an einem Faden (hängend) und eine Person, die den Faden durchschneidet. Aufgrund der Konstellation („hängend") klassifizieren wir den Gegenstand als „schwerer als Luft" und schließen (=erwarten),

daß er nach dem Durchschneiden des Fadens zu Boden fallen wird. Falls ein anderer Gegenstand durch einen Faden in einer schwebenden Position festgehalten wird (Ballon), werden wir eine andere Erwartung haben.

- Ein wenig anders gelagert ist das folgende, klassische Fallbeispiel aus der KI: Angenommen, wir sehen ein zweibeiniges Tier mit Federn. Wir klassifizieren es als Vogel, und wundern uns nicht, wenn es die Flügel bewegt und davonfliegt.

Der relevante Unterschied zwischen den beiden Fällen liegt nicht in unterschiedlichen Funktionen der Konzepte, sondern darin, daß die verwendeten Schlüsse in einem Fall stets, im anderen Fall nur für „typische Fälle" gültig sind.

Die Diskussion über Konzepte, die in der kognitiven Psychologie und der Künstlichen Intelligenz in den letzten 25 Jahren eine zentrale Position eingenommen hat, soll im weiteren nur im Hinblick auf die Skizzierung einiger besonders wichtiger Erklärungsversuche nachgezeichnet werden.

Innerhalb der KI wird insbesondere der Prototypen-Ansatz von E. Rosch [1973; 1975; 1977] häufig zitiert und als motivierende und forschungsleitende Konzeption genannt. Während klassische Konzept-Theorien die Zugehörigkeit einer Instanz zu einem Konzept mit der Elementbeziehung und die Beziehung Subkonzept – Konzept mit der Inklusion gleichsetzten, geht der Prototypen-Ansatz hier subtiler vor. Insbesondere wird zwischen zentralen und peripheren Subkonzepten bzw. Instanzen einer Konzepts unterschieden. Durch zahlreiche Experimente konnte nachgewiesen werden, daß verschiedene Subkonzepte eines Konzepts als unterschiedlich typisch angesehen werden [Rosch, 1973]. Versuchspersonen, die Vogelarten auf einer Skala von 1 (sehr typisch) bis 7 (vollständig untypisch) einstufen sollten, vergaben – im Mittel – für Amsel eine 1.1 und für Huhn eine 3.8. (Aber: selbst Fledermäuse wurden nicht einstimmig mit 7 bewertet.) Rosch führte entsprechende Einschätzungsexperimente im Hinblick auf zahlreiche Konzepte und Domänen durch.

Die Bedeutung dieser Befunde liegt nun darin, daß der erfragte Typikalitätswert mit weiteren Phänomenen korreliert: In weiteren Experimenten [Rosch, 1975] konnte gezeigt werden, daß bei der Benennung von Abbildungen das Bild einer Amsel schneller als Vogel benannt wird, als das Bild eines Huhnes. Dieses ist ein Hinweis unter vielen, daß das theoretische Konstrukt der *Typikalität* auch im Hinblick auf Phänomene, die nicht direkt mit Typikalität zu tun haben, relevant ist. Im vorliegenden Fall, daß in Erbringung der Leistung von Sehen und Benennen, „typische Vertreter einer Kategorie" (hier Vogel) schneller verarbeitet werden. Experimente und Theorien zur Typikalität sind ausschlaggebend für die Strukturierung des Begriffssystems in semantischen Netzen gewesen (vgl. Abschnitt 2.6.1).

2.3.2 Schlüsse: Der Spezialfall der Deduktion

Die Sichtweise „Denken ist Problemlösen über internen Repräsentationen" weist Schlußprozessen eine zentrale Rolle zu: Das Ausgangsmaterial für derartige Prozesse sind Wissensentitäten, die aus Wahrnehmungen, aus der Kommunikation oder aus dem Vorwissen stammen, und die durch Schlußfolgerungen zu neuen Wissensentitäten führen. Im vorliegenden Abschnitt werden einige Phänomene des menschlichen Schlußfolgerns vorgestellt, wobei die Frage, in welchem Repräsentationsformat Wissensentitäten in Schlußprozessen vorliegen, detailliert erst im Abschnitt über Wissensrepräsentation (2.6) behandelt wird.

Ausgangspunkt für weitere Darstellungen und auch für viele Ansätze der Psychologie ist die Logik, die als terminologischer Rahmen für die weiteren Beschreibungen verwendet wird. Hiermit soll nicht gesagt werden, Menschen würden, sollten oder müßten stets und ausschließlich „logische Schlüsse" durchführen; im weiteren werden sogar einige Untersuchungen angeführt, die zeigen, daß Menschen durchaus „unlogisch" schließen. Andererseits ist es nicht die primäre Aufgabe der Logik, die Schlußverfahren zu untersuchen, die dem menschlichen *common sense reasoning* zugrunde liegen. Die „Aufgabenverteilung" und die Beziehungen zwischen Logik und Psychologie – insbesondere im Hinblick auf Schlußverfahren – sind ausführlich dargestellt in Macnamara [1986]. Unabhängig davon, wie die Frage nach der kognitiven Realität logischer Schluß- bzw. Argumentationsfiguren beantwortet wird, stellt die Logik den Rahmen dafür dar, die für jede Theorie des Schliessens grundlegenden Begriffe der Korrektheit und der Folgerung zu spezifizieren (vgl. [Rips, 1994; Rips, 1995].

Deduktive Schlußfiguren haben schon in der griechischen Philosophie, z.B. bei Aristoteles, eine zentrale Stellung. Gültige deduktive Inferenzen, etwa gewisse kategorische Syllogismen, stellen sicher, daß das Schlußresultat, die Konklusion, wahr ist, falls wahre Prämissen, d.h. auf die Welt zutreffende Ausgangsinformation, vorliegen. Anders formuliert: Bei Verwendung gültiger Inferenzen kann – im Inferenzprozeß – nichts schiefgehen; wenn ein falsches Resultat herauskommt, muß dies an den Eingangsinformationen liegen, d.h. eine der Prämissen muß falsch gewesen sein. Wie korrekt ist nun das menschliche Schlußverhalten? Wie gut ist das Wissen um die Gültigkeit von Schlußfiguren? Der *Modus Ponens*, die einfachste und zugleich wohl grundlegendste aussagenlogische Schlußfigur

Wenn p, dann q. Und: p. Also: q

kann ohne Zweifel als essentieller Bestandteil des menschlichen Schlußvermögens angesehen werden: Einerseits werden entsprechende Schlüsse häufig durchgeführt und andererseits werden Beispiele für derartige Schlüsse von Versuchspersonen als korrekt eingestuft (vgl. [Rips, 1994; Rips, 1995]).

Entsprechendes gilt nicht im gleichen Maße für andere Schlußfiguren, wie in verschiedenen empirischen Untersuchungen gezeigt werden konnte. Wason & Johnson-Laird [1972] legten Versuchspersonen jeweils vier Karten vor, auf denen Buchstaben oder Zahlen abgebildet waren, und zwar auf einer Seite ein Buchstabe, auf der anderen Seite eine Zahl. Die Aufgabe bestand darin, zu prüfen, ob bei vorgegeben vier Karten die folgende Regel gültig war:

Wenn auf einer Seite ein Vokal abgebildet ist, so ist auf der anderen Seite eine gerade Zahl gedruckt.

Bei der Überprüfung sollten die Versuchspersonen so wenige Karten wie möglich umdrehen. Als Beispiel sei die Konfiguration

A D 4 7

betrachtet. An dieser Stelle sollten die Leser dieses Artikels zuerst einmal selbst versuchen, diese Aufgabe spontan, d.h. ohne Rückgriff auf ihre Logikkenntnisse, zu lösen. Die Versuchspersonen wählten die folgenden Möglichkeiten des Umdrehens:

A und 4	*46%*
A	*33%*
A und 7	*4%*
andere	*17%*

Die zu überprüfende Regel kann als „Wenn Vokal, dann gerade Zahl." formuliert werden. A, ein Vokal, muß umgedreht werden, um zu prüfen, ob auf der Rückseite eine gerade Zahl steht. Wie sieht es bei der 4 aus? Wenn auf der Rückseite ein Konsonant gefunden wird, betrifft dies die Regel nicht, und wenn ein Vokal gefunden wird, ist alles in Ordnung; das heißt, es ist unnötig, die 4-Karte zu prüfen. Hingegen ist es wichtig, die 7-Karte umzudrehen, denn wenn sich auf deren Rückseite ein Vokal befindet, so wäre die zu überprüfende Regel verletzt. Also: A-Karte und 7-Karte müssen zur Prüfung gedreht werden. Der in die Kartenaufgabe involvierte Schluß ist der *Modus Tollens*:

Wenn p, dann q. Und: nicht q. Also: nicht p.

Auch in anderen Experimenten hat sich gezeigt, daß der Modus Tollens einer der „Schwachpunkte" des menschlichen Schlußfolgern ist; die Resultate des hier dargestellten Experiments weisen darauf hin, daß die Schwierigkeit im Hinblick auf den *Modus Tollens* darin liegen dürfte, daß – wie auch in anderen Fällen – eine Tendenz besteht, Konditionale (Schlüsse in einer Richtung) als Bikonditionale (Schlüsse in Hin- und Rückrichtung) aufzufassen. In weniger abstrakten Situationen, d.h. in Alltagssituationen, wird von den Versuchspersonen der *Modus Tollens* häufiger korrekt eingesetzt [Johnson-Laird und Wason, 1977]. Eine ausführliche aktuelle Untersuchung über den Einfluß der Aufgabenstellung im Hinblick auf entsprechende Problemlösungssituationen ist durch Gigerenzer & Hug [1992] durchgeführt worden.

Über einfache Schlußfiguren der Aussagenlogik hinaus wurden auch zahlreiche Untersuchungen zu prädikatenlogischen Inferenzen durchgeführt. Auch hier hat sich gezeigt, daß beim menschlichen Schlußfolgern nicht nur logisch korrekte Inferenzen durchgeführt werden. Insbesondere dann, wenn Schlußketten lang werden oder die Anzahl der möglichen Konsequenzen groß ist, werden häufiger nicht korrekte Schlüsse durchgeführt. Darüber hinaus ist zu beobachten, daß die Schlußrichtung (Vorwärtsschließen vs. Rückwärtsschließen), ein Parameter der durch die aktuelle Aufgabenstellung bestimmt ist, zu unterschiedlichen Resultaten führen kann Rips [1994; 1995].

Im Gegensatz zu Rips, der einen an deduktiven Schlußfiguren orientierten kognitionspsychologischen Ansatz, dem eine Operationalisierung des „natürlichen Schliessens" zugrunde liegt, vertritt, gehen Johnson-Laird & Byrne [1991] davon aus, daß menschliches Schlußverhalten wesentlich auf der Verwendung von Modellen – sogenannten „mentalen Modellen" – basiert. In diesen Modellen werden interne Stellvertreter (für Objekte der Realität) angenommen und die Folgerungsmöglichkeiten werden innerhalb des mentalen Modells ermittelt, obwohl strenggenommen eigentlich nur eine Beispielklasse berücksichtigt wird. Fehler treten – unter der Perspektive des Ansatzes Mentaler Modell, vgl. auch Johnson-Laird [1983] – insbesondere dann auf, wenn das untersuchte Modell nicht in der Lage ist, den allgemeinen Fall hinreichend darzustellen. Durch experimentelle Untersuchungen zum räumlichen Schliessen haben Johnson-Laird & Byrne [1991] der Konzeption der Mentalen Modelle eine umfangreiche empirische Basis gegeben.

2.3.3 Induktion, Generalisierung und Analogiebildung

Induktive Schlüsse basieren im Gegensatz zu deduktiven Schlüssen auf Einzelerfahrungen und liefern als Resultat üblicherweise Aussagen über Klassen; d.h. induktive Schlüsse sind die Basis von Generalisierungen und somit von zahlreichen Verfahren des Lernens. Aufgrund der Schlußrichtung „vom Einzelfall zum Allgemeinen" sind induktive Schlüsse nie sicher, sondern weisen stets einen hypothetischen Charakter auf. Auch wenn wir 10, 100, 1000 oder mehr schwarze Raben oder weiße Schwäne beobachtet haben, können wir nur überzeugt, nicht jedoch – in einem strikten deduktiv-logischen Sinne – sicher sein, daß alle Raben schwarz bzw. alle Schwäne weiß sind. Die philosophischen Probleme der Induktion gehören seit mehreren Jahrhunderten zu den am heftigsten diskutierten der Philosophie; vgl. etwa Swinburne [1974].

Aus kognitionswissenschaftlicher Sicht stellt sich insbesondere die Frage, aufgrund welcher und wie vieler Einzeldaten Menschen gewillt sind, Generalisierungen vorzunehmen, bzw. welcher Art diese Generalisierungen sind. Nisbett et al. [1983] haben festgestellt, daß Vorwissen über „ähnliche Situationen bzw. Konfigurationen" – von ihnen als statistisches Vorwissen bezeichnet – herangezogen wird, um induktive Schlüsse durchzuführen. Insbesondere sind Annahmen über die Analogiebeziehung dann beteiligt, wenn schon aus einer sehr geringen Anzahl von Fallbeispielen eine Generalisierung abgeleitet wird. So erhielten die Versuchspersonen des von Nisbett et al. durchgeführten Experiments Informationen der folgenden Art:

> *Stellen Sie sich vor, daß Sie als Forscher eine kleine Insel im Südpazifik entdeckt haben und dort auf bisher unbekannte Eingeborene, Tiere, Pflanzen, Mineralien und anderes stoßen. Sie beobachten (untersuchen) einige Exemplare im Hinblick auf gewisse Eigenschaften und stellen Erwartungen an, wie üblich diese Eigenschaften bei den anderen Eingeboren, Tieren, usw sind. – Nehmen Sie an, sie beobachten eine neue Vogelart, den Shreeble. Er ist blau. Wieviel Prozent der Shreebles sind ihrer Erwartung nach blau?– Warum erwarten Sie diesen Prozentsatz?*

Dieser Fragetyp wurde im Hinblick auf verschiedene „neue Arten" (u.a. einen Eingeborenstamm und ein seltenes Element) durchgeführt und zwar jeweils in den drei Varianten: ein beobachtetes Exemplar, drei beobachtete Exemplare und zwanzig beobachtete Exemplare. In einigen Konstellationen, z.B. im Hinblick auf physikalische Eigenschaften des seltenen Elements, wurde von den Versuchspersonen schon nach einem Beispiel eine Generalisierung vorgenommen. Begründungen für die frühe Generalisierung war üblicherweise von der Art „in ähnlichen Fällen, z.B. bei anderen X, ist die Eigenschaft P sehr homogen." Anders ausgedrückt: Wissen bzw. Annahmen über die beteiligten Konzepte, z.B. über die weitgehende Eigenschaftskonstanz innerhalb der Kategorie chemischer Elemente, ist für das Generalisierungsverhalten ausschlaggebend.

Analogien werden nicht nur bei der Generalisierung, d.h. im Zusammenhang von induktiven Prozessen verwendet, sondern auch als Grundlage spezifischer Schlüsse in Problemlösungen (vgl. [Duncker, 1935; Gick und Holyoak, 1980; Holyoak, 1995]). Der – bisher nur partielle verstandene – Mechanismus der Analogieschlüsse ist auch ein wesentliches Element im fallbasierten Schließen (vgl. Abschnitt 2.5.1)

2.3.4 Problemlösen als heuristische Suche

Die – durch Newell & Simon [1972] eingeführte – Sichtweise des „Problemlösens als Suche in einem Problemraum" geht davon aus, ein Problem als ein Paar, bestehend aus einem Startzustand und einem Zielzustand anzusehen; Startzustand und Zielzustand sind die Repräsentationen der Ist-Situation bzw. der Soll-Situation in einem Problemraum (problem space), der durch sämtliche Repräsentationen, die zu möglichen Situationen in der Domäne korrespondieren, gebildet wird. Ein Verfahren zur Problemlösung, das durch die Bezeichnung *means-end-analysis* bekannt geworden ist, besteht unter dieser Perspektive darin, durch geeignete Operationen eine Folge von Zuständen im Problemraum zu finden, so daß der Zielzustand durch Operatoranwendung vom Startzustand aus erreicht werden kann. Problemlösen ist somit eine – möglichst geschickte – Suche im Problemraum (vgl. auch [Holyoak, 1995]).

Diese Sichtweise läßt sich gut am strukturähnlichen Fall der Routenplanung erläutern. Gesetzt den Fall, X befinde sich in Hamburg und beabsichtige, nach Freiburg zu gelangen. Der Startzustand ist somit: X ist in Hamburg, der Zielzusstand: X ist in Freiburg. Die Operatoren – ausgewählt im Hinblick auf die InterCity-Domäne der Deutschen Bundesbahn – haben die folgende, hier nur skizzierte Wirkung:

> *Falls A und B an einer IC-Linie benachbarte Stationen sind, ist „X ist in B" ein Nachfolgezustand von „X ist in A" und umgekehrt.*

Unter Verwendung der aufgrund des IC-Netzes zulässigen Operationen kann als Problemlösung die Zustandsfolge

> *⟨X ist in Hamburg-Hbf., X ist in Hamburg-Harburg, X ist in Hannover,...,X ist in Karlsruhe, X ist in Offenburg, X ist in Freiburg⟩*

angesehen werden. Beim Versuch diese Lösung zu finden, könnten durchaus „Irrwege" aufgetreten sein, etwa Folgezustände in der falschen Richtung, etwa „über Bremen", „zurück nach Karlsruhe und dann..." Eine – häufig erfolgreiche – Heuristik besteht darin, stets zu versuchen, den Abstand zum Zielzustand – im vorliegenden Fall: zum Ziel – zu verringern. (Ein Problem mit dieser Heuristik läßt sich leicht erkennen: Auf dem Weg von Europa nach Amerika würde zuerst einmal der Landweg bis zum Amerika-nächsten Punkt des Festlandes angetreten. Dieses Problem besteht generell beim Problemlösen und ist nicht nur auf Routenplanung beschränkt.)

Die oben aufgeführte Problemlösung, die alle Zwischenstationen beinhaltet, kann durch Verwendung von Makro-Operatoren, die jeweils größere Teilprobleme betreffen, verbessert werden; im vorliegenden Fall etwa führt ein derartiges Vorgehen, das den Prinzipien der hierarchischen Problemlösung folgt (vgl. [Sacerdoti, 1977]), zu maximalen IC-Verbindungen ohne Umsteigen. Eine andere grundlegende Eigenschaft hierarchischer Problemlösungen soll an einem weiteren Beispiel aus dem Bereich der Routenplanungen skizziert werden: Die Aufgabe sei etwa, eine Reise von X's Wohnung in Hamburg zur University of California at Santa Barbara zu planen. Das Problem läßt sich in die folgenden Teilprobleme zerlegen:

1. *Plane die Anreise von der Wohnung zum Start-Flughafen.*
2. *Suche eine Flugverbindung von Deutschland nach Californien.*
3. *Plane die Anfahrt vom Ziel-Flughafen nach Santa Barbara.*

Jedes dieser Probleme kann nun in der oben skizzierten Weise gelöst werden. Es ist jedoch offensichtlich, daß die Teilprobleme nicht unabhängig voneinander sind, und daß die Lösungsreihenfolge durchaus Beachtung verdient. Hier wäre es sicherlich gut, zuerst (2) zu bearbeiten, denn die Flughafentransfers sollten erst dann geplant werden, wenn eine geeignete transatlantische Verbindung gefunden ist. Umgekehrt, können die Transferbedingungen auch die Bewertung einer Flugverbindung beeinflussen.

Das hier an Routenplanungen skizzierte Vorgehen ist auch in anderen Bereichen anwendbar; *means-ends-analysis* ist von Newell & Simon [1972] u.a. für „die Türme von Hanoi" und Aufgaben der Krypto-Arithmetik, angewandt worden. Shallice [1982] hat die Tower-of-London Aufgabe, eine zu den Türmen von Hanoi verwandte Aufgabe, bei der drei farbige Kugeln (rot, gelb und blau) auf drei Stäben unterschiedlicher Länge in eine vorgegebene Zielkonstellation gebracht werden sollen, verwendet, um die Planungs- und Problemlösungsfähigkeiten von Patienten mit unterschiedlichen Hirnschädigungen zu untersuchen (vgl. Abb. 2.1). Die Befunde von Shallice sprechen dafür, daß bei diesem Typ von Planungsaufgaben für erfolgreiches Problemlösungsverhalten einerseits gute Erinnerungsleistungen im Bereich des Arbeitsgedächtnisses benötigt werden (siehe Abschnitt 2.2.2) und andererseits die Bestimmung von Teilzielen und deren Reihenfolge für den Erfolg ausschlaggebend ist.

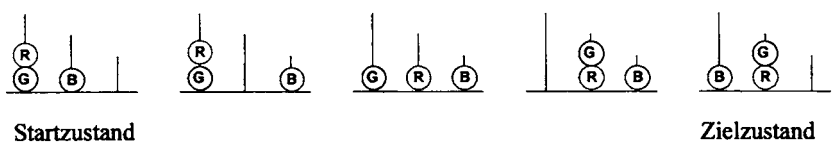


Abbildung 2.1: Tower of London: Problemlösung in vier Schritten

2.4 Aufmerksamkeit und kognitive Ressourcen

Allgemein wird die Annahme geteilt, daß unser kognitives System viele autonome und teilautonome Systeme zur Ausführung bestimmter Funktionen besitzt, die z.B. der Verarbeitung sensorischer Reize in den verschiedenen Sinnesmodalitäten dienen, der Regulation der Körperfunktionen, sowie der motorischen Steuerung beim Gehen, Greifen, Sprechen, Sehen (Okulomotorik), und daß außerdem ein zentrales, diese Teilsysteme partiell koordinierendes und überwachendes System vorhanden ist, an dessen Funktion auch das Phänomen des Bewußtseins gebunden ist (*supervisory attentional system*, oder *central executive*).

Funktionell ist die auffälligste Konsequenz dieser Architektur menschlicher Kognition die dem Handeln auferlegte Beschränkung auf grundsätzlich nur eine bewußt ausführbare Tätigkeit. Je nach Perspektive wird dies als „selektive Aufmerksamkeit" oder als „beschränkte Verarbeitungskapazität" bezeichnet. Der selektive Charakter der Aufmerksamkeit wurde schon im späten 19. Jh. bei William James hervorgehoben. Heute wird Selektion im Dienste der Handlungssteuerung von Psychologen wieder als positives und die Überlebensfähigkeit von Organismen steigerndes Merkmal beschrieben.

Traditionell entspricht dem Konstrukt der selektiven Aufmerksamkeit die Methapher des Lichtkegels, der einen Ort im Umgebungsraum erhellt. Ergebnisse der Aufmerksamkeitsforschung aus den fünfziger Jahren entsprechen dieser Ansicht. So sind wir in der Lage, auf einer Party, wo alle durcheinander reden, uns auf das zu konzentrieren, was eine bestimmte Person sagt. Oder wir können uns im Labor beim *dichotic listening* (über Kopfhörer werden links und rechts unterschiedliche Sprecher gleichzeitig abgespielt) willentlich (fast) ganz auf eine Seite konzentrieren (Cocktail-party-Phänomen, [Cherry, 1953]). Inzwischen ist aber nachgewiesen worden, daß wir bis zu sieben beliebige Objekte aus einer größeren Menge sich bewegender Objekte gleichzeitig zu verfolgen vermögen: Aufmerksamkeit ist nicht notwendig räumlich zu definieren [Pylyshyn, 1989].

Nach gegenwärtigem Stand wäre Aufmerksamkeit zu definieren als die Fähigkeit, relevante Objekte (der Außenwelt oder auch des Gedächtnisses) zu fokussieren, d.h. die mit ihnen einhergehenden Merkmale (Position wie Eigenschaften) ausgiebig zu verarbeiten zu Lasten der übrigen, nun vernachlässigten Objekte. Dabei bestimmt sich Relevanz nach den gerade verfolgten Zielsetzungen. Aufmerksamkeit pflegt also dynamisch zu wechseln.

2.4.1 Kognitive Ressourcen

Die Kehrseite fokussierter Aufmerksamkeit ist unsere Beschränktheit, mehrere Dinge gleichzeitig zu tun, wie es in hochspezialisierten technischen Umgebungen (Düsenjet) wünschenswert wäre. Hierfür hat sich die Redeweise eingebürgert, Aufmerksamkeit sei als generelle kognitive Ressource zu betrachten [Kahneman, 1973]. Da unter bestimmten Umständen (unterschiedliche Ein- und Ausgabemodalitäten, Synchronisierbarkeit) doch erfolgreiche Mehrfachtätigkeit möglich ist, wurde in der Folgezeit das Konzept zu einer Theorie multipler Ressourcen erweitert [Navon und Gopher, 1979].

Aufmerksamkeit ist nicht für alle Tätigkeiten gleichermaßen notwendig. Die Ausführung gewohnter Handlungsschemata hat keinen oder nur geringen Ressourcenbedarf; eine solche Tätigkeit wird als „automatisiert" bezeichnet [Schneider *et.al.*, 1984]. Kognitive Prozesse können durch Übung automatisiert werden, sie verbrauchen dann keine zentralen Ressourcen. Diese Ansicht paßt zu der Theorie von Anderson [1982], wonach ausgedehnte Übung zu einer „Kompilierung" kognitiver Prozesse und damit zu einer drastischen Leistungssteigerung führt.

Charakteristisch ist, daß automatisierte Handlungen an Auslösereize geknüpft sind. Mangelt es in einer Situation an Aufmerksamkeit (z.B., weil man „in Gedanken" ist), so können leicht situationsgemäße Handlungsschemata unwillentlich ausgeführt werden (*action slips:* [Norman, 1981]; *capture errors:* [Reason, 1990]). Dies ist die häufigste Quelle aller nicht auf mangelnder Kenntnis beruhenden Bedienungsfehler.

Aufmerksamkeit ist nicht die einzige kognitive Ressource. Verarbeitungszeit ist eine weitere. Auch physiologische Faktoren spielen insofern eine Rolle, als Ermüdung oder umgekehrt zu hohe, meist mit ausgeprägter Emotion (z.B. Angst) einhergehende Aktiviertheit die geistige Leistungsfähigkeit beeinträchtigen (so schon [Yerkes und Dodson, 1908]). In neuerer Zeit ist auch die schädliche Rolle von meist negativ getönter Selbstreflexion untersucht worden (*worry cognitions*, Lageorientierung: [Kuhl, 1983]). Hinzu treten strukturelle Beschränkungen, wovon die wesentlichste die knappe Kapazität des menschlichen Arbeitsgedächtnisses darstellt.

2.4.2 Beschränkte Kapazität des Arbeitsgedächtnisses

Während für die Kapazität des Langzeitgedächtnisses keine Beschränkungen bekannt sind, gehört die enge Begrenztheit des unmittelbaren Behaltens zu den auffälligsten und am besten bekannten Phänomenen menschlicher Kognition. Die klassische Arbeit stammt von Miller [1956], der darin die Kapazität zur „magischen Zahl 7 ± 2" bestimmt. Diese Zahl bezieht sich aber nicht auf bit, sondern auf bedeutungshaltige Einheiten (chunks).

Diese Einheiten können in ihrer Komplexität stark variieren. Während in einer Folge wie „g x o m k b o f" jeder einzelne Buchstabe eine Einheit darstellt, die sozusagen einen eigenen Platz im Kurzzeitgedächtnis benötigt, kann die wesentlich längere Zeichenfolge „Donaudampfschiffahrtsgesellschaft" ohne Mühe als eine Einheit gemerkt werden, wenn das Wort selbst bereits im (langzeitlich gespeicherten) Wissen vorhanden ist. Gedächtniskünstler machen von komplizierten Codierungen langer Ziffernreihen Gebrauch, wenn sie (wie bei [Ericsson, 1985]) etwa Folgen von über 80 Ziffern unmittelbar nach Vorgabe korrekt reproduzieren.

Bis zur Mitte der siebziger Jahre herrschte die Ansicht vor, daß es sich beim Kurz- und Langzeitgedächtnis um unterschiedliche Speicher handle, ähnlich wie bei CPU-internen Registern und RAM (Mehrspeichermodell, [Atkinson und Shiffrin, 1968]). Hingegen formulierten Schneider und Detweiler [1987] die Ansicht, daß die Inhalte des Kurzzeitgedächtnisses lediglich die gerade aktivierten Inhalte des Langzeitgedächtnisses seien; die Beschränkung auf ca. 7 Einheiten wird dann mit Problemen des Übersprechens (*crosstalk*) begründet, also damit, daß zuviele gleichzeitig aktivierte Inhalte einander stören.

Das bekannteste psychologische Modell des Arbeitsgedächtnisses stammt von Baddeley [1986]. Er unterscheidet ein zentrales System (*central executive*) und mehrere modalitätsspezifische Speicher, nämlich einen phonologischen Puffer (*articulatory loop*) und einen Arbeitsspeicher für bildhafte Vorstellungen (*visuo-spatial scratchpad*), der nach Logie [1995] möglicherweise in einen visuellen und einen räumlichen Speicher unterteilt werden muß. Gegenwärtig steht vor allem das zentrale System im Mittelpunkt der Forschung. Die aktuellste und beste Übersicht gibt der Sammelband von Miyake und Shah [1999].

2.5 Wissen und Expertise

Der Wissensbegriff wird nicht einheitlich gebraucht. Für Philosophen ist es (was bis zu Plato zurückverfolgt werden kann) selbstverständlich, daß Wissen mit Wahrheit zu tun hat: falsches Wissen ist keines, kann höchstens falscher Glaube (belief) sein. Wenn hingegen Psychologen vom Wissen reden, meinen sie in der Regel das, was jemand für wahr hält. So kann es vorkommen, daß in der Wissenspsychologie davon gesprochen wird, jemand habe „fehlerhaftes Wissen" erworben.

2.5.1 Problemlösen als Anwendung spezifischen Wissens

Die frühe KI war vornehmlich mit der Suche nach allgemein anwendbaren Methoden zur Lösung kognitiver Probleme befaßt. Gegen Anfang der siebziger Jahre erkannte man, in

welch starkem Maße wir selbst, wenn wir Probleme lösen, auf Wissen zurückgreifen, das gerade nicht allgemein, sondern spezifisch ist für den jeweiligen Problembereich (*domain-specific knowledge*). Die Trennung von allgemeinen Inferenzverfahren und der je spezifischen Wissensbasis, wie sie der Technologie von Expertensystemen zugrundeliegt, spiegelt diesen Umschwung wider.

Woraus besteht bereichsspezifisches Wissen? Anderson [1983] hat, ein Begriffspaar von Winograd aufgreifend, deklaratives und prozedurales Wissen unterschieden. Deklarativ ist terminologisches Wissen und das Wissen über Sachverhalte; ein Wissen, über das in der Regel auch verbal Auskunft gegeben werden kann. Im Gegensatz dazu ist prozedurales Wissen Handlungswissen, das keineswegs in allen Fällen verbalisierbar ist (oder können Sie ganz genau beschreiben, wie Sie radfahren?). Hinzu kommen bereichstypische Problemlösestrategien, die das Repertoire allgemeiner Heuristiken ergänzen oder ersetzen. Schließlich werden durch Erfahrung in einem Bereich auch typische Problemstellungen samt den zugehörigen Lösungsmustern erworben, sowie ganz konkrete, bereits früher gelöste Probleme samt Lösungsweg im Gedächtnis gespeichert. Der Abruf solcher „Fertiglösungen" oder auch die Anpassung der erinnerten Lösung für ein der aktuellen Aufgabe ähnliches Problem wird als fallbezogenes Schließen (*case-based reasoning*, [Kolodner, 1993]) bezeichnet. Notwendige Anpassungen erinnerter Lösungen scheinen nach ähnlichen Prinzipien zu erfolgen, wie sie für die Konstruktion von Analogien gelten (z.B. [Holyoak und Thagard, 1994]).

2.5.2 Expertise

Als Experte darf derjenige gelten, der über umfassendes Wissen in einem Bereich verfügt, so daß er bereichstypische Probleme auch höchsten Schwierigkeitsgrades lösen kann. Auf acht bis zehn Jahre schätzt man die Zeit, bis jemand in irgendeinem Gebiet zum Experten wird. Doch sind langjährige Erfahrung und selbst umfangreiches deklaratives Wissen nicht hinreichend für Expertise, und diese ist auch unabhängig von allgemeiner Intelligenz oder sozialem Hintergrund [Ceci und Liker, 1986]. Fortgesetzte Übung erbringt auch nach vielen Jahren noch weitere, wenn auch immer geringere Leistungssteigerungen [Newell und Rosenbloom, 1981].

Experten besitzen nicht nur ein sehr umfängliches Sachwissen, sie haben es auch besonders gut organisiert, so daß sie bei Bedarf sehr schnell darauf zugreifen können [Ericsson und Smith, 1991]. Auch die phänomenale Gedächtnisleistung von Spitzenspielern im Schach (die selbst komplizierte Stellungen nach nur 20 sec Inspektionszeit aus der Erinnerung korrekt nachstellen können) beruht auf dem Wissen über geschätzte dreißig- bis fünfzigtausend Schachpartien; von sinnlosen, d.h. in Partien nicht vorkommenden Stellungen werden auch von Experten nicht mehr als 4-5 Figuren korrekt nachgestellt [Chase und Simon, 1973]. Insbesondere sind die Organisationskriterien auf bereichsspezifische Lösungsprinzipien bezogen, die oft nichts mit der oberflächlichen Formulierung der Aufgabenstellungen zu tun haben [Chi *et.al.*, 1988]. Ferner haben Experten in hohem Maße „metakognitive" Fertigkeiten entwickelt, d.h. sie überwachen ihre eigene Tätigkeit und können recht gut einschätzen, wie nahe sie der Lösung sind [Gruber und Strube, 1989]. Daß allgemeine Intelligenz Mangel an Spezialwissen nur in recht geringem Maße kompensieren kann, hat Schmalhofer [1982] gezeigt. Und menschliche Expertise ist unterfüttert

durch allgemeines und kulturelles Weltwissen, den berühmten *common sense* – nach wie vor ein Problem für KI-Expertensysteme.

2.5.3 Wissensdiagnose und Knowledge Engineering

Für die Konstruktion wissensbasierter Systeme ist eine kognitiv adäquate Organisation der Wissensbasis ebenso entscheidend wie eine den Transfer in ein maschinelles System unterstützende Methodologie (z.B. KADS: [Schreiber *et.al.*, 1993]). Praktisch alle in der KI angewandten Methoden basieren auf mehr oder weniger standardisierten Interview- und anderen Techniken, die jeweils explizit das Wissen des Experten abzufragen suchen. Insbesondere die Interviewtechniken setzen aber voraus, daß Experten zur Introspektion in ihr eigenes Wissen und dessen Einsatz fähig sind, was gerade in Bereichen wie z.B. der medizinischen Diagnose fraglich ist.
Von der anfänglichen Annahme, Experten seien als passive Wissensquellen zu betrachten, ist man jedenfalls abgekommen; ebenso von der abwegigen Zielsetzung, es gälte, die Gesamtheit des Wissens eines Experten zu erheben. *Knowledge engineering* beschränkt sich heute auf die jeweils zu bearbeitende (routinisierte) Aufgabenstellung und faßt den Prozeß der Wissenserhebung als zwischen Experten und Systemdesignern gemeinsame Konstruktion auf [Strube 1996a; Strube 1996b].

2.6 Wissensrepräsentation und Gedächtnis

Wissen kann in verschiedenen Formen repräsentiert werden; Probleme können in unterschiedlicher Art und Weise präsentiert werden. Ob wir ein Problem lösen können, bzw. wie wir es lösen, hängt wesentlich davon ab, in welcher Form das Problem dargestellt wird.

Ein Beispiel (angelehnt an Kleene) mag dies verdeutlichen: müssen wir eine Multiplikationsaufgabe etwa der Art

43 mal 118

lösen, so befinden wir uns im Hinblick auf die zugrundeliegenden Repräsentationen – hier für Zahlen – in einem ungeheuren Vorteil gegenüber etwa den Bürgern des Römischen Reiches, für die das entsprechende arithmetische Problem durch

XLIII mal CXVIII

formuliert wurde, denn wir können in unseren gebräuchlichen Rechenverfahren die Stelligkeit der Ziffern im Dezimalsystem nutzen. Der Fortschritt der Rechenkunst ist wesentlich mit dem Fortschritt der Zahlsysteme verbunden [Krämer, 1988].

Ein weiteres Beispiel aus der Problemlöseforschung betrifft die folgende von Raphael [1976] und Wickelgren [1974] ausführlich dargestellte Situation: *Gegeben ist ein Quadrat mit einer internen 8 x 8 Einteilung in 64 quadratische Felder. Außerdem existieren rechteckige Steine (Dominosteine), die in ihrer Größe genau zwei der kleinen Quadrate überdecken. Überdecke mit 32 Steinen das Quadrat.*

Dieses Problem zu lösen, dürfte keinerlei Schwierigkeiten aufwerfen. Nun wird die Problemstellung abgewandelt: *Aus dem großen Quadrat werden nun zwei der kleinen*

Quadrate herausgeschnitten, und zwar die beiden – diagonal gegenüber liegenden – Eckquadrate in der linken oberen und der rechten unteren Ecke. Überdecke mit 31 Steinen das Quadrat. (An dieser Stelle möchten wir die Leser bitten, zuerst einmal selbst die Problemlösung zu versuchen, und erst in etwa fünf Minuten weiter zu lesen.)

Das nun vorliegende Problem erweist sich als wesentlich unangenehmer. Es ist zu erwarten, daß Versuchspersonen zuerst einmal systematische Überdeckungsversuche unternehmen (und in diesen scheitern). Die volle Problematik der Aufgabenstellung wird dann deutlich, wenn das Quadrat als Schachbrett dargestellt wird. Unter dieser Darstellungsweise ergibt sich als „zusätzliche Information", daß die beiden Quadrate, die entfernt wurden, die gleiche Farbe besitzen, z.B. beide schwarz waren. Andererseits ist offensichtlich, daß ein Dominostein stets ein weißes und ein schwarzes Feld überdeckt. Zieht man diese beiden Fakten zusätzlich in Betracht, so ergibt sich, daß nach jeder Überdeckung durch einen Stein, die Anzahl der freien weißen und freien schwarzen Felder gleichermaßen um eins reduziert ist. Da wir mit 30 schwarzen und 32 weißen Feldern beginnen, müßte – falls wir überhaupt soweit kommen – nach dem Legen des vorletzten Dominosteins eine Konstellation mit zwei weißen Feldern vorliegen. Spätestens dann muß ein weiterer Überdeckungsversuch scheitern.

Als Fazit läßt sich festhalten, daß eine adäquate Repräsentation des Problems – und entsprechendes gilt für das Vorwissen – eine wesentliche Voraussetzung für eine gute Problemlösung darstellt. Betroffen sind somit einerseits das Arbeitsgedächtnis, d.h. die Gedächtniskomponente, die aktuelle Gedächtnisentitäten für die Problemlösung bereitstellt, und andererseits das Langzeitgedächtnis, das langfristig sowohl Strategien als auch Fakten für Problemlösungen zur Verfügung halten muß (Vgl. [Jonides, 1995], zur Stellung des Arbeitsgedächtnisses beim Problemlösen und [Lehman *et.al.*, 1998], zu einer Beschreibung des Zusammenspiels von Problemlösung und Gedächtnis innerhalb der SOAR-Konzeption.)

2.6.1 Begriffliche Repräsentation

2.6.1.1 Semantische Netze

Die Beziehung „Begriff – Oberbegriff" konstituiert einen grundlegenden Typ von Begriffssystemen: Konzepthierarchien; Konzepthierarchien sind dem semantischen Gedächtnis [Tulving, 1972] zuzurechnen, d.h. dem Teilsystem des Gedächtnisses, das – im Gegensatz zum episodischen Gedächtnis – die strukturellen Beziehungen von Bedeutungen betrifft. Die hierachische Organization des Konzeptsystems legt es nahe, dieses durch baumartige Strukturen zu repräsentieren. Dementsprechend hat Quillian [1968] ein Gedächtnismodell vorgelegt, das auf zwei grundlegenden Typen von Beziehungen basiert, einer Beziehung zwischen Konzeptknoten, der *IS-A*-Beziehung, die die Ober-Konzept – Unter-Konzept – Relation betrifft, und einer Beziehung Has-Prop, die zwischen Konzepten und Eigenschaften besteht. (Vgl. hierzu Abb. 2.2; diese Abbildung, die an Darstellungen von Quillian angelehnt ist, wird im weiteren zur Erläuterung verschiedener Konzeption verwendet werden. Es sind nicht ausschließlich Knoten aufgenommen, die in den Originalarbeiten von [Collins und Quillian, 1969], verwendet wurden.) Aufgrund der netzartigen Struktur werden derartige Gedächtnismodelle als „semantische Netze" bezeichnet.

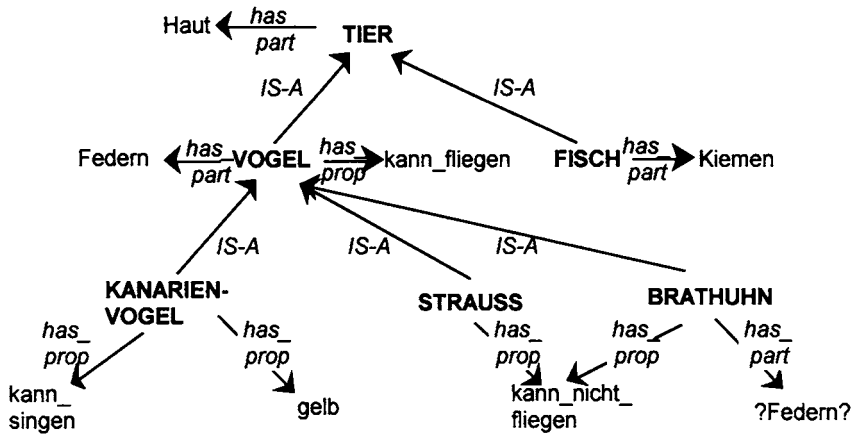


Abbildung 2.2: Semantisches Netz (Ausschnitt): Wirbeltiere, speziell Vögel

In der Folgezeit hat dieses Gedächtnismodell gerade in der Künstlichen Intelligenz großen Einfluß gehabt; es ist später insbesondere von Brachman und Mitarbeitern zu einem Wissensrepräsentationsmodell weiterentwickelt und mit einer „logischen Basis" unterlegt worden: Konzepthierarchien sind der Kern der T-Box in KL-ONE [Brachman und Schmolze, 1985]. (Zu weiteren Entwicklungen im Bereich der KL-ONE-artigen Repräsentationssysteme und insbesondere zur „logischen Fundierung" dieser Systeme vgl. Kapitel 5 dieses Buches).

Aus kognitionswissenschaftlicher Sichtweise vorangig ist die Frage nach der empirischen Evidenz für semantische Netze als Modelle des menschlichen Gedächtnisses. Collins & Quillian [1969] berichten über zwei Reaktionszeitexperimente, die die beiden Kantentypen im Netz betreffen. Zum einen wurde die Kategorienzugehörigkeit abgefragt, etwa durch die Verifikation von Aussagen der Art:

- *Ein Kanarienvogel ist ein Vogel.*
- *Ein Kanarienvogel ist ein Tier.*

Zum anderen wurde die Verifikation von Aussagen, die Eigenschaften betreffen, getestet, etwa im Hinblick auf:

- *Kanarienvögel können singen.*
- *Kanarienvögel haben Flügel.*
- *Kanarienvögel haben Haut.*
- *Kanarienvögel haben Kiemen.*

In beiden Experimenten konnte eine verlängerte Reaktionszeit nachgewiesen werden, wenn Oberkonzepte oder Eigenschaften von Oberkonzepten betroffen waren. Anders ausgedrückt: der im Modell vorhergesagte erhöhte Verarbeitungsaufwand, der durch das „Ansteigen in der Hierarchie" („Begehen von IS-A-Kanten") verursacht ist, konnte von Collins & Quillian [1969] experimentell bestätigt werden. Weitere – in der Folgezeit durchgeführte – Experimente, die verschiedene Aspekte des Basis-Modells der Semantischen Netze betrafen, haben dieses nicht in allen Punkten bestätigen können. Im weiteren sollen

einige „kritische Punkte" skizziert werden. Insbesondere im Zusammenhang mit Konzepttheorien, die das Phänomen der Prototypikalität betrafen (vgl. 2.3.1), wurde die Frage aufgeworfen, ob die Verbindungen innerhalb Semantischer Netze immer von der „gleichen Stärke" sind; da die Verifikationszeit für Rotkehlchen kürzer ist als etwa die von Huhn oder Strauß ist davon auszugehen, daß die *IS-A*-Kanten zwischen den entsprechenden Knoten unterschiedliche Eigenschaften besitzen. Rips, Shoben & Smith [1973] etwa untersuchten Reaktionszeiten im Hinblick auf unterschiedliche Unterkonzept - Oberkonzept - Ketten, z.B. zwischen (1) Hund und Tier einerseits und (2) Hund und Säugetier anderseits. Hierbei stellte sich heraus, daß durchaus kürzere Wege im Netz, (2), eine längere Verarbeitungszeit benötigen. Als Antwort auf diese Einwände wurde von Collins & Loftus [1975] eine Modifikation der Konzeption entwickelt, die durch unterschiedliche Typen von Kanten, ein Konzept der semantischen Distanz und ein Verarbeitungsmodell, das dem Prinzip der Aktivationsausbreitung genügt, gekennzeichnet ist. Dieses Modell ist jedoch nicht mehr überprüfbar. Es enthält so viele Parameter, daß es sich nahezu jeder erdenklichen Datenbasis anpassen ließe.

Auch wenn die Ansätze von Collins, Loftus und Quillian in den Details nicht als kognitiv adäquat bestätigt werden konnten, und daher in den letzten Jahren zahlreiche Alternativen und Modifikationen entwickelt wurden, besteht ihre Bedeutung nach wie vor darin, daß sie den generellen Rahmen für zahlreiche Ansätze der Psychologie und Künstliche Intelligenz gesetzt haben: Die Aktivationsausbreitungsprozesse, die Collins und Loftus [1975] beschreiben, können durchaus als - wenn auch lokalistische - Vorläufer von Prozessen über verteilten Repräsentationen aufgefaßt werden (vgl. [Hinton *et al.*, 1986]). Dieser Entwicklung folgend wird in der neueren Konzeptforschung verstärkt eine Integration perzeptueller und konzeptueller Begriffssysteme vorgeschlagen, die sich insbesondere auch in einer Kombination von regelbasierten und netzartigen Repräsentationen niederschlägt [Schyns *et.al.*, 1998; Goldstone und Barsalou, 1998; Barsalou, 1999].

2.6.1.2 Schemata und Frames

Eine wichtige, frühe Konzeption der Psychologie, die in der kognitiven Psychologie und Künstlichen Intelligenz seit Beginn der 70er Jahre eine zunehmend bedeutendere Rolle spielt ist die Schemakonzeption von Bartlett [1932]. Schon in den 30er Jahren hat Bartlett experimentell nachgewiesen, daß neue Information (neues Material) auf der Basis von im Gedächtnis vorhandenen Strukturen organisiert und erinnert wird. Derartige Strukturen werden als Schemata bezeichnet. Gegenüber „einfachen" Konzepttheorien, wie sie sich in semantischen Netzen des Collins-Quillian-Typs widerspiegeln, wird in Schemaansätzen eine bedeutend reichere interne Struktur angenommen.

Dieser Grundauffassung folgend stellte Minsky [1975] das Konzept der *frames* (Rahmen) vor. Ein *frame* ist eine komplexe Wissensstruktur, in der unterschiedliche Rollen realisiert sind. Zu den Rollen eines Konzeptes gehören an prominenter Stelle solche, die die Verweise zu Ober- und Unterkonzepten herstellen, also eine Konzepthierarchie im Sinne von Collins & Quillian aufspannen. Darüber hinaus sind auch Eigenschaften (*has-prop*-Beziehungen) über Rollen realisiert. Eine besonders einflußreiche Idee innerhalb der *frame*-Konzeption ist sicherlich das Konzept der *defaults*. Im oben (vgl. Abb 2.2) skizzierten Beispiel ist die Eigenschaft *kann-fliegen* dem Konzept Vogel zugewiesen: Was

besagt dies nun für die Subkonzepte? An diesem Beispiel kann ein zentrales Problem, bzw. Phänomen, bei der Verarbeitung begrifflichen Wissens erläutert werden:

Eigenschaften, die einem Konzept zugeordnet sind, können generell auf die Subkonzepte vererbt werden. Somit wird die Eigenschaft *hat-federn* auf alle Subkonzepte von Vogel vererbt, d.h. allen Subkonzepten kann diese Eigenschaft zugewiesen werden. Für die Eigenschaft *kann-fliegen* des Konzepts Vogel sollte jedoch anders vorgegangen werden: Sie sollte nicht – wenigstens nicht im gleichen strikten Sinne – auf alle Subkonzepte vererbt werden. *default*-Eigenschaften sind nun gerade solche, die standardmäßig angenommen und insofern vererbt werden können; wenn gegenteilige Information bekannt wird, d.h. explizit vorliegt, z.B. im Standardbeispiel für Strauße oder Pinguine, so wird eine Vererbung blockiert. (Vgl. hierzu: Nicht-monotones Schließen, Vererbungsnetze im Kapitel 7 dieses Buches.)

Die Aufnahme von Brathuhn in die in Abb. 2.2 dargestellte Konzepthierarchie macht es notwendig, auch *hat-federn* als *default*-Eigenschaft aufzufassen; wird diese Eigenschaft als strikte, nicht-*default*-Eigenschaft angesehen, so muß Brathuhn an anderer Stelle in das konzeptuelle Wissen eingebunden werden, etwa als „ex-vogel", wobei die Beziehungen zwischen Vögeln und Ex-Vögeln dann in anderer Weise zu spezifizieren sind.
Das Konzept der *frames*, d.h. der intern reich strukturierten Wissensentitäten, liegt – in der einen oder anderen Weise – den meisten gegenwärtigen Wissensrepräsentationssystemen der KI zugrunde. In derartigen taxonomischen Systemen wird der Aspekt der Standardannahmen (der *default*-Eigenschaften) häufig nicht berücksichtigt, insbesondere deswegen, weil eine Fundierung auf der Prädikatenlogik 1. Stufe für nicht-monotone Verfahren nicht konfliktfrei möglich ist (vgl. auch [Minsky, 1975; Minsky, 1981] und [Nebel, 1990]. Die beiden hier angesprochenen Aufsätze von Minsky – 1975, 1981 – gehen auf ein gleichnamiges MIT-Memo aus dem Jahr 1974 zurück, für dessen externe Veröffentlichung unterschiedliche Zusammenstellungen ausgewählt wurden.)

2.6.2 Repräsentation von Ereignissen und Texten: Propositionen und Scripts

Wenn in Konzepttheorien von „begrifflichem Wissen" die Rede ist, dann ist hiermit normalerweise „Wissen über Objekt-Kategorien" gemeint. Über „Objektwissen" hinaus spielt die Verarbeitung von Wissen über Situationen (diese übergreifende Bezeichnung wird im vorliegenden Abschnitt für Ereignisse, Prozesse, Aktionen u.a. verwendet, ohne daß hier die unterschiedlichen Subtypen näher erläutert werden sollen) eine besondere Rolle. Unter Verwendung des Situations-Konzepts können die folgenden Phänomen- / Problembereiche angegangen werden:

- Über Objekt-Konzepte hinaus können Situations-Konzepte untersucht werden. So werden etwa von Miller und Johnson-Laird [1976] detaillierte Analysen für Bewegungs- und Besitzwechsel-Situationen vorgelegt. Insbesondere wird deutlich, daß auch – entsprechen zu den Objekt-Konzepten – im Bereich der Situationen Konzepthierarchien anzunehmen sind [Morris und Murphy, 1990].
- Auf der Ebene der Situations-Konzepte bzw. der Instantierung durch einzelne Situationen ist die Repräsentation von individuellen „Erfahrungen" und somit die Modellierung des episodischen Gedächtnisses [Tulving, 1972] möglich.

- Die Repräsentation von Episoden ermöglicht es, die Bedeutung von Texten, die Ereignisabläufe in der Welt betreffen, zu erschließen und andererseits Episoden sprachlich zu beschreiben. Auf dieser Basis ist eine Untersuchung von Prozessen des Textverstehens und der Textproduktion möglich [Habel und Tappe, 1999].

Im weiteren werden einige – ebenfalls, wie im Fall der Objektbegriffe – schemaorientierte Ansätze, die die Repräsentation von Situationen und Text(-Bedeutungen) betreffen, vorgestellt: scripts aber auch story-grammars wurden für die Untersuchung von Textverstehensprozessen entwickelt und darüberhinaus auch zur Beschreibung weiterer kognitiver Aufgabenstellungen eingesetzt.

Die in der KI – aber auch der Kognitionswissenschaft im allgemeinen – einflußreichste Konzeption in diesem Themenbereich betrifft *scripts*, die durch Schank & Abelson [1977] ausführlich dargestellt sind. Ausgangspunkt für *scripts* ist die Feststellung, daß bei der Beschreibung der Welt (Situationen in der Welt) vermittels natürlicher Sprache von Vorwissen über den „normalen Ablauf der Ereignisse" intensiv Gebrauch gemacht wird: es ist für die Textproduzentin nicht notwendig alle Details eines Ereignisses zu beschreiben, da der Textrezipient aufgrund von Erfahrungen viele Details durch Standardannahmen erschließen kann. Entsprechendes Vorwissen liegt in *scripts* vor, die in ihrem inneren Aufbau an „Drehbüchern" orientiert sind; ein Beispiel soll dies verdeutlichen (wir verzichten hier auf das Standardbeispiel des *restaurant-scripts*, das wohlbeschrieben in zahlreichen Darstellungen veröffentlicht wurde):

Name:	*Wissenschaftlicher Vortrag*
Inventar:	*Tische, Stühle*
	Projektor, Leinwand
	Folien, Folienstifte
Rollen:	*Vortragender, Zuhörer*
Voraussetzungen:	*Vortragender ist vorbereitet*
Ergebnis:	*Zuhörer haben Neues erfahren*
Szene 1:	*Zuhörer betreten Vortragsraum.*
	Zuhörer suchen Platz und setzen sich.
	Vortragender betritt Raum.
	Vortragender geht zum Projektor.
	Vortragender legt Folien und Stifte bereit.
	Vortragender schaltet Projektor ein.
Szene 2:	*Vortragender begrüßt Zuhörer.*
	Vortragender legt Folie auf.
	Vortragender spricht.
	(Wiederholung möglich!)
	Vortragender beendet Vortrag mit Dank.
Szene 3:	*Zuhörer stellt Frage, gibt Kommentar.*
	Vortragender antwortet.
	(Wiederholung möglich!)
	Vortragender verabschiedet sich.
Szene 4:	*Vortragender verläßt Raum*
	Zuhörer verlassen Vortragsraum.

Das hier skizzierte *script* betrifft einen speziellen Typ von Vortrag (in der Terminologie von Schank und Abelson: *track*); so ist etwa bei „Kolloquiums- oder Festvorträgen" eine eigene (Kurz-)Szene 2.a „Begrüßung des Vortragenden" anzusetzen, in anderen Disziplinen wird auf Folien verzichtet, Diaprojektoren oder die Tafel treten an diese Stelle.

Was können nun derartige *scripts* leisten? Zuerst einmal geben sie den Rahmen des Üblichen vor; Ereignisse, die dem *script* folgen, werden vom Hörer erwartet und müssen daher nicht in allen Details verbalisiert werden. So wäre ein Bericht, der jeden Folienwechsel erwähnt als überausführlich zu bezeichnen. Stattdessen werden Darstellungen der Art „Sie ist wieder einmal in hohem Tempo durch ihren Folienstapel gegangen." gegeben. Ebenso werden Szene 1 und Szene 4 normalerweise nicht behandelt, solange wenigstens nicht, solange keine Probleme, also nichts unerwartetes passiert: „Der Vortragsraum war bis zwei Minuten vor Beginn abgeschlossen." Eine wesentliche Aufgabe von *scripts* besteht darin, aus dem Text eine kohärente Bedeutungsstruktur aufzubauen. Beginnt eine Vortragsbeschreibung mit:

Der Vortrag begann eine viertel Stunde zu spät. Erst war der Strom ausgefallen, und dann stellte sich heraus, daß die Birne durchgebrannt war,

so wird der Bedeutungszusammenhang erst dann klar, wenn die kausalen Beziehungen, die zwischen den Einzelaussagen bestehen, aufgedeckt sind. Der verspätete Beginn und der Stromausfall können auf der Grundlage des *scripts* und zusätzlichen Wissens über Projektoren etc. (in Form von *frame*-artigen Strukturen) als eine kausale Begründung aufgefaßt werden. Die durchgebrannte Birne ist ebenfalls nur auf der Grundlage des *scripts* als eindeutig beschreibbares Objekt – und daher durch einen bestimmten Artikel ausgezeichnet – identifizierbar.

Über die hier skizzierte Funktion von *scripts* in Prozessen der Sprachverarbeitung hinaus sind sie auch im Bereich der Planung einsetzbar. Schank & Abelson [1977] und Abelson [1981] weisen darauf hin, daß das Wissen über den normalen Ablauf von Ereignissen dazu verwendet wird, in vielen Fällen anstelle von Planungsprozessen einen Abruf von Plänen bzw. Planschemata vorzunehmen. Eine derartige Verwendung von Wissenseinheiten setzt ein Gedächtnismodell voraus, das dynamischer ist, als es das ursprüngliche *script*-Konzept von Schank & Abelson [1977] war. Von Schank [1982] vorgestellte Weiterentwicklungen, u.a. MOPs (memory organization packets) und TOPs (thematic organization points) sind grundlegend für die Ansätze des fallbasierten Schließens [Kolodner, 1993].

Die durch Beispiele in der oben auch verwendeten Art motivierte kognitive Plausibilität konnte in Experimenten bestätigt werden: Bower, Black & Turner [1979] berichten über eine Serien von Untersuchungen, die dem Nachweis der kognitiven Realität von *scripts* galten. Versuchspersonen, die aufgefordert wurden, einen typischen Restaurant-Besuch (oder andere Standardsituationen) zu beschreiben, produzierten Texte, die den von Schank und Abelson [1977] angenommenen *scripts* weitgehend entsprachen. Wurden in Texten Szenen oder Ereignisse aus Szenen nicht explizit erwähnt, so wurden einige von ihnen mit großer Häufigkeit bei einer „Nacherzählung" trotzdem produziert. Traten in Texten Reihenfolgeabweichungen gegenüber dem *script* auf, so wurden diese später gut erinnert bzw. bei Nacherzählungen zum Teil „repariert". Diese Resultate zeigen eine hohe Übereinstimmung zu Bartlett's [1932] Untersuchungen zu Schemata.

Ungefähr zeitgleich zur Entwicklung der *script*-Konzeption wurde insbesondere durch Rumelhart [1977] die Konzeption der *story grammar* ausgearbeitet; am Beispiel von meist mündlich überlieferten Märchen wurde dafür argumentiert, daß auf der Textebene, d.h. jenseits der Satzgrenze/-ebene, regelmäßige linguistische Strukturen vorliegen, die beim Verstehen und Produzieren von Texten verwendet werden. So werden etwa von Rumelhart Ersetzungsregeln der Art :

Ereignis -> Ereignis | Ereignis {und / dann / deswegen } Ereignis

vorgeschlagen. Der zweite Teil dieser Regel besagt, daß komplexe Ereignisse als Summation von Ereignissen, als zeitliche Aufeinanderfolge von Ereignissen oder als kausale Kette von Ereignissen aufgefaßt und beschrieben werden können. Hierdurch wird deutlich, daß Geschichten-Grammatiken nicht nur die Sprachverarbeitung betreffen, sondern generell Repräsentation und Verarbeitung von Wissen. Die Basisidee der Geschichten-Grammatiken kann in analoger Weise auch für andere kommunikativ verwendete externe Repräsentationen nutzbar gemacht werden, z.B. für Skizzen (vgl. [Habel und Tappe, 1999]).

Abschließend soll kurz auf die interne Struktur der bisher in diesem Abschnitt vorgestellten Repräsentationen eingegangen werden: es handelt sich um „propositionale Repräsentationen", also solche, die eine Operator-Operanden-Struktur aufweisen; vgl. auch Habel [1986]. Neben der linearen Darstellungsweise, die an LISP-Ausdrücke bzw. Ausdrücke der Logik angelehnt ist, etwa bei Rumelhart [1977], van Dijk & Kintsch [1983] und den Repräsentationen im Rahmen von Beschreibungslogiken, finden sich – formal äquivalente – netzartige Beschreibungen als semantische Netze für Bedeutungsrepräsentationen durch Collins & Quillian [1969] oder Brachman & Schmolze [1985] oder auch als *conceptual dependency nets* [Schank, 1972]. Obwohl „formale Äquivalenz" auf der Repräsentationssebene vorliegt, erklären netzartige Darstellungsformate auf der Prozeßebene Assoziationsphänomene (vgl. hierzu insbesondere [Collins und Quillian, 1969] und [Anderson, 1983]).

2.6.3 Bildhafte (analoge) Repräsentationen

Wenn die Leserin oder der Leser dieses Kapitels zwei Stunden nach einem – z.B. dem im letzten Abschnitt skriptmäßig behandelten – Vortrag gefragt wird, wie die / der Vortragende gekleidet war, wird meistens eine recht zutreffende Beschreibung der Kleidung erfolgen, Art, Stil und Farbe von Anzug, Rock, Pullover oder Hemd werden bei entsprechenden Fragen noch gut erinnert. Die Antwortenden geben – auf Nachfrage – meist an, die Antwort aufgrund einer bildhaften Vorstellung bzw. Erinnerung gegeben zu haben.

Phänomene der hier skizzierten Art haben dazu geführt, daß über die Existenz bildhafter Vorstellungen, auch als Mentale Bilder bezeichnet, in den letzten Jahren wieder intensiv diskutiert wurde. Die unter der Bezeichnung *imagery debate* [Block, 1981] geführte Diskussion betrifft die Frage, ob mentale Bilder existieren, d.h. kognitiv real sind, oder ob es sich bei ihnen um „Epiphänomene" handelt. Für den Bereich der Künstlichen Intelligenz und Kognitionspsychologie stellt sich diese Frage in leicht veränderter Form: „Sind mentale Bilder eine geeignete Form zur Repräsentation räumlichen Wissens?"

Bevor hier die Frage nach der Existenz mentaler Bilder näher diskutiert wird, seien die beiden kontroversen Richtungen innerhalb der imagery debate skizzenhaft gegenübergestellt: Die Deskriptionalisten, als Hauptvertreter kann hier Pylyshyn [1981] gelten, gehen von der Existenz eines – und zwar propositionalen – Repräsentationsformats aus. Im Gegensatz hierzu nehmen die Depiktionalisten, zB. Kosslyn [1980] zwei Repräsentationsformate an, ein propositionales und ein depiktionales/bildhaftes (auch als depiktives bezeichnet). Wie diese Gegenüberstellung zeigt, gehen beide Ausrichtungen von propositionalen Repräsentationen aus, die unterschiedlichen Einstellungen betreffen im wesentlichen die Frage, ob und in welchen Fällen, zusätzlich auch nicht-propositionale Repräsentationen verwendet werden. Kombinationen von propositionalen und depiktionalen Repräsentationen werden schon seit langem im Hinblick auf konzeptuelles Wissen untersucht; entsprechende Ansätze werden – Paivio folgend – der *dual coding theory* zugeordnet (vgl. [Paivio, 1986]). Die Annahme der Existenz nicht-propositionaler, perzeptionsnaher Repräsentationen in kognitiven Prozessen bzw, in mentalen Modellen, bildet – vgl. hierzu Kosslyn [1980; 1994], Habel [1988; 1998], Barsalou [1999] eine Brücke zwischen Kognition und Perzeption.

Evidenz für die Existenz bildhafter Repräsentationsformate sind u.a. die Experimente zur mentalen Rotation, die insbesondere von Shepard, Metzler und Cooper durchgeführt wurden [Metzler und Shepard, 1974; Shepard und Cooper, 1982]. Shepard und Metzler forderten Versuchspersonen auf, ein Paar von nacheinander präsentierten Objekten – Buchstaben oder Ziffern, die entweder normal oder gespiegelt präsentiert wurden (vgl. Abb. 2.3) – daraufhin zu prüfen, ob sie durch Rotation ineinander überführbar sind. Anders ausgedrückt: Sind zwei Darstellungen Exemplare desselben Objekts, die durch Rotation auseinander hervorgehen, oder handelt es sich um unterschiedliche Objekte. Die Versuchspersonen sahen nacheinander zwei Stimuli mit der Aufgabe, die Überführbarkeit zu entscheiden. In anderen Versuchsreihen wurden Strichzeichnungen drei-dimensionaler Objekte präsentiert (Abb. 2.3, rechts).

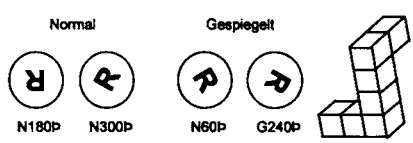


Abbildung 2.3: Mentale Rotation: Stimuli entsprechend den Experimenten von Metzler & Shepard [1974] [N = Normal, G = Gespiegelt] (links) bzw. Shepard & Cooper [1982] (rechts)

Gemessen wurde die Zeit, die für die Entscheidung benötigt wurde. Unterschieden wurden drei Typen von Stimulus-Paaren (für Buchstabenexperimente der in Abb. 2.3 dargestellten Art entfällt der Typ der „Rotation im Raum"):

- Überführbar durch Rotation in der Ebene (2D)
- Überführbar durch Rotation im Raum (3D)
- Nicht ineinander überführbar

Für 2D- und 3D-Rotation ergab sich hierbei, daß die Antwortzeit (abgesehen von einer Basisverzögerung, die der Überprüfung identischer Objekte entspricht) proportional zum Rotationswinkel ist. Dieses Phänomen spricht, dafür, daß Repräsentationen verwendet

werden, in denen Winkel analog repräsentiert werden. Propositionale Repräsentationen würden entsprechende Resultate schwerlich beschreiben und erklären.

Kosslyn [1980] hat in zahlreichen kognitionspsychologischen Experimenten die Eigenschaften Mentaler Bilder untersucht; insbesondere konnte er Evidenz dafür finden, daß derartige depiktive Repräsentationen räumliche Eigenschaften besitzen. So konnten Belege für einen „Mentalen Sehwinkel" gefunden werden; ebenso konnte gezeigt werden, daß mentale Bilder im Zentrum eine bessere Auflösung als in der Peripherie besitzen. Aus diesen und weiteren neuropsychologischen Befunden herrscht mittlerweile weitgehend Konsens, daß es eine Verarbeitungsebene gibt, in der bildhafte Wahrnehmung und bildhafte Vorstellung gemeinsame Prinzipien und Verarbeitungsmechanismen besitzen; eine ausführliche Zusammenfassung der kognitionspsychologischen und neurowissenschaftlichen Forschungen zu Mentalen Bildern gibt Kosslyn [1994].

Bildhafte Repräsentationen sind ein spezieller Typ analoger Repräsentationen. Das Attribut analog bezieht sich hierbei auf die Abbildung zwischen Repräsentiertem und Repräsentierendem (vgl. [Palmer, 1978]): Im Fall Mentaler Bilder etwa ist die intrinsische Räumlichkeit der Repräsentationen Voraussetzung für eine Analogiebeziehung. Die Experimente zur mentalen Rotation zeigen – wie oben erläutert – daß der „mentale Drehwinkel" linear proportional in die Verarbeitungszeit eingeht. Somit ist davon auszugehen – so Shepard und Metzler – daß in der Repräsentation ein Analogon zum realen Drehwinkel existiert.
Trotz dieser überzeugenden Evidenz für die Annahme bildhafter bzw. räumlicher Repräsentationen sind gegenwärtig einige zentrale Fragen in Hinsicht auf diesen Typ von Repräsentationen weiterhin offen:

- In welcher Hinsicht sind mentale Bilder „bildhaft" in welcher Weise sind sie „räumlich"?
- In welcher Weise interagieren propositionale und bildhafte/räumliche Repräsentationen in komplexen, multimodalen Systemen. Diese Frage betrifft u.a. die Aufgabenverteilung und Integration durch die Zentrale Exekutive des Baddeleyschen Modells des Arbeitsgedächtnisses (vgl. 2.4.2)?
- In welcher Weise kann in bildhaften/räumlichen Repräsentationen Unterbestimmtheit repräsentiert und verarbeitet werden (vgl. [Habel, 1998])?

2.7 Sprache

Die Fähigkeit zur Verwendung von Sprache gehört zu den herausragenen Eigenschaften menschlicher Kognition. Die Erforschung menschlicher Sprache und des menschlichen Sprachverhaltens ist Gegenstand aller kognitionswissenschaftlichen Einzeldisziplinen. Die Kognitionswissenschaft der Sprache wird oft als „Psycholinguistik" bezeichnet und umfasst auch Erkenntnisse aus den Neuro- und Computerwissenschaften (Computationale Psycholinguistik).

Das Hauptproblem der Sprachverarbeitung ist das hohe Maß an Ambiguität auf allen Ebenen: Wörter können mehrere Bedeutungen haben (z.B. „Bank"), mehreren syntaktischen Kategorien angehören (z.B. „sein"), morphologisch unterschiedlich dekomponiert werden (Stau-becken vs. Staub-ecken). Bei gesprochenem Text ist der Grad an Mehrdeutigkeit

um einiges höher, da hier noch Probleme der Segmentierung und der phonologischen Enkodierung hinzukommen. Jenseits der Wortebene sind es vor allem strukturelle, in der Grammatik begründete Ambiguitäten, in zunehmenden Maße aber auch semantische (z.B. Bedeutung und Skopus von Quantifizierern wie „manche", „nur wenige", „alle" etc. oder referenzielle Ambiguitäten; ein Überblick findet sich in [Frazier, 1999]) die Gegenstand psycholinguistischer Forschung sind. Eine Klassifikation findet sich bei Konieczny et al. [2000]. Da sich Ambiguitäten im Verlauf der Verarbeitung vom Anfang zum Ende eines Satzes bzw. Textes multiplizieren, stehen wir vor einem hochkomplexen Suchproblem. Interessanterweise hängt die beobachtbare Verarbeitungszeit bei Menschen in der Regel nicht vom Grad der Ambiguität im Stimulusmaterial ab und nimmt auch zum Ende eines Satzes nur unwesentlich oder gar nicht zu, manchmal sogar ab. Gleichzeitig werden Ambiguitäten in den seltensten Fällen bewusst. Auf der anderen Seite ist wiederholt gezeigt worden, dass wir Sätze hochgradig inkrementell (d.h. Wort für Wort) so verarbeiten, dass jedes neue Wort sofort im Kontext der vorangegangen syntaktisch analysiert und interpretiert wird (z.B. [Hemforth, 1993]). Offensichtlich entscheiden wir uns im Falle von auftretenden Ambiguitäten sehr früh für eine, oder nur eine geringe Auswahl der möglichen Alternativen. Die Tatsache, dass uns Ambiguitäten nur selten bewusst werden, nämlich meist dann, wenn sich im späteren Verlauf der Verarbeitung herausstellt, dass wir uns falsch entschieden haben (der sog. „Holzweg" oder „Garden-path" Effekt), deutet darauf hin, dass wir uns in den weitaus meisten Fällen richtig entscheiden. Eine der Hauptfragen der Psycholinguistik lautet daher: Welche Prinzipien oder Mechanismen leiten die frühzeitige Auflösung, oder Einschränkung, von Ambiguitäten? Die Beantwortung dieser Frage kann nur auf der Basis genauer Annahmen über das zugrunde liegende Wissen und die Verfahren seiner Nutzung erbracht werden.

Während wir auf der einen Seite über äußerst effiziente Mechanismen zur Verarbeitung ambigen Materials verfügen, scheitern wir gelegentlich an eindeutigen, aber zu komplexen Sätzen. Ein gängiges Beispiel sind Zentraleinbettungen:

Der Hund, der die Katze, die die Maus gefangen hatte, jagte, stolperte.

Hier sind drei Sätze so ineinander verschachtelt, dass die Verben, die die Sätze abschließen, erst nach dem jeweils eingebetteten Relativsatz verarbeitet werden können.

Erklärungen für dieses Phänomen basieren meist auf der Annahme, dass die Kapazität des menschlichen Arbeitsgedächnisses überschritten wird. Jedoch unterscheiden sich die Ansätze darin, welche Konstrukte sie für modellierungsrelevant halten. Die Vorschäge reichen vom Stapeln von Phrasenstrukturregeln [Yngve, 1960], über lokalitäts- oder entfernungsbasierte Ansätze [Hawkins, 1994; Gibson *et al.*, 1996], bis hin zu interferenzbasierten Modellen [Lewis, 1998]. Alle Modelle beinhalten genaue Metriken zur Berechnung und Vorhersage der lokalen Verarbeitungskomplexität, die sich in psycholinguistischen on-line Experimenten überprüfen lassen.

2.7.1 Psycholinguistische Methoden

Psycholinguistische Daten werden entweder off-line, d.h. im Anschluss an die vollständige Verarbeitung eines Textes, oder on-line, d.h. während der Verarbeitung, erhoben. On-line Techniken haben den Vorteil, dass sie detaillierte Verarbeitungsprofile ermöglichen, die weit über die globale Beurteilung eines Satzes hinausgehen. Off-line Techniken hingegen

sind in ihrer Anwendung wesentlich weniger aufwendig und werden oft in Normierungsstudien eingesetzt oder um einen ersten Anhaltspunkt über globale Präferenzen zu erhalten.

Typische off-line Techniken sind Fragebögen mit Akzeptabilitätsurteilen, Satzvervollständigung oder einer multiple-choice Aufgabe. Typische on-line Techniken sind das selbst-getaktete Lesen (self-paced reading: Versuchpersonen drücken eine Taste um einen neuen Textabschnitt lesen zu können) sowie die Aufzeichnung von Blickbewegungen während des Lesens (z.B. [Rayner und Sereno, 1994]). Mit beiden Techniken werden Lesezeiten für jeden einzelnen Textabschnitt ermittelt, die den aktuellen kognitiven Aufwand widerspiegeln. In jüngerer Zeit hat sich eine weitere Klasse von Techniken als nützlich erwiesen: die Aufzeichnung von Potential- und/oder Magnetfeldveränderungen an der Schädeloberfläche während des Lesens oder Hörens von Texten. Mit Hilfe solcher Aufzeichnungen ist es möglich geworden, verschiedene Prozessklassen (wie syntaktische versus semantische Verarbeitung) in den Daten qualitativ zu unterscheiden (positive Shifts in bestimmten Zeitfenstern reflektieren eher syntaktische, negative eher semantische Prozesse), und sie in verschiedenen Arealen des Gehirns zu lokalisieren (z.B. [Friederici und Mecklinger, 1996; Hagoort *et.al.*, 1993; Osterhout und Holcomb, 1992]). Darüber hinaus werden zunehmend Korpus-Daten herangezogen, um Hinweise auf etwaige Präferenzen zu bekommen (z.B. [Gibson *et al.*, 1996; Uszkoreit *et al.*, 1998]).

Da jedes kognitive Modell letztlich auf die neurophysiologischen Gegebenheiten des Gehirns abbildbar sein muß, werden auch aus der Neurolinguistik wesentliche Rahmenbedingungen gesetzt. Gerade die Untersuchung der Beeinträchtigungen des Sprachverhaltens bei hirnorganischen Funktionsstörungen ermöglicht Aussagen über die „normale" Funktionsweise des Systems [Martin *et al.*, 1994; Caplan und Waters, 1999].

Die Funktion der Sprache schließlich, als Vermittlerin zwischen dem Menschen und seiner Umwelt, das Verhältnis also von Denken, Sprache und Wirklichkeit wird eingehend in der Philosophie diskutiert [Grice, 1975; Searle, 1969]. Der Rolle der Sprache im menschlichen Handeln oder als wirklichkeitskonstituierende Strukturierungshilfe bleibt auch für die Entwicklung von Modellen menschlicher Sprachverarbeitung wichtig [Hörmann, 1976].

2.7.2 Syntax, Weltwissen und Diskurs

2.7.2.1 Die Rolle der Syntax

Die Rolle der Syntax für die Verarbeitung natürlicher Sprache wird bis heute kontrovers diskutiert. Vielfach wurde angenommen, daß syntaktische Verarbeitung nur als letzter Ausweg eingesetzt wird, wenn eine semantische Analyse nicht ausreicht [Bever, 1970; Herrmann, 1985; Schank und Riesbeck, 1981]. *Solche Syntax-as-last-resort*-Ansätze werden heute aus den verschiedensten Gründen kaum noch vertreten. So konnte beispielsweise [Flores d'Arcais, 1987] in psycholinguistischen Experimenten zeigen, daß eindeutig ungrammatische Sätze mehr Verarbeitungszeit benötigen als grammatische, selbst wenn die syntaktischen Verletzungen nicht bemerkt werden (s. auch [Hemforth, 1993; Konieczny, 1996]). Dies läßt sich nur erklären, wenn in jedem, auch im einfachsten Fall eine syntaktische Analyse vorgenommen wird.

Kontraintuitiv scheint es auch, daß Sätze wie (1a,b) wirklich für identisch gehaltenwerden, wie der *SALR*-Ansatz vorhersagen würde. Der Satz

(1a) *The red tomato ate the little boy.*

kann nach einigem Überlegen schon so verstanden werden, daß es wohl der Junge sein wird, der die Tomate ißt. Es ist jedoch auch offensichtlich, daß (1a) dann korrekterweise die Form von (1b) haben müßte [Pulman, 1986].

(1b) *The little boy ate the red tomato.*

Wäre es nun tatsächlich so, daß wir Sprache nahezu ausschließlich aufgrund unserer Kenntnis von der Welt verstehen, könnten wir nichts ausdrücken, was wir nicht schon wissen. Das Verhältnis von Struktur und Bedeutung ermöglicht es erst, kreative, neue Dinge über die Welt oder fiktive Welten zu sagen (in denen beispielsweise Tomaten kleine Jungen verspeisen) (vgl. [Chomsky, 1981]).

Obwohl die psychologische Realität von Syntax heute selten angezweifelt wird, besteht wenig Einigkeit in der Frage, wie syntaktisches Wissen kognitiv repräsentiert ist. Die meisten Psycholinguisten folgen der von Chomsky [1965] nahegelegten Trennung von Kompetenz und Performanz, die sich im einfachsten Fall in einer Trennung von deklarativem und prozeduralem Wissen niederschlägt. Auf der Seite der linguistischen Theorie (Kompetenz) wird meist auf das Prinzipien-und-Parameter-Framework (Chomsky, [1986], in jüngster Zeit auch das Minimalistische Programm, Chomsky, [1995]) zurückgegriffen. Während die LFG (*Lexical Functional Grammar*, [Bresnan und Kaplan, 1982]) in den Achtzigern auch aus der Motivation entstanden ist, kognitiv adäquate Verarbeitung der Grammatik direkt, d.h. ohne Zwischenschritte, die die Grammatik in ein verarbeitbarares Format überführen, zu ermöglichen (im Sinne der *strong competence hypothesis*, [Bresnan und Kaplan, 1982]), spielt sie heute in psycholinguistisch motivierten Ansätzen kaum noch eine Rolle. Vereinzelt sind auch Modelle auf der Basis von Kategorialgrammatik (*Combinatory Categorial Grammar*, [Ades und Steedman, 1982]), der GPSG (*Generalized Phrase Structure Grammar*, z.B. von [Ford und Dalrymple, 1988]) oder der HPSG (*Head-driven Phrase Structure Grammar*, [Pollard und Sag, 1994], hier existiert ein Modell von [Konieczny, 1996]), vorgeschlagen worden.

Bis auf wenige Ausnahmen übernehmen die meisten Ansätze eher eine schwache Kompetenzannahme. Das verwendete sprachliche Wissen entspricht demnach den Beschränkungen der jeweiligen linguistischen Theorie; es liegt jedoch in einem kognitive leichter verarbeitbaren Format vor. Hier sind vor allem die perzeptuellen Strategien [Bever, 1970], aber auch aus lexikalisierten Grammatiken kompilierte und meist generalisierte Phrasenstrukturregeln zu erwähnen.

Diesen Ansätzen, die meist in der Linguistik begründet sind, stehen eher psychologisch oder computational motivierte Modelle gegenüber, in denen die Trennung von Kompetenz und Performanz vollständig aufgehoben ist. Hier ist vor allem die Klasse der probablistischen Modelle zu nennen, zu denen auch lernfähige konnektionistische Modelle gehören. Im Extremfall, etwa in einer rekurrenten Netzwerkarchitektur [Elman, 1991] erwirbt ein System durch bloße assoziative Verkettung von Wörtern in Trainingskorpora implizites Wissen über syntaktische und semantische Kategorien. Konnektionistische Systeme heben auch die funktionale Trennung von Wissen und Arbeitsgedächtnis auf. Verarbeitungschwierigkeiten bei komplexem eindeutigem Material können hier auf die geringere Regularität in Korpora zurückgeführt werden; interindividuelle Unterschiede in der Gedächtniskapazität werden als Unterschiede in der Erfahrenheit begriffen (MacDonald und Christiansen, eingereicht).

2.7.2.2 Weltwissen und Diskurs

Crain & Steedman [1985] zeigten anhand von Sätzen wie (2a,b), dass *Holzweg*-Sätze leichter zu verstehen sind, wenn durch das Weltwissen die weniger präferierte Lesart nahegelegt wird.

(2a) *The teachers taught by the Berlitz method passed the test.*
(2b) *The children taught by the Berlitz method passed the test.*

Das Wissen darüber, dass Kinder eher lernen als lehren, erleichtert es z.B in (2b), die Phrase *taught by the Berlitz method* als Relativsatz („who were taught...") zu verstehen. In (2a) hingegen fehlt ein solcher Hinweis des Weltwissens, und in der Tat bleiben Leser länger bei der zunächst präferierten Hauptverb-Lesart „Die Lehrer lehrten...".

Deutlicher wird der Einfluß allgemeinen Wissens noch, wenn nicht nur isolierte Sätze, sondern komplexere Texte (3) betrachtet werden.

(3) *Eine Frau ging in ein Restaurant. Der Kellner führte sie zu ihrem Tisch. Sie bestellte Shrimps und eine Flasche Champagner. Als sie ging, half der Kellner ihr in den Mantel.*

Bemerkenswerter ist bei solchen Texten, die syntaktisch und semantisch keinerlei Probleme bereiten dürften, daß sie von Lesern auch als völlig kohärent betrachtet werden. Obwohl hier weder etwas davon zu lesen ist, daß die Frau das Bestellte verzehrt und auch bezahlt, wird dies vom Leser automatisch als für einen Restaurantbesuch typisch inferiert. Dies geht so weit, daß schon relativ bald nach Lesen des Textes nicht mehr zu unterscheiden ist, welche Informationen tatsächlich gegeben wurden und welche aus dem Situationsskript [Schank und Abelson, 1977; Lehnert, 1982] abgeleitet wurden. Es ist also notwendig für das Verstehen von Texten, Wissen über typische Situationsabläufe zu integrieren (vgl. Abschnitt 2.6.2).

2.7.3 Die Architektur des menschlichen Sprachverarbeitungssystems

Eine der Hauptfragen für eine Theorie der Sprachverarbeitung betrifft das Zusammenspiel oder die Interaktion aller Wissensquellen. Hier lassen sich zunächst zwei Extrempole ausmachen:

- die *modulare* Auffassung des Sprachverarbeitungssystems (SVS), nach der zumindest die syntaktische Analyse als autonomes Modul [Fodor, 1983] von der Verarbeitung weiterer Information abgekoppelt ist. Dabei wird angenommen, daß sie der Verarbeitung auf höherer Ebene zeitlich und logisch vorgeschaltet ist und sich von keinerlei Information nachgeschalteter Prozesse beeinflussen läßt.
- die *interaktive* Auffassung, gemäß der Information aller Ebenen direkt oder indirekt die Verarbeitung auf jeder anderen Ebene leiten kann.

Vertreter der Auffassung, dass die Syntax weitgehend autonom ist, sind neben Fodor [1983] vor allem Frazier [1987] und Kollegen, die mit ihrem Verarbeitungsmodell, der sogenannten *sausage machine*, und später mit ihrem *Garden-path*-Modell der Syntaxanalyse den Vorrang vor der Verarbeitung auf höheren Ebenen gibt.

Interaktive Modelle variieren enorm in der Art und Weise der angenommenen Interaktion, d.h. im angenommenen Informationsfluß zwischen den verschiedenen Ebenen der Verarbeitung. Marslen-Wilson und Tyler [1980; 1987] haben in sog. *Shadowing*-Experimenten gezeigt, dass Versuchpersonen, die einen Text noch während des Hörens nach- oder mitsprechen sollten, nur wenige zehntel Sekunden hinter dem Original herhinkten und dabei Korrekturen von gezielt eingestreuten kleinen Fehlern vornahmen (*compsiny* statt *company*). Solche schnellen „konzeptgesteuerten" Verbesserungen deuten auf einen extrem schnellen Informationsaustausch zwischen allen Ebenen der Verarbeitung hin, wie er vor allem von interaktiven Modellen vorhergesagt wird (Auf der anderen Seite kann ein modulares System nicht allein durch Betrachtung der Geschwindigkeit des Informationsaustausches widerlegt werden).

Zu den wichtigsten interaktiven Modellen zählen die von Just und Carpenter [1992], Juliano & Tanenhaus [1994] sowie MacDonald et al. [1994]. Meist handelt es sich dabei um konnektionistische oder hybride Modelle, bei denen multiple Constraints zu jedem Zeitpunkt miteinander im Wettstreit stehen. Widersprechen sich die Constraints verschiedener Ebenen etwa bei der Auflösung einer Ambiguität, wird mehr Zeit für die Auflösung benötigt.

2.7.3.1 Der Klassiker: Das Holzweg-Modell

Frazier [1987; 1987] postuliert Verarbeitungsprinzipien, die sich ausschließlich an strukturellen Eigenschaften des Satzes orientieren. Ein Satz wird inkrementell von links nach rechts so verarbeitet, daß jedes zu verarbeitende Wort nach den Regeln der Grammatik in die bis dahin aufgebaute Satzstruktur integriert wird. Erlaubt die Grammatik mehr als eine Integrationsmöglichkeit, wird die Ambiguität auf der Basis der Prinzipien „Minimal Attachment" (MA) und „Late Closuer" (LC) aufgelöst.

(p1) *minimal attachment principle:*
Postuliere keine Knoten, die sich später möglicherweise als unnötig erweisen.

(p2) *late closure* [Frazier, 1987]:
Ordne nach Möglichkeit jedes neue Item der Phrase zu,
die momentan verarbeitet wird.

Wird beispielsweise der Satz *The spy saw the cop with binoculars* gelesen, so wird nach den ersten Worten *The spy* und *saw* die Struktur [S [NP *The spy*] [VP saw... aufgebaut. Die Anbindung des nun folgenden Artikels *the* kann aber auf zweierlei Weise geschehen: Erstens, indem eine Nominalphrase postuliert wird, die sich nach der Regel „VP → V NP PP" in dieVerbalphrase integrieren läßt, oder aber indem nach der Regel „NP → NP PP" zunächst eine komplexe Nominalphrase postuliert wird, die erst dann in die Verbalphrase integriert werden kann. MA entscheidet hier zugunsten des Aufbaus der sparsameren Struktur, so daß später die Präpositionalphrase *with binoculars* unmittelbar in die Verbalphrase integriert werden muß, da für die Anbindung an die vorangegangene NP die bisherige Struktur revidiert und erst ein zusätzlicher NP-Knoten eingefügt werden müßte.
Scheitert im weiteren Verlauf die gewählte Analyse, etwa weil sie sich als unplausibel erweist, wird eine Reanalyse eingeleitet, in der die komplexere, aber in diesem Fall plausiblere Struktur aufgebaut wird. Das *Garden-Path*-Modell sagt also für die Verarbeitung

von Sätzen mit letztlich nicht minimaler Struktur einen meßbar höheren Verarbeitungsaufwand vorher, der auf eine erzwungene Reanalyse zurückzuführen ist.

Obwohl sich Minimal Attachment in einer Vielzahl von Experimenten bestätigt gefunden hat, sind in letzter Zeit vermehrt Gegenbefunde vorgelegt worden (z.B. [Konieczny *et al.*, 1997]). Frazier und Clifton [1996] haben im Rahmen ihrer „Construal Theory" jüngst den Wirkunsbereich struktureller Prinzipien wie MA stark eingeschränkt.

2.7.4 Universalität menschlicher Sprachverarbeitungsprinzipien

Prinzipien wie *Minimal Attachment* gelten als universell; d.h. es wird postuliert, dass sie für Sprecher/Hörer beliebiger Sprachen gültig sind. Diese Annahme wurde in den letzten Jahren insbesondere aufgrund von Evidenzen aus sprachvergleichenden Studien angezweifelt. Scheinbar abweichend von den universellen Vorhersagen *des Late-closure* Prinzips (p2) und Befunden aus dem Englischen findet sich bei Relativsätzen (4) in vielen Sprachen (so im Spanischen, im Französischen, im Niederländischen und im Deutschen) eine Präferenz zur „hohen" Anbindung, d.h. zur Anbindung an den Kopf der komplexen NP (*Dienerin*).

(4) *Jemand erschoß die Dienerin der Schauspielerin, die auf dem Balkon war.*

Diese und ähnliche Befunde stehen im Widerspruch zu einer universellen Erklärung menschlicher Verarbeitungspräferenzen [Mitchell, 1994]. Die beobachtbaren Unterschiede zwischen verschiedenen Sprachen werden häufig auf die statistische Verteilung der Strukturen in der jeweiligen Sprache bzw. in der individuellen Lerngeschichte der einzelnen Sprecher/Hörer zurückgeführt (p3).

(p3) *tuning hypothesis* [Mitchell, 1994]:
Individuelle Strategien entwickeln sich, weil sie in der Vergangenheit häufiger erfolgreich waren. Strategien sind *exposure based.*

Da sich die sprachspezifischen Befunde bisher auf fakultative Satzelemente beschränken, wird aus universeller Perspektive die Besonderheit von Relativsätzen und ähnlich modifizierenden Elementen (Adjunkten) in den Blickpunkt gerückt. So wird in der Nachfolgetheorie des *Garden-path*-Modells, der *Construal Theory* [Frazier und Clifton, 1996; Gilboy *et al.*, 1995], angenommen, daß nur primäre, thematisch lizensierte (Argument-)Relationen auf der Basis universeller Prinzipien innerhalb des Syntaxmoduls angebunden werden. Die Anbindung von Adjunkten folgt dagegen allgemeineren semantisch/konzeptuellen oder pragmatischen Prinzipien. Das Gricesche Prinzip der Klarheit könnte beispielsweise die Unterschiede zwischen dem Englischen und vielen anderen Sprachen erklären: Da die hohe Anbindung im Englischen durch Verwendung des sächsischen *Genitivs the actress's servant* auch eindeutig ausgedrückt werden könnte, sollte hier im ambigen Fall *the servant of the actress* die Anbindung an die „tiefe" Modifier-NP präferiert werden. Die Unterschiede zwischen den Verarbeitungspräferenzen in verschiedenen Sprachen ließen sich so auch im Rahmen eines universellen Modells auf der Basis einzelsprachlicher Besonderheiten erklären.

2.7.5 Der Aufbau semantischer Repräsentationen beim Textverstehen

Die Funktion des Sprachverstehens ist die Abbildung einer sprachlichen Äußerung auf ihre Bedeutung. Wie aber könnte eine solche Bedeutungsrepräsentation tatsächlich aussehen? Auch für diese Frage gibt es bisher sicher keine eindeutige Lösung. In der Sprachpsychologie ist wohl der Versuch am bedeutsamsten, die Satzbedeutung durch Propositionen zu repräsentieren, also als Prädikat-Argument-Strukturen [Allen, 1987; Chomsky, 1965; Kintsch und Dijk, 1978; Kintsch, 1997]. Einem Satz wie (5) könnten beispielsweise die drei Propositionen (a,b,c) zugrundeliegen.

(5) *Der sorgsame Hausmann gießt den rosa Weihnachtsstern.*

a) *sorgsam (Hausmann)*
b) *rosa (Weihnachtsstern)*
c) *gießen (Hausmann, Weihnachtsstern)*

Es konnte gezeigt werden, daß sich die Anzahl an Propositionen, die ein Satz enthält, im Verarbeitungsaufwand niederschlägt [Kintsch und Dijk, 1978]. Ebenso fanden McKoon & Ratcliff [1981] die Rolle propositionaler Netzwerke als Repräsentation für Texte darin bestätigt, dass eher propositionale Nähe als Oberflächennähe einen Einfluss auf Verarbeitungsprozesse aufweist.

2.7.6 Die Rolle der computationalen Psycholinguistik

Die empirische Entscheidung zwischen verschiedenen Verarbeitungsmodellen und den Repräsentationsmöglichkeiten für die notwendigen Informationsquellen ist bis heute ausgesprochen problematisch.

Zum einen ist die Integration eines Kenntnissystems in ein Verarbeitungssystem keinesfalls eindeutig. Dies führt beispielsweise dazu, daß der Output eines auf einer bestimmten Grammatik basierenden Parsers von vielen Faktoren bestimmt wird, die nicht der Grammatik inhärent sind. Darüber hinaus sind die Konsequenzen von Verarbeitungsoperationen für beobachtbares sprachliches Verhalten nicht immer klar. Es ist unklar, wieviel kognitiven Aufwand die verschiedenen Operationen erfordern, bzw. ob sie sich überhaupt beobachtbar niederschlagen.
Nur eine detaillierte Spezifikation und Formalisierung der verschiedenen Kenntnissysteme und der darauf basierenden Verarbeitungsprinzipien, sowie eine konsequente empirische Prüfung der Systemparameter kann einen Ausweg aus diesem Dilemma anzeigen. Erst der interdisziplinäre Zugang zur Sprache im Rahmen der Kognitionswissenschaft ermöglicht eine solche Herangehensweise.

Ohnehin scheint ein tiefes Verständnis der Sprache ohne Berücksichtigung des kognitiven Verabeitungsapparates nicht möglich. Sprache und kognitive Architektur haben sich im Laufe der Evolution im Einklang miteinander entwickelt und aufeinander abgestimmt. Ohne eine Kenntnis der kognitiven Beschränkungen des menschlichen Sprachverarbeitungssystems bleiben viele Aspekte der Sprache daher ein Rätsel (vgl. [Hawkins, 1994]).

Einen guten allgemeinen Überblick über psycholinguistische Modelle des Satzverstehens bietet Mitchell [1994], speziell für Prozesse der Satzverarbeitung im Deutschen Hemforth und Konieczny [2000].

2.8 Kognitionswissenschaft

Wir alle haben einen intuitiven Begriff davon, was „Kognition" bedeutet, nämlich zu denken und Probleme zu lösen, sprachlich zu kommunizieren und die Welt um uns herum zu erkennen. Wissenschaftlich ist der Begriff kaum besser definiert. Eine besonders enge Definition beschränkt Kognition auf das bewußte Denken, eine sicherlich zu weite Definition ist die Formulierung von Maturana und Varela [1980], alles Lebendige sei bereits kognitiv. Überhaupt ist festzuhalten, daß „Kognition" erst in der Psychologie des 19. Jahrhunderts zum Fachbegriff wird (abgeleitet aus dem lateinischen *cognoscere* = erkennen, vgl. [Prinz, 1976]).

Die moderne Kognitionswissenschaft ist eine gerade 25 Jahre alte Disziplin, erwachsen aus einer Konvergenz der Forschungsinteressen in unterschiedlichen Fächern und gefördert durch ein Programm der Sloan Foundation 1975 [1978], das nach Inhalt und Bezeichnung (nämlich *Cognitive Science*) ihren Beginn markiert. Damals fanden sich Linguisten, Philosophen, Computerwissenschaftler, Psychologen, Anthropologen und Neurowissenschaftler zusammen, um Methoden und Konzepte bei der Erforschung des menschlichen Geistes zu diskutieren und nach einem gemeinsamen theoretischen Fundament zu suchen. Doch gilt, daß diese „Mutterdisziplinen" der Kognitionswissenschaft, die sich zum Teil beträchtlich mit ihr überlappen, ihren spezifischen Forschungsstil zumeist beibehalten; dennoch sind sie soweit aufeinander bezogen, daß die jeweils praktizierten Methoden sich gegenseitig ergänzen [Habel *et.al.*, 1990]. In diesem Sinn wird auch von „den Kognitionswissenschaften" gesprochen [Wilson und Keil, 1999].

Die *cognitive science*, also die Kognitionswissenschaft (im Singular!) hat als ihre theoretische Grundannahme formuliert, daß kognitive Prozesse als Berechnungsvorgänge zu charakterisieren seien, die auf mentalen Repräsentationen als Datenstrukturen operieren [Gardner, 1990]. Insofern ist die Kognitionswissenschaft die der KI besonders nahestehende unter denjenigen Disziplinen, deren Forschungsgegenstand (auch) Kognition ist. Für eine ausführliche Darstellung wird auf Strube [Strube 1996a] verwiesen.

2.8.1 Grundlegende Annahmen der Kognitionswissenschaft

In den meisten Bereichen der Kognitionswissenschaft werden bei der Entwicklung und Überprüfung theoretischer Annahmen und Konzepte Computersysteme zu Hilfe genommen. Legitimiert wird diese computerunterstützte Modellbildung über kognitive Prozesse durch die zentrale Annahme, daß Mensch und Computer in vergleichbarer Weise Informationen speichern und verarbeiten (jedenfalls, wenn man beide auf der für die Erforschung der kognitiven Prozesse relevanten Beschreibungsebene betrachtet). Demzufolge können Computersimulationen zu Einzelaspekten der menschlichen Kognition über deren strukturelle Besonderheiten Aufschluß geben. Alle kognitiven Prozesse werden in der Kognitionsforschung als informationsverarbeitende Prozesse begriffen.

Die verbindende Grundannahme in der Kognitionsforschung kann demnach als Informationsverarbeitungsparadigma bezeichnet werden. Diesem Ansatz sind (nach [Stillings *et al.*, 1995]) folgende Prämissen zuzuordnen:

- informationsverarbeitende Prozesse lassen sich als formale Prozesse beschreiben, also als Algorithmen, die auf mentalen Repräsentationen operieren.
- Kognition läßt sich unabhängig von der Ebene ihrer materiellen Basis betrachten (d.h. auf einer höheren, funktionalen Ebene, der Wissensebene: [Newell, 1982; Marr, 1982]); doch ist die Relation zur neuronalen „Implementierung" Hauptgegenstand der kognitiven Neurowissenschaft.
- Kognitionswissenschaft ist als Basisdisziplin aufzufassen: es können grundlegende Prinzipien des Geistes eruiert werden, die für unterschiedlichste kognitive Vorgänge gleichermaßen gelten.

Das Informationsverarbeitungsparadigma ist auch das gemeinsame Fundament der Kognitionswissenschaft und der kognitionswissenschaftlich orientierten KI.

Der Symbolverarbeitungsansatz ist von Newell und Simon [1976; 1980] als „physical symbol systems hypothesis" expliziert worden. Kognitive Prozesse sind Transformationen von Symbolstrukturen. Symbolstrukturen wiederum sind aus elementaren Symbolen als den bedeutungstragenden Einheiten (Symbole stehen für etwas in der Welt) gemäß syntaktischer Regeln zusammengesetzt. Die Symbole müssen selbst in einer Trägermaterie codiert sein, z.B. Bitmuster in Computerspeichern oder Aktivitätsmuster von Neuronenverbänden. Damit wird das Symbolsystem zum materiell verankerten, eben zum „physical symbol system". Die These von Newell und Simon besagt nun, daß ein derartiges System, gleich welche Materie es zur Darstellung seiner Symbole benutzt, notwendige und hinreichende Voraussetzung für intelligentes Verhalten ist. Dabei ist das „hinreichend" lediglich prinzipiell gemeint, Newell [1980] bemerkt, daß auch ein gewisses Mindestmaß an Komplexität des Systems erforderlich ist.

Allgemeiner Konsens ist es, daß Berechnung notwendig ist; hingegen bleibt umstritten, ob kognitive Prozesse restlos als solche der (symbolischen) Informationsverarbeitung erklärt werden können. Auch das konnektionistische Berechnungsparadigma, also künstliche neuronale Netze, ist nach einigen Jahren zum Teil heftiger Auseinandersetzung [Fodor und Pylyshyn, 1988; Rumelhart und McClelland, 1986; Smolensky, 1988] in die Kognitionswissenschaft integriert worden.

2.8.2 Philosophische Grundlagen

Insbesondere im Kontext der Symbolverarbeitungshypothese läßt sich ein philosophischer Hintergrund für die Kognitionsforschung identifizieren: Die sogenannte Computertheorie des Geistes. Sie geht auf den von Putnam und Fodor vertretenen Funktionalismus zurück, demzufolge das Mentale als irreduzibler Realitätsbereich aufgefaßt werden kann. Geistige Prozesse können also unabhängig von der Basis ihrer materiellen Realisierung betrachtet werden.

Zentral für die computationale Theorie des Geistes ist das Konzept der mentalen Repräsentation. Mentale Repräsentationen werden syntaktisch strukturierten Symbolen gleichgesetzt - es sind also logische Entitäten (Propositionen), zu denen ein Organismus bzw. ein System in einer bestimmten intentionalen Beziehung steht: ein Subjekt

S weiß (glaubt, hofft, wünscht), daß P, (etwa daß der Himmel blau ist). Mentale Repräsentationen sind demnach propositionale Einstellungen.

Drei zentrale Konzepte, die schon Chomsky [Chomsky 1959b] für menschliche Sprache formuliert hat, charakterisieren menschliche Kognition:

- Produktivität: Wir sind dazu in der Lage, aus einer begrenzten (kleinen) Anzahl einfacher Elemente eine unbegrenzte Anzahl an mentalen Ausdrücken zu generieren und zu verstehen.
- Systematizität: Wenn wir einen Begriff in einem Satz verstanden haben, so verstehen wir ihn auch in einer großen Zahl ähnlich konfigurierter Sätze; d.h. wenn wir den Satz „John liebt Mary" verstehen, dann verstehen wir auch die Sätze „Mary liebt Klaus" oder „John wird von Anna geliebt".
- Kompositionalität: Die Bedeutung komplexer mentaler Ausdrücke kann als Funktion der Bedeutung der einzelnen Bestandteile dieses Ausdrucks und ihrer syntaktischen Relation aufgefaßt werden; d.h. die formalen syntaktischen Eigenschaften von Repräsentationen sowie die Bedeutung der atomaren Einheiten determinieren die semantische Struktur eines Ausdrucks.

Das Konzept kompositionaler mentaler Repräsentationen mündet bei Fodor [1975] in die Vorstellung einer „Sprache des Geistes". Diese ist charakterisiert durch die Existenz elementarer psychophysischer Bedeutungsträger, aus denen sich die Semantik mentaler Repräsentationen und damit auch der kognitiven Prozesse ableiten läßt. Diese innerhalb der Kognitionswissenschaft bereits klassische Sicht hat selbst innerhalb der analytischen Philosophie des Geistes Kritik erfahren [Kemmerling, 1991]; weit stärker haben indes die Thesen von der Situiertheit und sozialen Determiniertheit der Kognition (s.u.) die Fodorschen Vorstellungen heute in den Hintergrund treten lassen.

Ein weiteres, ebenfalls schwieriges und kontrovers diskutiertes Problem ist das Verhältnis von Kognition und Bewußtsein. Searle [1992] geht so weit zu behaupten, es gäbe lediglich Bewußtsein und die im Gehirn ablaufenden neurophysiologischen Prozesse – aber nichts sonst, insbesondere keine Informationsverarbeitung. Die überwiegende Meinung sieht jedoch das Bewußtsein als ein Subsystem der (menschlichen) Kognition an, das auf der Grundlage seiner Repräsentation der Welt und des eigenen, darin handelnden Selbst zum Medium der Reflexion und Planung wird [Dennett, 1991; Metzinger, 1993].

2.8.3 Die „neue" KI und Kognitionswissenschaft

Das alte theoretische Leitbild in beiden Disziplinen reduzierte den Forschungsbereich vornehmlich auf Denkvorgänge in individuellen und von ihrer Umwelt isolierten kognitiven Systemen; bei Fodor [1980] wurde dies sogar als „methodologischer Solipsismus" propagiert. Kurze Zeit später ereignete sich eine Wende.

Kritiker der alten Ansätze hatten von jeher betont, daß menschliche Kognition im Dienste des Handelns steht, also situationsbezogen ist. Dies zeigt sich beispielsweise beim Planen einer Fahrtroute: Anstatt alles bis ins Kleinste voraus zu planen, machen wir in erheblichem Maße von je situativ verfügbaren Hinweisen (z.B. Wegweiser) Gebrauch. Auch in der technisierten Arbeitswelt konnte Ähnliches beobachtet werden [Suchman, 1987]. Daraus resultierte die Forderung, diese Situiertheit der Kognition (die gleichwohl bereits bei Simon, [1969], thematisiert worden war) ernst zu nehmen:

- die Analyse der System-Umwelt-Beziehung auch bezüglich unmittelbarer „Reaktivität" (d.h. nicht kognitiver, aber adaptiver Prozesse) samt deren Interaktion mit kognitiven Prozessen; dies hat seine Parallele in der Robotik [Brooks, 1991],
- die Bedeutung externer Repräsentationen (wie Notizen, Wegweiser, Landkarten oder Markierungen) für menschliches Problemlösen [Zhang, 1997],
- die soziale Genese von Wissen und soziale Prozesse, die dessen Gebrauch bestimmen (insbesondere stehen hier *shared knowledge*, also gemeinsam geteiltes Wissen, und die Kooperation zwischen Systemen, die über komplementäres Wissen verfügen, im Vordergrund); ein Beispiel stellt das *knowledge engineering* dar [Strube 1996a], ein anderes das kooperative Lernen [Plötzner, 1998],
- die Analyse von kognitiven Systemen, die nicht aus einem einzigen Organismus oder Roboter bestehen, sondern aus mehreren, einschließlich gemischter Mensch-Maschine-Systeme (z.B. ein Cockpit: [Hutchins, 1995; Hutchins, 1995]); hier kommt neuerdings in Gestalt der Sozionik [Malsch *et.al.*, 1996] auch die Soziologie ins Spiel,
- die Bedeutung der Leiblichkeit, sowohl phänomenologisch [Becker, 1998], als auch biologisch in der *cognitive neuroscience* (einführend: [Kandel *et.al.*, 1995]).

Für ausführlichere Darstellungen sei auf Strube [in press, in Druck a] verwiesen.

2.8.4 Zur Methodologie der Kognitionswissenschaft

Die Methoden innerhalb der Kognitionswissenschaft entstammen den Disziplinen, aus denen sie entstanden ist. Dies erklärt die Breite und Heterogenität der angewandten Verfahren. Wenigstens drei Gruppen von Methoden lassen sich unterscheiden:

- Theoretische Analyse kognitiver Funktionsbereiche und Aufgabenstellungen mit dem Ziel der Formalisierung, sowie der Bestimmung von Regularitäten und grundlegenden Beschränkungen (Methoden der Geistes- und Formalwissenschaften, z.B. Philosophie, Logik und Linguistik),
- Empirische Untersuchung der Organisation der Informationsverarbeitung bei Mensch und Tier, sowie der biologischen Grundlagen natürlicher kognitiver Systeme (Methoden der Naturwissenschaften, z.B. Psychologie und Neurowissenschaften, also die strenge Methode des naturwissenschaftlichen Experiments, nicht etwa irgendwelches „Herumexperimentieren"),
- Modellierung kognitiver Prozesse durch Konstruktion virtueller kognitiver Maschinen, also Computerprogrammen, mit den Mitteln der Informatik (Methoden der Ingenieurwissenschaften, z.B. im Bereich Künstliche Intelligenz).

Die Ergebnisse, die über einen Inhaltsbereich (etwa „Sprache") mit diesen Methoden gewonnen werden können, müssen für eine kognitionswissenschaftliche Betrachtung aufeinander bezogen werden. Dies ist ein anspruchsvolles Ziel, das mehr die Ausrichtung der Arbeit anzeigt, als daß es für gewöhnlich erreicht würde. Zumindest in den empirisch ausgerichteten Fachgebieten ist jedoch insofern eine gewisse Einigkeit festzustellen, als die Computersimulation eine vorrangige Rolle bei der Theoriebildung spielt.

In der Tat kann die kognitive Modellierung (Strube, in Druck b) insofern als die kognitionswissenschaftliche Methodologie angesehen werden, als sie nicht einfach Computersimulation ist, sondern auf theoretischen Analysen und bekannten empirischen Re-

sultaten aufbauend zur Modellierung kognitiver Prozesse fortschreitet und das Modell durch weitere empirische Untersuchungen überprüft.

2.8.4.1 Theoretische Analyse kognitiver Funktionsbereiche

Hier kann die Kognitionswissenschaft auf die reiche Tradition der Philosophie, insbesondere der Erkenntnistheorie, zurückgreifen. Hinzu kommen Denkrichtungen, die vom Hauptstrom abendländischen Denkens abweichen. Hinzu kommen ferner die spezielleren Disziplinen, die sich im Laufe der letzten Jahrhunderte aus der Philosophie ausgegliedert haben: (theoretische) Psychologie, (philosophische und mathematische) Logik sowie die Formalwissenschaften generell, von denen als eine kognitionswissenschaftlich besonders bedeutsame die theoretische Linguistik hier erwähnt sei. Diese Wissenschaften haben wesentliche Beiträge zur inhaltlichen und formalen Analyse von Produkten der Kognition geleistet: von sprachlichen Äußerungen und sozialer Kommunikation, von Kunstwerken und anderen kulturellen Leistungen, von Erkenntnisprozessen selbst und von Argumentationsstrukturen. Infolgedessen dürfen Untersuchungen menschlicher Sprachverarbeitung die von der Sprachwissenschaft beschriebenen Phänomene und Theorien, insbesondere aber linguistische Methoden zur Überprüfung sprachbezogener Modellvorstellungen nicht ignorieren, ohne kognitionswissenschaftlich zu kurz zu greifen. Entsprechend muß die Konstruktion einer Konzepthierarchie für ein wissensbasiertes System, wenn sie kognitionswissenschaftliche fundiert sein soll, auf philosophische Untersuchungen zur Ontologie und auf psychologische Untersuchungen (s.u.) gestützt sein. Es fehlt hier der Platz, entsprechende Methoden auch nur in Auswahl vorzustellen. Auch existiert kein Kanon dessen, was ein Kognitionswissenschaftler unbedingt an Methoden beherrschen sollte (wenn auch wohl Konsens besteht, daß eine Mindestbeherrschung formaler Logik und Grundkenntnisse in der formalen Linguistik unverzichtbar sind). Ein möglichst breiter Hintergrund und die Integration solcher inhaltlichen und methodischen Kenntnisse mit informatischen und experimentell-naturwissenschaftlichen Methoden (s. die folgenden Abschnitte) ist für die Kognitionswissenschaft charakteristisch.

2.8.4.2 Die Methodologie empirischer Kognitionsforschung

Von Systemen der künstlichen Intelligenz wissen wir genau, wie sie funktionieren, oder kennen, wo das nicht mehr praktikabel ist (bei lernenden Systemen, konnektionistischen Netzen usw.), wenigstens die Prinzipien, nach denen diese Systeme arbeiten. Natürliche kognitive Systeme hingegen wollen erst in ihrer Funktionsweise aufgeklärt sein. Dies ist ein äußerst mühsames Unterfangen; genau genommen ist es sogar unmöglich, denn immer existiert eine unendliche Menge von Strukturen und auf ihnen arbeitenden Prozessen, die jeweils dasselbe Systemverhalten hervorbringen können. Braitenberg [1984] nennt dies *„the law of downhill synthesis and uphill analysis"*: Analyse ist Sisyphusarbeit, mühsam und nie zu Ende zu bringen. Modellierung (also Synthese) vermeidet dieses Problem, muß aber durch Analyse immer wieder an der Empirie – hier vor allem an der menschlichen Kognition – gemessen werden. Die Methode des Experiments ist der Weg, hier in kleinen Schritten bergauf zu gehen.

Der Grundgedanke des Experiments ist der, das Verhalten eines Systems unter zwei (oder mehreren) Bedingungen zu vergleichen, wobei sich die Bedingungen lediglich in

einer kleinen Variation der Aufgabenstellung unterscheiden und alle übrigen Randbedingungen konstant gehalten werden. Nur dadurch wird der Experimentator in die Lage versetzt, Veränderungen des Systemverhaltens eindeutig auf die experimentelle Variation zurückzuführen und somit zu Kausalerklärungen zu gelangen. Diesen einfachen Grundgedanken so zu realisieren, daß mögliche Fehler (von denen es viele gibt) vermieden werden, ist schwierig genug, so daß die entsprechende Ausbildung einen großen Teil des Studiums in den empirisch arbeitenden Mutterdisziplinen der Kognitionswissenschaft, vor allem Psychologie und Neurowissenschaften, ausmacht. Um unbeabsichtigte Bedingungsvariationen auszuschalten (z.B. Effekte der Reihenfolge verschiedener Aufgabenstellungen) bedarf es einer ausgeklügelten Versuchsplanung und Materialerstellung. Um systematische Verhaltensunterschiede von bloß zufälligen Schwankungen abzugrenzen, muß unter Umständen erheblicher Aufwand an statistischer Analyse betrieben werden. Außerdem unterscheiden sich Individuen zuweilen beträchtlich voneinander (z.B. in der Geschwindigkeit, mit der sie eine bestimmte Aufgabe erledigen), so daß allgemein-artspezifische Effekte von interindividuellen Differenzen getrennt werden müssen. Für die experimentell arbeitende Psychologie sei hier auf Erdfelder [1996], sowie auf Irtel [1993] und dort angeführte Literatur verwiesen.

Die kognitive Psychologie verwendet vor allem zwei Arten von Verhaltensmessungen: Reaktionszeiten (Ausführungszeiten) und Fehlerraten. Wo beide Maße eine Rolle spielen, muß zudem ihr Zusammenhang berücksichtigt werden (speed-accuracy trade-off: eine Aufgabe kann langsamer und genauer oder schneller und weniger genau bearbeitet werden). Die Logik des Reaktionszeit-Experiments wurde schon von Donders (1868) beschrieben: Man entwerfe zwei Aufgaben, die sich nur dadurch unterscheiden, daß die eine einen zusätzlichen Verarbeitungsschritt benötigt. Der Mehrverbrauch an Zeit, so Donders, ist dann dem zusätzlichen Verarbeitungsschritt zuzuordnen. (Dies setzt natürlich Verfahren der Aufgabenanalyse voraus; außerdem liegt dem Konzept der Verarbeitungsschritte das Verarbeitungsmodell einer sequentiellen Architektur zugrunde. Die heute verwendeten Methoden sind weitaus komplizierter.)

Ein Beispiel hier darzustellen ist auf dem vorgegebenen knappen Raum unmöglich. Ein berühmtes, knapp gefaßtes, leicht zugängliches und ohne Vorkenntnisse (empfohlen wird allerdings die Lektüre von Abschnitt 2.6.3) verständliches Reaktionszeit-Experiment ist das von Shepard und Metzler [1971] zur mentalen Rotation.

Ein weiterer empirischer Zugang ist die Untersuchung des Verhaltens gestörter Systeme. In den Neurowissenschaften, vor allem der Neuropsychologie, wird versucht, die mit bestimmten Schädigungen des Zentralnervensystems einhergehenden Verhaltensänderungen zu bestimmen und auf diese Weise zu Annahmen über die funktionelle Struktur des Gehirns zu kommen. Im folgenden werden wir Kognition ausschließlich auf einer funktionalen Ebene betrachten. Ein allgemeinverständliches Nachschlagewerk mit neurowissenschaftlichem Schwerpunkt ist Gregory [1990]; weitergehende Informationen bei Gazzaniga [1994].

2.8.4.3 Kognitive Modellierung durch Computersimulation

Computersimulation im Rahmen der Kognitionswissenschaft ist das Zentralstück der kognitiven Modellierung [Lewis, 1999; Simon und Wallach, 1999; Strube, in press]. Ziel

dieses Verfahrens ist es nicht nur, Einsicht in die Art und Weise der menschlichen Informationsverarbeitung zu gewinnen, sondern auch dazu beizutragen, intelligente maschinelle Systeme zu schaffen. Zwei Aspekte der Computersimulation können unterschieden werden:

- Computersimulation als Hilfsmittel zur Präzisierung und logischen Strukturierung bestehender Theorien
- Computersimulation als Anregung zur Formulierung einer neuen Theorie über kognitive Prozesse.

Einerseits gilt also die Computersimulation als geeignet, um Annahmen über kognitive Strukturen und Prozesse zu testen. Dazu werden einzelne Komponenten eines theoretischen Modells in ein entsprechend konzipiertes Programm übertragen, um zunächst ihre Funktionsfähigkeit zu prüfen. Weiter sollen die einzelnen Verarbeitungsschritte in ihrer Komplexität (z.B. Zeitbedarf relativ zu anderen Verarbeitungsschritten) und ihren Zwischenresultaten möglichst genau mit dem übereinstimmen, was von menschlicher Kognition bekannt ist. Hierzu bedarf es natürlich erst der empirischen Forschung, so daß die oben genannten drei Methodengruppen stets aufeinander angewiesen sind. Die empirische Forschung hat dann den Charakter der Validierung von Theorien, die dazu erst formalisiert bzw. implementiert werden müssen.

Computersimulation als Hilfsmittel zur Strukturierung und Prüfung bestehender Theorien gestattet die Realisierung eines traditionellen wissenschaftstheoretischen Anspruchs, der in ausschließlich experimentell geprüften Konzepten oftmals nicht eingelöst werden kann: der logischen Widerspruchsfreiheit eines theoretischen Entwurfs. Dabei ist die in der Computersimulation notwendige Konkretisierung aller Details eines Modells geradezu heilsam, da auf diese Weise ein höherer Grad an Differenziertheit erreicht wird. Die durch die Computersimulation erzwungene Eindeutigkeit der benutzten Begriffe gewährleistet ein hohes Maß an Präzision des so entwickelten Modells.

Für die unterschiedlichen Berechnungsmodelle (Symbolverarbeitung, Konnektionismus) stellt sich der theoretische Erklärungswert von Computermodellen verschieden dar. In Systemen, die auf dem Symbolverarbeitungsansatz basieren, besteht die Möglichkeit einer funktionalen Dekomposition: Einzelteile des entsprechenden Programms können eliminiert werden, um ihre Bedeutung im Gesamtkonzept zu ermitteln. Damit lassen sich Aussagen über den funktionalen Zusammenhang der verschiedenen Komponenten tätigen, woraus sich letztlich das theoretische Modell ableiten läßt. Dagegen ist die analytische Zergliederung eines konnektionistischen Netzwerks, wenn überhaupt möglich, ungleich schwieriger, da die Einzelelemente erst durch ihre Verknüpfung bedeutungstragend werden. Allerdings läßt sich im Konnektionismus ein deutlicher Trend zu differenziert strukturierten Architekturen erkennen [McClelland, 1999].

Der praktische Aufwand für die kognitive Modellierung ist hoch. Daher empfiehlt es sich, einen von zwei Wegen zu gehen:

- Modellierung auf der Grundlage einer sogenannten kognitiven Architektur. Darunter sind umfangreiche, in jahrelanger Arbeit auf der Grundlage allgemein anerkannter Annahmen über die menschliche Informationsverarbeitung entwickelte Programme (oft samt Entwicklungsumgebung) zu verstehen, die dem Anwender ersparen, von Grund auf neu zu beginnen. Die bekanntesten sind ACT-R [Anderson

und Lebiere, 1998] und SOAR [Laird *et.al.*, 1987]. Beide stammen von Produktionssystemen ab, wobei sie inzwischen auch bemüht sind, perzeptuelle und motorische Komponenten (bzw. Interfaces zu solchen) zu integrieren.

- Modellierung *from scratch*, wobei diese auf überschaubare Teilbereiche der Kognition beschränkt sein muß. Ein Beispiel ist der Aufbau mentaler Modelle beim räumlich-relationalen Schließen [Schlieder, 1995; Behrendt, 1996; Knauff *et.al.*, 1998]. Der Vorteil eines solchen Ansatzes besteht darin, daß man nicht die Festlegungen einer existierenden „kognitiven Architektur" übernehmen muß. Inzwischen gibt es mit COGENT [Cooper und Fox, 1998] auch ein *toolkit*, das die Entwicklungsarbeit erleichtert.

Gerade die Methodologie der kognitiven Modellierung hat zu besonders überzeugenden Einsichten in die menschliche Informationsverarbeitung geführt, weil theoretisch abgestützte und empirisch überprüfte Computersimulationen zu generativen Theorien führen. Denn während herkömmliche empirische Verfahren lediglich Aufschluß darüber geben, daß bestimmte Bedingungen einen Einfluß auf kognitive Prozesse haben, können Computersimulationen detailliert darstellen, auf welchem Wege eine solche Beeinflussung erfolgt, und sie können die kognitiven Phänomene, um deren Erklärung es geht, selbst nachahmend hervorbringen. Dabei können auch neue Phänomene beobachtet werden, die zu weiterer empirischer Forschung Anlaß geben. Bisher 25 Jahre kognitionswissenschaftlicher Forschung haben den Erfolg dieser KI-nahen Methodologie bestätigt.

2.8.5 Anwendungen

Im Bereich der Informatik sind vor allem Anwendungen auf die Gestaltung der Mensch-Computer-Interaktion von Bedeutung, wobei neben der „kognitiven Ergonomie" von Interfaces auch besondere Fragestellungen, etwa der Gestaltung von ressourcenadaptiven Benutzerschnittstellen [Wahlster *et.al.*, 1998] eine Rolle spielen. Auch die Gestaltung von Entwicklungsumgebungen ist hier zu nennen (Schlieder & Hagen, i.Druck). Ein weiterer Bereich ist das *knowledge engineering* für wissensbasierte Systeme.

Weitere Anwendungen finden sich beim Design von Informationssystemen (Indexierung), in der Arbeitsgestaltung bei der computergestützten Modellierung betrieblicher Abläufe, in der Medizin und in vielen anderen Bereichen. Vor allem ist hier die Pädagogik zu nennen, wo es gegenwärtig um die Optimierung von Wissenserwerb mit Hilfe moderner elektronischer Medien geht.

Literaturverzeichnis

[Abelson, 1981] Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist, 36*, 715-729.

[Ades und Steedman, 1982] Ades, A., & Steedman, M. (1982). On the order of words. *Linguistics and Philosophy, 4*, 517-558.

[Allen, 1987] Allen, J. (1987). Natural language understanding. In A. Apt & J. Weisel (Hrsg.),. Sand Hill Road, Menlo Park, California: The Benjamin/Cummings Publishing Company.

[Anderson, 1976] Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale N.J.: Erlbaum.

[Anderson, 1982] Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

[Anderson, 1983] Anderson, J. R. (1983). *The architecture of cognition.* Cambridge MA: Harvard Univ. Press.

[Anderson und Lebiere, 1998] Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought.* Hillsdale, NJ: Erlbaum.

[Arbib, 1972] Arbib, M. A. (1972). *The metaphorical brain: An introduction to cybernetics as artificial intelligence and brain theory.* New York: Wiley-Interscience.

[Atkinson und Shiffrin, 1968] Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In O. Spence & O. Spence (Eds.), *Advances in the psychology of learning and motivation* (vol. 2, pp. 89-195). New York: Academic Press.

[Baddeley, 1986] Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford Univ. Press.

[Balkenius, 1995] Balkenius, C. (1995). *Natural intelligence in artificial creatures* (Lund University Cognitive Studies, vol. 37). Lund: Lund University.

[Barsalou, 1983] Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11,* 211-227.

[Barsalou, 1999] Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577-660.

[Bartlett, 1932] Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press.

[Becker, 1998] Becker, B. (1998). Leiblichkeit und Kognition. In P. Gold & A. K. Engel (Hrsg.), *Der Mensch in der Perspektive der Kognitionswissenschaften* (pp. 270-288). Frankfurt: Suhrkamp (stw 1381).

[Behrendt, 1996] Behrendt, B. (1996). Explaining preferred mental models in Allen inferences with a metrical model of imagery. *Proceedings of the 18th Annual Conference of the Cognitive Science Society,* 489-494.

[Bever, 1970] Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279-362). New York: Wiley

[Biederman *et.al.*, 1982] Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143-177.

[Block, 1981] Block, N. (1981). *Imagery.* Cambridge, MA: MIT Press.

[Bower *et.al.*, 1979] Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11,* 177-220.

[Brachman und Schmolze, 1985] Brachman, R. J., & Schmolze, J. G. (1985). An overview of the KL-ONE knowledge representation system. *Cognitive Science, 9,* 171-216.

[Braitenberg, 1984] Braitenberg, V. (1984). *Vehicles.* Cambridge, MA: MIT Press.

[Bresnan und Kaplan, 1982] Bresnan, J., & Kaplan, R. M. (1982). Lexical Functional Grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The mental representation of grammatical relations.* Cambridge, MA: MIT Press.

[Brooks, 1991] R. A. Brooks. Intelligence without representation. *Artificial Intelligence,* 47:139–159, 1991.

[Caplan und Waters, 1999] D. Caplan und G. S. Waters. Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences,* 22:77–94, 1999.

[Ceci und Liker, 1986] Ceci, S. J., , Liker, & J. K.. (1986). A Day at the Races: A Study of IQ, Expertise, and Cognitive Complexity. *Journal of Experimental Psychology; General, 115 (3),* 255 - 266.

[Chase und Simon, 1973] Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

[Cherry, 1953] Cherry, C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America, 25,* 975-979.

[Chi *et.al.*, 1988] Chi, M., Glaser, R., & Farr, M. (Eds.). (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

[Chomsky, 1965] Chomsky. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

[Chomsky 1959a] Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control, 2,* 137-167.

[Chomsky 1959b] Chomsky, N. (1959). Verbal behavior. *Language, 35,* 26-58.

[Chomsky, 1981] Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

[Chomsky, 1986] Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use.* New York: Praeger.

[Chomsky, 1995] Chomsky, N. (1995). *The minimalist program.* Cambridge, MA: MIT Press.

[Collins und Loftus, 1975] Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407-428.

[Collins und Quillian, 1969] Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240-247.

[Cooper und Fox, 1998] Cooper, R., & Fox, J. (1998). COGENT: a visual design environment for cognitive modeling. *Behavior Research Methods, Instruments & Computers, 30*, 553-564.

[Crain und Steedman, 1985] Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. R. Dowty, L. Karttunen & A. R. Zwicky (Eds.), *Natural language parsing* (pp. 320-358). Cambridge: Cambridge University Press.

[Dennett, 1991] Dennett, D. C. (1991). *Consciousness explained.* Boston: Little, Brown & Co.

[Dijk, 1983] Dijk, T. A., van, Kintsch, & W.. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

[Donders 1868] Donders, F. C. (1868). Over de snelheid van psychische processen. *Onderzoekingen gedann on het Physiologisch Laboratorium der Utrechtschen Hoogeschool, 2*, 92-120.

[Duncker, 1935] Duncker, K. (1935). *Zur Psychologie des produktiven Denkens.* Berlin: Springer.

[Elman, 1991] Elman, J. L. (1991). Incremental learning, or the importance of starting small. *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 443-448). Hillsdale, NJ: Lawrence Erlbaum Associates.

[Erdfelder, 1996] Erdfelder, E. (1996). Experiment. In G. Strube et.al. (Hrsg.), *Wörterbuch der Kognitionswissenschaft* (pp. 164-169). Stuttgart: Klett-Cotta.

[Ericsson, 1985] Ericsson, K. A. (1985). Memory skill. *Canadian Journal of Psychology, 39*, 188-231.

[Ericsson und Smith, 1991] Ericsson, K. A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits.* Cambridge: Cambridge University Press.

[Erman *et.al.*, 1980] Erman, L. D., Hayes-Roth, F., Lesser, V. R., & Reddy, D. R. (1980). The Hearsay-II speech-understanding system: integrating knowledge to resolve uncertainty. *Computer Surveys, 12*, 213-253.

[Flores d'Arcais, 1987] G. B. Flores d'Arcais. Syntactic processing during reading comprehension. In M. Coltheart (Hg.), *The psychology of reading*, Seite 619-633. Lawrence Erlbaum, Hove/London/Hillsdale, 1987.

[Fodor, 1975] Fodor, J. A. (1975). *The language of thought.* New York: Crowell.

[Fodor, 1980] Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *The Behavioral and Brain Sciences, 3*, 63-109.

[Fodor, 1983] Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

[Fodor und Pylyshyn, 1988] Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3-71.

[Ford und Dalrymple, 1988] Ford, M., & Dalrymple, M. (1988). A note on some psychological evidence and alternative grammars. *Cognition, 29*, 63-71.

[Frazier, 1987] Frazier, L. (1987). Sentence processing: a tutorial review. In M. Coltheart (Ed.), *The psychology of reading* (pp. 559-586). Hove: Lawrence Erlbaum.

[Frazier, 1987] Frazier, L. (1987). Theories of sentence processing. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language processing* (pp. 293-309). Cambridge, Massachusetts: MIT Press.

[Frazier, 1999] Frazier, L. (1999). *On sentence interpretation.* Dordrecht: Kluwer Academic Press.

[Frazier und Clifton, 1996] Frazier, L., & Clifton, C. (1996). *Construal.* Cambridge, MA: MIT Press.

[Friederici und Mecklinger, 1996] Friederici, A. D., & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing processes. *Journal of Psycholinguistic Research, 25(1)*, 157-176.

[Gallistel, 1993] Gallistel, C. R. (1993). *The organization of learning.* Cambridge, MA: MIT Press.

[Gardner, 1990] Gardner, H. (1990). *Dem Denken auf der Spu.* Stuttgart: Klett-Cotta.

[Garfield, 1987] Garfield, J. L. (Ed.). (1987). *Modularity in knowledge representation and natural language understanding.* Cambridge, MA: MIT Press.

[Gazzaniga, 1994] Gazzaniga, M. S. (Ed.). (1994). *The cognitive neurosciences.* Cambridge, MA: MIT Press.

[Gazzaniga und Hutsler, 1999] Gazzaniga, M. S., & Hutsler, J. J. (1999). Hemispheric specialization. In R. A. Wilson & F. C. Keil (Hrsg.), *The MIT encylopedia of the cognitive sciences* (pp. 369-372). Cambridge, MA: MIT Press.

[Gibson *et al.*, 1996] E. Gibson, N. Pearlmutter, E. Canseco-Gonzalez und G. Hickock. Recency preference in the human sentence processing mechanism. *Cognition*, 59:23–59, 1996.

[Gick und Holyoak, 1980] Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 15*, 1-38.

[Gigerenzer und Goldstein, 1996] Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-669.

[Gigerenzer und Hug, 1992] Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43*, 127-171.

[Gilboy *et al.*, 1995] E. Gilboy, J. . M. Sopena, C. Clifton und L. Frazier. Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, 54:131–167, 1995.

[Godden und Baddeley, 1975] D. R. Godden und A. D. Baddeley. Context–dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66:325–331, 1975.

[Goldstone und Barsalou, 1998] Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition, 65*, 197-230.

[Goschke, 1996] Goschke, T. (1996). Wille und Kognition: Zur funktionalen Architektur der intentionalen Handlungssteuerung. In J. Kuhl & H. Heckhausen (Hrsg.), *Motivation, Volition und Handlung* (pp. 583-663). Göttingen: Verlag für Psychologie.

[Gregory, 1990] Gregory, R. L. (Ed.). (1990). *The Oxford companion to the mind.* Oxford: Oxford University Press.Grice, H. P. (1975). William James Lectures, Harvard University, 1967. In P. Cole & J. L. Morgan (Hrsg.), *Syntax and semantics* (vol. 3). New York: Seminar Press.

[Grice, 1975] H. P. Grice. Logic and conversation. In P. Cole und J. L. Morgan (Hg.), *Syntax and Semantics*, Band 3, Seite 41–58. Academic Press, New York, 1975.

[Gruber und Strube, 1989] Gruber, H., & Strube, G. (1989). Zweierlei Experten. Eine experimentelle Untersuchung zur Expertise im Schachspiel und beim Lösen von Schachproblemen. *Sprache und Kognition, 8*, 72-85 (Extended English Abstract in: German Journal of Psychology).

[Habel, 1986] Habel, C. (1986). Stories - an Artificial Intelligence perspective (?). *Poetics, 15*, 111-125.

[Habel, 1988] Habel, C. (1988). Repräsentation räumlichen Wissens. In G. Rahmstrof (Hrsg.), *Wissensrepräsentation in Expertensystemen* (pp. 98-131). Berlin: Springer.

[Habel, 1998] Habel, C. (1998). Piktorielle Repräsentationen als unterbestimmte räumliche Modelle. *Kognitionswissenschaft, 7*, 58-67.

[Habel *et.al.*, 1990] Habel, C., Kanngießer, S., & Strube, G. (1990). Editorial „Kognitionswissenschaft". *Kognitionswissenschaft, 1*, 1-3.

[Habel und Tappe, 1999] Habel, C., & Tappe, H. (1999). Processes of seg-men-ta-tion and linearization in describing events. In R. Klabunde & C. Stutterheim (Eds.), *Representations and processes in language production* (pp. 117-152). Wiesbaden: Deutscher Universitäts-Verlag.

[Hagoort *et.al.*, 1993] Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*(4), 439-483.

[Hassoun, 1995] Hassoun, M. H. (1995). *Fundamentals of artificial neural networks.* Cambridge, MA: MIT Press.

[Hawkins, 1994] Hawkins, J. A. (1994). *A performance theory of order and constituency* (Cambridge studies in Linguistics). Cambridge University Press.

[Hebb, 1949] Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

[Hemforth, 1993] Hemforth, B. (1993). *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens.* Sankt Augustin: Infix.

[Hemforth und Konieczny, 2000] Hemforth, B., & Konieczny, L. (2000). *German Sentence Processing.* Dordrecht, NL: Kluwer Academic Press.

[Herrmann, 1985] Herrmann, T. (1985). *Allgemeine Sprachpsychologie. Grundlagen und Probleme.* München: Urban & Schwarzenberg.

[Hewitt, 1977] Hewitt, C. (1977). Viewing control structures as patterns of passing messages. *Artificial Intelligence, 8*, 323-364.

[Hinton *et al.*, 1986] G. E. Hinton, J. McClelland und D. E. Rumelhart. Distributed representations. In D. E. Rumelhart und J. McClelland (Hg.), *Parallel distributed processing*, Band 1, Seite 77–109. MIT Press, Cambridge, MA, 1986.

[Hinton, 1989] Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence, 40*, 185-234.

[Hörmann, 1976] Hörmann, H. (1976). *Meinen und Verstehen. Grundzüge einer psychologischen Semantik.* Frankfurt: Suhrkamp.

[Hoffmann, 1986] Hoffmann, J. (1986). *Die Welt der Begriffe. Psychologische Untersuchungen zur Organisation des menschlichen Wissens.* Berlin: Deutscher Verlag der Wissenschaften.

[Holyoak, 1995] Holyoak, K. J. (1995). Problem solving. In E. E. Smith & D. N. Osherson (Hrsg.), *Thinking (An invitation to cognitive science, vol. 3)* (2. Aufl., pp. 267-296). Cambridge, MA: MIT Press.

[Holyoak und Thagard, 1994] Holyoak, K. J., & Thagard, P. (1994). *Mental leaps.* Cambridge, MA: MIT Press.

[Hutchins, 1995] Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: MIT Press.

[Hutchins, 1995] Hutchins, E. (1995). How a cockpit remembers its speed. *Cognitive Science, 19*, 265-288.

[Irtel, 1993] Irtel, H. (1993). *Experimentalpsychologisches Praktikum.* Berlin: Springer.

[Johnson-Laird, 1983] Johnson-Laird, P. N. (1983). *Mental Models. Towards a cognitive science of language, inference, and consciousness.* Cambridge: Cambridge University Press.

[Johnson-Laird und Byrne, 1991] Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hove: Erlbaum.

[Johnson-Laird und Wason, 1977] Johnson-Laird, P. N., & Wason, P. C. (1977). A thoretical analysis of insight into a reasoning task. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 143-157). Cambridge: Cambridge University Press.

[Jonides, 1995] Jonides, J. (1995). Working memory and thinking. In E. E. Smith & D. N. Osherson (Eds.), *Thinking: An invitation to cognitive science, Vol. 3* (pp. 215-265). Cambridge, MA, US: MIT Press.

[Juliano und Tanenhaus, 1994] Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research, 23*, 459-471.

[Just und Carpenter, 1992] Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

[Kahneman, 1973] Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, N.J.: Prentice-Hall.

[Kahneman *et.al.*, 1982] Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

[Kahnemann und Tversky, 1973] Kahnemann, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, No. 4, 80*, 237-251.

[Kandel *et.al.*, 1995] Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (1995). *Essentials of neural science and behavior.* London: Prentice Hall International (dt. 1996, Heidelberg: Spektrum).

[Karmiloff-Smith, 1999] Karmiloff-Smith, A. (1999). Modularity of mind. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 558-560). Cambridge, MA: MIT Press.

[Kemmerling, 1991] Kemmerling, A. (1991). Eine Verteidigung des Repräsentationalismus? *Kognitionswissenschaft, 2*(2), 99-104.

[Kintsch, 1997] Kintsch, W. (1997). *Comprehension.* Cambridge: Cambridge University Press.

[Kintsch und Dijk, 1978] Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

[Knauff *et.al.*, 1998] Knauff, M., Rauh, R., Schlieder, C., & Strube, G. (1998). Mental models in spatial reasoning. In C. Freksa, C. Habel & K. F. Wender (Hrsg.), *Spatial cognition (Lecture Notes in Artificial Intelligence)* (vol. 1404, pp. 267-291). Berlin: Springer.

[Kolb und Wishaw, 1990] Kolb, B., & Wishaw, I. Q. (1990). *Fundamentals of human neuropsychology.* New York: Freeman.

[Kolodner, 1993] Kolodner, J. (1993). *Case-based reasoning.* San Mateo, CA: Morgan Kaufmann.

[Konieczny, 1996] Konieczny, L. (1996). *Human sentence processing: a semantics-oriented parsing approach* (IIG-Berichte 3/96 (Phil. Diss.)). Freiburg: IIG, Universität Freiburg.

[Konieczny *et al.*, 1997] L. Konieczny, B. Hemforth, C. Scheepers und G. Strube. The role of lexical heads in parsing: evidence from German. *Language and Cognitive Processes*, 12:307–348, 1997.

[Konieczny *et al.*, 2000] L. Konieczny, B. Hemforth und C. Scheepers. Head position and clause boundary effects in reanalysis. In B. Hemforth und L. Konieczny (Hg.), *German sentence processing*, Seite 247–278. Kluwer Academic Press, Dordrecht, 2000.

[Kosslyn, 1980] Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

[Kosslyn, 1994] Kosslyn, S. M. (1994). *Image and brain.* Cambridge, MA: MIT Press.

[Krämer, 1988] Krämer, S. (1988). *Symbolische Maschinen. Die Idee der Formalisierung in geschichtlichem Abriß.* Darmstadt: Wissenschaftliche Buchgesellschaft.

[Kuhl, 1983] Kuhl, J. (1983). *Motivation, Konflikt und Handlungskontrolle.* Berlin: Springer.

[Laird *et.al.*, 1987] Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence, 33*, 1-64.

[Lehman *et.al.*, 1998] Lehman, J. F., Laird, J. E., & Rosenbloom, P. (1998). A gentle introduction to Soar: an architecture for human cognition. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, and conceptual issues (An invitation to cognitive science, vol. 4)* (pp. 211-251). Cambridge, MA: MIT Press.

[Lehnert, 1982] Lehnert, W. (1982). Plot units. A narrative summarization structure. In W. G. Lehnert & M. H. Ringle (Eds.), *Strategies for Natural Language Processing.* Hillsdale, NJ: Lawrence Erlbaum.

[Lewis, 1998] Lewis, R. (1998). *A new computational model of sentence processing: interference-limited working memory and Non-competitive Ambiguity Resolution.* Proceedings of AMLaP-98, U. of Freiburg.
`http://www.iig.uni-freiburg.de/cognition/events/amlap98/amlap98d.htm`.

[Lewis, 1999] Lewis, R. L. (1999). Cognitive modeling, symbolic. In R. A. Wilson & F. C. Keil (Hrsg.), *MIT encyclopedia of the cognitive sciences* (pp. 141-143). Cambridge, MA: MIT Press.

[Logie, 1995] Logie, R. H. (1995). *Visual-spatial working memory.* Hove, UK: Erlbaum.

[MacDonald *et al.*, 1994] M. MacDonald, N. Pearlmutter und M. S. Seidenberg. The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101:676–703, 1994.

[Macnamara, 1986] Macnamara, J. (1986). *A border dispute - the place of logic in psychology.* Cambridge, MA: MIT Press.

[Maes, 1994] Maes, P. (1994). Modeling adaptive autonomous agents. *Artificial Life, 1*, 135-162.

[Malsch *et.al.*, 1996] Malsch, T., Florian, M., Jonas, M., & Schulz-Schaeffer, I. (1996). Sozionik: Expeditionen ins Grenzgebiet zwischen Soziologie und Künstlicher Intelligenz. *KI*(2), 6-12.

[Marr, 1982] Marr, D. (1982). *Vision. A computational investigation into the human representation and processing of visual information.* San Francisco: Freeman.

[Marslen-Wilson und Tyler, 1980] Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1-71.

[Marslen-Wilson und Tyler , 1987] Marslen-Wilson, W. D., & Tyler, L. K. (1987). Against modularity. In J. Garfield (Hrsg.), *Modularity in knowledge representation and natural language understanding.* Cambridge, MA: MIT Press.

[Martin *et al.*, 1994] R. C. Martin, J. R. Shelton, L. Yaffee und S. Language processing and working memory: neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, 33:83–111, 1994.

[Maturana und Varela, 1980] Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition: The Realisation of the Living.* London: Reidel.

[McClelland, 1999] McClelland, J. L. (1999). Cognitive modeling: connectionist. In R. A. Wilson & F. C. Keil (Hrsg.), *The MIT encyclopedia of the cognitive sciences* (pp. 137-141). Cambridge, MA: MIT Press.

[McFarland und Bösser, 1996] McFarland, D., & Bösser, T. (1996). *Intelligent behavior in animals and robots.* Cambridge, MA: MIT Press.

[McKoon und Ratcliff, 1981] McKoon, G., & Ratcliff, R. (1981). The Comprehension Processes and Memory Structures Involved in Instrumental Inference. *Journal of Verbal Learning and Verbal Behavior* (vol. 20, pp. 671-682). New York, London: Academic Press.

[Meltzoff und Moore, 1983] Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development, 54*, 702-709.

[Metzinger, 1993] Metzinger, T. (1993). *Subjekt und Selbstmodell.* Paderborn: Schöningh.

[Metzler und Shepard, 1974] Metzler, J., & Shepard, R. N. (1974). Transformational studies of the internal representation of three-dimensional objects. In R. L. Solso (Ed.), *Theories in cognitive psychology: the London symposium* (pp. 147-201). Potomac, MD: Erlbaum.

[Meyer und Kieras, 1997] Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple task performance: Part 1. Basic mechanisms. *Psychological Review, 104*, 3-65.

[Miller, 1956] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-96.

[Miller und Johnson-Laird, 1976] Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception.* Cambridge: Cambridge University Press.

[Minsky, 1975] Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.

[Minsky, 1981] Minsky, M. (1981). A framework for representing knowledge. In J. Haugeland (Ed.), *Mind design* (pp. 95-128). Cambridge, MA: MIT Press.

[Mitchell, 1994] Mitchell, D. C. (1994). Sentence parsing. In M. A. Gernsbacher (Hrsg.), *Handbook of psycholinguistics* (pp. 375-409). San Diego: Academic Press.

[Miyake und Shah, 1999] Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory. Mechanisms of active maintenance and executive control.* Cambridge: Cambridge University Press.

[Morris und Murphy, 1990] Morris, M. W., & Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition, 18*, 407-418.

[Moyer und Landauer, 1967] Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215*, 1519-1520.

[Müller, 1996] Müller, J. P. (1996). *The design of intelligent agents.* Berlin: Springer (LNAI 1177).

[Navon und Gopher, 1979] Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86*, 214-255.

[Nebel, 1990] Nebel, B. (1990). *Reasoning and revision in hybrid representation systems.* Berlin: Springer.

[Neisser, 1976] Neisser, U. (1976). *Cognition and reality.* San Francisco: Freeman.

[Newell, 1980] Newell, A. (1980). Physical Symbol Systems. *Cognitive Science, 4*, 135-183.

[Newell, 1982] Newell, A. (1982). The knowledge level. *Artificial Intelligence, 18*, 87-127.

[Newell, 1990] Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

[Newell und Rosenbloom, 1981] Newell, A., Rosenbloom, P. S., &. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, N. J.: Lawrence Erlbaum Ass.

[Newell und Simon, 1972] Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

[Newell und Simon, 1976] Newell, A., & Simon, H. A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the ACM, 19*, 113-126.

[Nisbett *et al.*, 1983] R. E. Nisbett, D. H. Krantz, D. Jepson und Z. Kunda. The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90:339–363, 1983.

[Norman, 1981] Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88*, 1-15.

[Osterhout und Holcomb, 1992] Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language., 31*, 785-806.

[Paivio, 1986] Paivio, A. (1986). *Mental representation: a dual coding approach.* New York: Oxford University Press.

[Palmer, 1978] Palmer, S. E. (1978). Fundamental aspects of cognitive representations. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 259-303). Hillsdale, NJ: Erlbaum.

[Pinker, 1984] Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

[Pinker, 1994] Pinker, S. (1994). *The language instinct.* New York: Morrow; 1995: Harper Perennial (dt. Der Sprachinstinkt. München: Kindler, 1996).

[Plötzner, 1998] Plötzner, R. (1998). *Flexibilität im Problemlösen und Lernen - Konstruktion, Anwendung und Koordination von Repräsentationssystemen.* Lengerich: Pabst.

[Pollard und Sag, 1994] Pollard, C., & Sag, I. A. (1994). *Head-driven phrase-structure grammar.* Chicago: University of Chicago Press.

[Prince und Smolensky, 1997] Prince, A., & Smolensky, P. (1997). Optimality: from neural networks to universal grammar. *Science, 275*, 1604-1610.

[Prinz, 1976] Prinz, W. (1976). Kognition, kognitiv. In J. Ritter & K. Gründer (Hrsg.), *Historisches Wörterbuch der Philosophie* (vol. 4, pp. 866-878). Basel: Schwabe & Co.

[Pulman, 1986] Pulman, S. G. (1986). Grammars, parsers, and memory limitations. *Language and Cognitive Processes* (vol. 1, No 3, pp. 197-225). VNU Science Press.

[Pylyshyn, 1981] Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88*, 16-45.

[Pylyshyn, 1989] Pylyshyn, Z. W. (1989). Computing in cognitive science. In M. I. Posner (Hrsg.), *Foundations of cognitive science* (pp. 49-91). Cambridge, MA: MIT Press (Bradford).

[Quillian, 1968] Quillian, M. R. (1968). Semantic memory. In M. Minsky (Hrsg.), *Semantic information processing* (pp. 227-270). Cambridge, Mass.: MIT-Press.

[Raphael, 1976] Raphael, B. (1976). *The thinking computer.* San Francisco: Freeman.

[Rapp und Caramazza, 1995] Rapp, B. C., & Caramazza, A. (1995). Disorders of lexical processing and the lexicon. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 901-913). Cambridge, MA: MIT Press.

[Rayner und Sereno, 1994] Rayner, K., & Sereno, S. (1994). Eye movements in reading: Psycholinguistic studies. In M. A. Gernsbacher (Hrsg.), *Handbook of psycholinguistics* (pp. 57-82). San Diego: Academic Press.

[Reason, 1990] Reason, J. (1990). *Human error.* Cambridge: Cambridge University Press.

[Rescorla und Wagner, 1972] Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II.* New York: Appleton-Century-Crofts.

[Rips, 1995] Rips, L. (1995). Deduction and cognition. In E. E. Smith & D. N. Osherson (Eds.), *Thinking (An invitation to cognitive science, vol. 3)* (2nd ed., pp. 297-343). Cambridge, MA: MIT Press.

[Rips, 1994] Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking.* Cambridge, MA: MIT Press.

[Rips *et.al.*, 1973] Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12,* 1-20.

[Rizzolatti *et.al.*, 1996] Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3,* 131-141.

[Rosch, 1973] Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York: Academic Press.

[Rosch, 1975] Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192-233.

[Rosch, 1977] Rosch, E. (1977). Human categorization. In N. Warren (Hrsg.), *Advances in cross-cultural psychology* (vol. 1, pp. 1-49).

[Rumelhart, 1977] Rumelhart, D. E. (1977). Understanding and summarizing brief stories. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: perception and comprehension.* Hillsdale, NJ: Erlbaum.

[Rumelhart und McClelland, 1986] Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge, MA: MIT Press.

[Sacerdoti, 1977] Sacerdoti, E. D. (1977). *A structure for plans and behavior.* New York: Elsevier North-Holland.

[Schank, 1972] Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology, 3,* 552-631.

[Schank, 1982] Schank, R. C. (1982). *Dynamic memory.* Cambridge: Cambridge Univ. Press.

[Schank und Abelson, 1977] Schank, R. C., & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding.* Hillsdale, NJ: Erlbaum.

[Schank und Riesbeck, 1981] Schank, R. C., & Riesbeck, C. K. (Eds.). (1981). *Inside computer understanding: Five programs plus miniatures (The Artificial Intelligence Series).* Hillsdale, NJ: Erlbaum.

[Schlieder, 1995] Schlieder, C. (1995). *The construction of preferred mental models in reasoning with the interval relations* (IIG-Bericht 3/95). Freiburg: Universität Freiburg, IIG.

[Schmalhofer, 1982] Schmalhofer, F. (1982). *Comprehension for a technical text as a function of expertise.* PhD Diss., Univ. Col. Boulder, Univ. Microfilms Int.

[Schmalhofer, 1999] Schmalhofer, F. (1999). *Constructive knowledge acquisition.* Mahwah, NJ: Erlbaum.

[Schneider und Detweiler, 1987] Schneider, W., & Detweiler, M. (1987). A connectionist-control architecture for working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (vol. 21, pp. 53-119). London: Academic Press, INC.

[Schneider *et.al.*, 1984] Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1984). Automatic and control processing and attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 1-27). New York: Academic Press.

[Schreiber *et.al.*, 1993] Schreiber, G., Wielinga, B., & Breuker, J. (1993). *KADS. A principled approach to knowledge-based systems development.* London: Academic Press.

[Schyns *et.al.*, 1998] Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21,* 1-54.

[Searle, 1983] Searle, J. (1983). *Intentionality: an essay in the philosophy of mind.* Cambridge: Cambridge University Press.

[Searle, 1969] Searle, J. R. (1969). *Speech acts: an essay in the philosphy of language.* Cambridge: Cambridge University Press.

[Searle, 1992] Searle, J. R. (1992). *The rediscovery of the mind.* Cambridge, MA: MIT Press.

[Shallice, 1982] Shallice, T. (1982). Specific impairment of planning. *Proceedings of the Royal Society, Series B, 298*, 199-209.

[Shepard und Cooper, 1982] Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

[Shepard und Metzler, 1971] R. N. Shepard und J. Metzler. Mental Rotation of Three–Dimensional Objects. *Science*, 171:701–703, 1971.

[Simon, 1969] Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

[Simon und Wallach, 1999] Simon, H., & Wallach, D. (1999). Cognitive modeling in perspective. *Kognitionswissenschaft, 8*, 1-4.

[Sloan Foundation, 1978] Sloan Foundation. (1978). *Cognitive Science, 1978. Report of the State of the Art Committee.* New York: Alfred P. Sloan Foundation.

[Smith, 1995] Smith, E. E. (1995). Concepts and categorization. In E. E. Smith & D. N. Osherson (Eds.), *Thinking (An invitation to cognitive science, vol. 3)* (2nd ed., pp. 3-33). Cambridge, MA: MIT Press.

[Smith und Medin, 1981] Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge: Cambridge Univ. Press.

[Smolensky, 1988] Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*, 1-74.

[Stillings *et al.*, 1995] N. A. Stillings, M. H. Feinstein, J. L. Garfield, E. L. Rissland, D. A. Rosenbaum, S. E. Weisler und L. Baker–Ward. *Cognitive science. An introduction.* MIT Press, Cambridge, MA, 2 Auflage, 1995.

[Strube 1996a] Strube, G. (1996). Knowledge-based systems from a socio-cognitive perspective. *Behaviour & Information Technology, 15*, 276-288.

[Strube 1996b] Strube, G. (1996). Kognition. In G. Strube, B. Becker, C. Freksa, U. Hahn, K. Opwis & G. Palm (Hrsg.), *Wörterbuch der Kognitionswissenschaft* (pp. 303-317). Stuttgart: Klett-Cotta.

[Strube *et.al.*, 1996] Strube, G., Becker, B., Freksa, C., Hahn, U., Opwis, K., & Palm, G. (Hrsg.). (1996). *Wörterbuch der Kognitionswissenschaft.* Stuttgart: Klett-Cotta.

[Strube *et.al.*, 1996] Strube, G., Janetzko, D., & Knauff, M. (1996). Cooperative construction of expert knowledge: The case of knowledge engineering. In P. B. Baltes & U. M. Staudinger (Eds.), *Interactive minds* (pp. 366-393). Cambridge: Cambridge University Press.

[Strube, 1998] Strube, G. (1998). Modelling motivation and action control in cognitive systems. In U. Schmid, J. Krems & F. Wysocki (Eds.), *Mind modelling: a cognitive science approach to reasoning, learning, and discovery* (pp. 89-108). Berlin: Pabst.

[Strube, in press] G. Strube. Cognitive modeling: research logic in cognitive science. In N. J. Smelser und P. B. Baltes (Hg.), *International encyclopedia of the social and behavioral sciences.* Elsevier Science, Oxford, in press.

[Suchman, 1987] Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication.* New York: Cambridge University Press.

[Swinburne, 1974] Swinburne, R. (Ed.). (1974). *The justification of induction.* Oxford: Oxford University Press.

[Tambe *et al.*, 1995] M. Tambe, W. L. Johnson, R. M. Jones, F. Koss, J. E. Laird, P. S. Rosenbloom und K. Schwamb. Intelligent agents for interactive simulation environments. *AI Magazine*, 16(1):15–39, 1995.

[Tang *et al.*, 1999] Y. . P. Tang, E. Shimizu, G. R. Dube, C. Rampon, G. Kerchner, M. Zhuo, G. Liu und J. Z. Tsien. Genetic enhancement of learning and memory in mice. *Nature*, 401:63–69, 1999.

[Tulving, 1972] Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). New York: Academic Press.

[Uszkoreit *et al.*, 1998] H. Uszkoreit, T. Brants, D. Duchier, B. Krenn, L. Konieczny, S. Oepen und W. Skut. Studien zur performanzorientierten Linguistik. Aspekte der Relativsatzextraposition im Deutschen. *Kognitionswissenschaft*, 7:129–133., 1998.

[Wahlster *et.al.*, 1998] Wahlster, W., Blocher, A., Baus, J., Stopp, E., & Speiser, H. (1998). Ressourcenadaptierende Objektlokalisation: Sprachliche Raumbeschreibung unter Zeitdruck. *Kognitionswissenschaft, 7*, 111-117.

[Wallesch *et.al.*, 1996] Wallesch, C.-W., Bartels, C., & Herrmann, M. (1996). Störungen höherer Hirnleistungen. In G. Strube et al. (Hrsg.), *Wörterbuch der Kognitionswissenschaft* (pp. 677-689). Stuttgart: Klett-Cotta.

[Wason und Johnson-Laird, 1972] Wason, P. C., & Johnson-Laird, P. N. (1972). *The psychology of reasoning.* Cambridge, MA: Harvard University Press.

[Wickelgren, 1974] Wickelgren, W. A. (1974). *How to solve problems. Elements of a theory of problems and problem solving.* New York: Freeman.

[Wilson und Keil, 1999] Wilson, R. A., & Keil, F. C. (Eds.). (1999). *The MIT encyclopedia of the cognitive sciences.* Cambridge, MA: MIT Press.

[Wittgenstein, 1953] Wittgenstein, L. (1953). *Philosophische Untersuchungen.* Frankfurt: Suhrkamp.

[Yerkes und Dodson, 1908] Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative and Neurological Psychology, 18,* 459-482.

[Yngve, 1960] Yngve, V. H. (1960). A model and hypothesis for language structure. *Proceedings of the American Philosophical Society, 104,* 444-466.

[Zhang, 1997] Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 21.*

Kapitel 3

Neuronale Netze

Hanspeter A. Mallot, Wolfgang Hübner und Wolfgang Stürzl

3.1 Motivation

Schon vom Namen her nimmt die *Künstliche Intelligenz* Bezug auf die Leistungen der biologischen Informationsverarbeitung, die ihr als Vorbild, als Existenzbeweis und als Beispiel für Teillösungen dient. In diesem Kapitel soll auf einen Aspekt biologischer Informationsverarbeitung näher eingegangen werden, der etwa in den letzten fünfzehn Jahren ein (erneutes) großes Interesse gefunden hat, die *Neuronalen Netze*. Die Neuronalen Netze bilden ein Teilgebiet der Neuroinformatik, deren Gegenstand die Erforschung biologischer Informationsverarbeitung mit den Methoden der Informatik und Informationstechnologie ist (vgl. [Mallot *et al.*, 1992]).

Biologische Informationsverarbeitung dient der Organisation von Verhalten. Der Zweck biologischer Informationsverarbeitung ist es, Verhalten (und interne Regulationsprozesse) in Abhängigkeit von der jeweiligen Umweltsituation so zu organisieren, dass das Lebewesen sich behauptet. Die Ausstattung mit sensorischen Fähigkeiten und Verhaltensrepertoire wirkt dabei auf die Komplexität dieser Aufgabe zurück. So hat zum Beispiel die Entwicklung der *Hand* als eines *Manipulations*organs und die damit verbundene Erweiterung des Verhaltensrepertoires weitreichende Konsequenzen auch für die Hirnentwicklung der Primaten gehabt. Biologische Informationsverarbeitung kann geradezu als die umweltbezogene Organisation von Verhalten definiert werden; Umwelt meint dabei im Sinne von Uexküll den Teil der Außenwelt, mit dem der Organismus interagieren (sich auseinandersetzen) kann [Uexküll, 1956].

Gehirne sind keine universellen Rechner. Im Vergleich zur technischen Informationsverarbeitung ergeben sich aus dieser Betrachtungsweise einige wichtige Unterschiede. Offensichtlich sind Gehirne keine Computer im gängigen Sinn: Sie sind nicht auf Universalität angelegt und können nicht „programmiert" werden; ihre Flexibilität gewinnen sie durch Anpassungen der Struktur („Plastizität"). Auf der anderen Seite zeigt ein Blick auf Abb. 3.1, dass die Probleme, zu denen man sich von der Neuroinformatik einen Beitrag erwarten kann (Sehen, Sprache, Manipulation, Lernen), große Bedeutung auch für die technische Informationsverarbeitung und die künstliche Intelligenz haben.

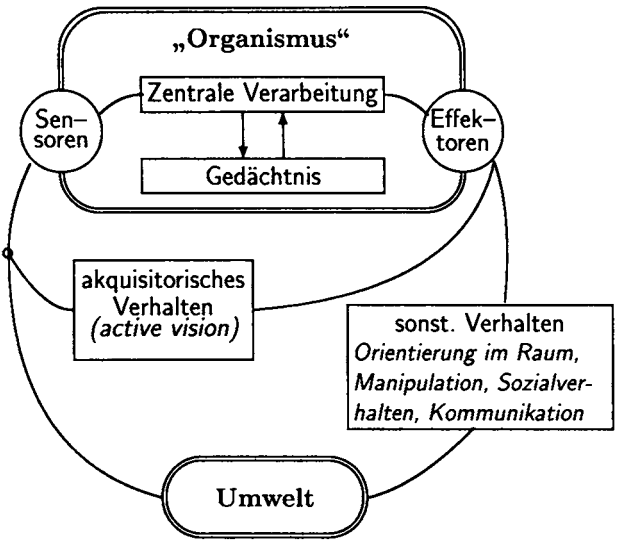


Abbildung 3.1: Biologische Informationsverarbeitung kann als „Informationswechsel“ aufgefasst werden, bei dem Verhalten in einer der jeweiligen Umweltsituation angepassten Weise erzeugt wird. Rückkopplungen des Verhaltens auf die wahrnehmbaren Sinnesreize finden entweder direkt („akquisitorisches Verhalten“, z.B. Augenbewegungen) oder über Wechselwirkungen mit der Umwelt statt. Zur „Umwelt“ gehören auch andere Organismen.

Die Arbeitsweise des Gehirns zeigt sich in seiner Struktur. Im Gegensatz zu *universellen* Rechnern sind Gehirne an ihre Funktionen angepasst. Nur aus diesem Grund ist es überhaupt möglich, im Sinne der neuronalen Netze von der Struktur (*„hardware“*) auf die Funktion zu schließen. Im Ansatz der künstlichen Intelligenz ist diese *„bottom-up“* Schlussweise explizit verboten (z.B. bei [Marr, 1982]). Umgekehrt folgt aus dieser Überlegung, dass künstliche neuronale Netze nur dann funktionieren können, wenn sie auf solchen Anpassungen aufbauen. Tun sie das nicht, so verlieren sie nicht nur die durch das Wort *neuronal* bezeichnete Motivation, sondern höchstwahrscheinlich auch ihre Anwendbarkeit: man kann vom Gehirn nur das lernen, was es selbst auch mit einiger Wahrscheinlichkeit tut. Abschnitt 3.2 fasst einige wichtige anatomische und physiologische Eigenschaften von Nervensystemen zusammen.

Informationsverarbeitungsprobleme für neuronale Netze. Da das Gehirn kein universeller Rechner ist, können auch nicht für alle Informationsverarbeitungsprobleme gute „neuronale“ Implementierungen existieren. Welche Probleme treten nun bei der Aufgabe „umweltbezogene Verhaltensorganisation“ auf? Interessante Fälle sind sicher die sog. *early vision* Probleme (Bildsegmentierung, Bewegungssehen, Tiefensehen, optische Navigation etc.), die sich biologischen Organismen tatsächlich stellen und die neurobiologisch z.T. auch schon sehr gut untersucht sind (insb. das Bewegungssehen). Ähnliche Probleme in anderen Sinnesmodalitäten sind z.B. das Richtungshören, die Unterscheidung verschiedener Schallquellen („Party-Effekt“) oder das Ertasten von Formen und Oberflächeneigenschaften. Im motorischen Bereich sind etwa Bewegungskoordination oder Manipulation zu nennen. Probleme, die primär auf das Erstellen von Repräsentationen abzielen oder für sich genommen bedeutungslose logische Probleme (wie z.B. das XOR-„Problem“) sind unter dem Gesichtspunkt der Verhaltensorganisation kaum gute Anwendungen oder auch nur Spielprobleme für die neuronale Informationsverarbeitung.

3.2 Natürliche neuronale Netze

In diesem Abschnitt werden einige wichtige Eigenschaften von natürlichen Nervensystemen kurz dargestellt, soweit sie für die Modellbildung bedeutsam sind. Dabei soll nicht bloß (wie heute in vielen Darstellungen üblich) eine Motivation für einen „neuronalen Stil" gegeben werden; vielmehr ist es gerade diese Anatomie und Physiologie, von der man für die Informationsverarbeitung lernen will. Für ein tiefergehendes Studium sei auf die zitierten Lehrbücher verwiesen.

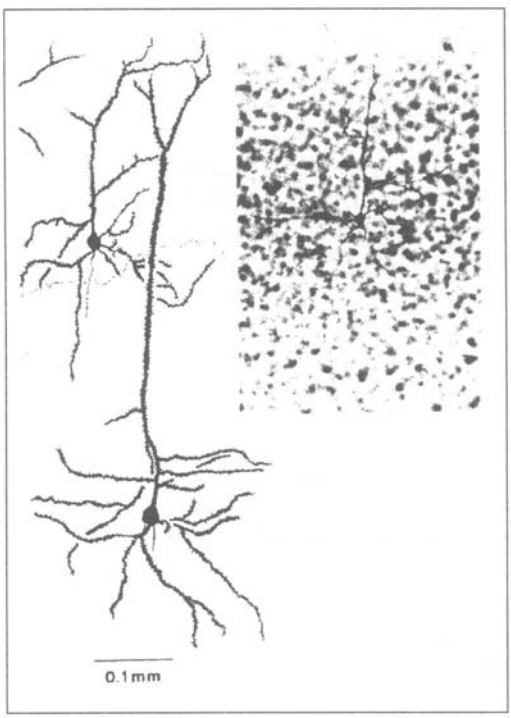
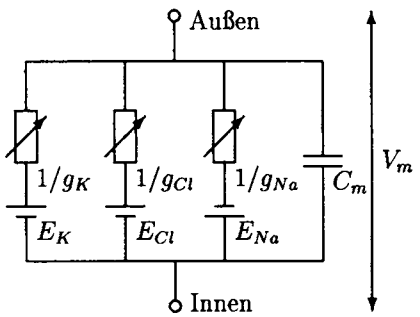


Abbildung 3.2:
Nervenzellen (Pyramidenzellen) aus dem visuellen Cortex. Die dicken, mit Dornen (*„spines"*) besetzten Fasern sind Dendriten, die dünnen Fasern im Bereich der linken Zelle Axone. Das Einschaltbild rechts zeigt ein Foto einer weiteren Pyramidenzelle vor dem Hintergrund der hier ebenfalls sichtbaren Somata der übrigen Nervenzellen; es vermittelt einen Eindruck von der Dichte des Gewebes. Aus [Braitenberg und Schüz, 1998]

3.2.1 Das Nervensystem besteht aus diskreten Zellen

Neuronentypen. Abb. 3.2 zeigt einige Nervenzellen (sog. Pyramidenzellen) aus dem visuellen Cortex einer Maus [Braitenberg und Schüz, 1998]. Sie bestehen aus folgenden Elementen: (1.) Der *Zellkörper (Soma)* enthält den Zellkern und wickelt die meisten biochemischen Synthesen ab. (2.) *Dendriten* sind relativ kurze und stark verzweigte Fortsätze des Somas mit (meist) nicht erregbarer Membran (s.u.). (3.) Das *Axon* ist die eigentliche leitende Nervenfaser, ein Fortsatz des Somas mit erregbarer Membran (Aktionspotentiale). Es kann bis zu mehreren Metern lang werden und ist wenig verzweigt. (4.) Die *Synapsen* sind die (gerichteten) Verbindungsstellen zwischen den Nervenfasern, genauer zwischen den axonalen Enden einer „präsynaptischen" und den Dendriten einer „postsynaptischen" Zelle. Durch den Erregungsfluss Dendrit → Soma → Axon erhält die Nervenzelle eine polare (gerichtete) Organisation.

Membranpotential. Wichtig für das Verständnis der Erregbarkeit von Nervenzellen ist die Tatsache, dass über der Membran der ganzen Zelle eine Spannungsdifferenz von ca. -70 mV (innen negativ) besteht, die auf die ungleiche Verteilung von Na^+, K^+, Ca^{2+} und Cl^--Ionen sowie der Anionen saurer Proteine zurückgeht. Jeder dieser Ionenverteilungen lässt sich über die Nernstsche Gleichung ein Gleichgewichtspotential zuordnen. Insgesamt befindet sich das System jedoch nicht im thermodynamischen Gleichgewicht; vielmehr werden die Ionenverteilungen von der sog. Kalium-Natrium-Pumpe unter Energieverbrauch aktiv erzeugt. Die Pumpe arbeitet gegen einen Leckstrom, der auf das Durchtreten einzelner Ionen durch die Membran zurückgeht. Man kann die elektrischen Verhältnisse am besten anhand eines Ersatzschaltbildes der Membran verstehen, in dem die Ionenverteilungen als Spannungsquellen, die Ionendurchtritte durch die Membran als Widerstände, und die Membran selbst als Kapazität dargestellt werden (Abb. 3.3a).



a.

	Außen	Membran	Innen
	Ionen-konz. [mM]	Nernst-potential [mV]	Ionen-konz. [mM]
K^+	5,5	+ 90 −	150
Cl^-	125	+ 70 −	9
Na^+	150	− 60 +	15

b.

Abbildung 3.3: **a.** Ersatzschaltbild der Membran der Nervenzelle nach [Katz, 1971] (g_x Leitfähigkeit, E_x Nernstpotential, C_m Membrankapazität, V_m Membranpotential). **b.** Konzentrationen der drei wichtigsten Ionenarten und damit verbundene Gleichgewichts- (Nernst-)potentiale. Nach [McGeer *et al.*, 1987]

3.2.2 Nervenzellen sind erregbar

Die wichtigste Eigenschaft von Nervenzellen ist die Erregbarkeit. Diese Erregung (Aktivierung) stellt das eigentliche Signal bei der neuronalen Informationsverarbeitung dar.

Spannungsabhängige Kanäle. Ausgelöst durch bestimmte Änderungen des Membranpotentials ändern sich bei der Aktivierung die Leitfähigkeiten der Membran (vgl. Abb. 3.3), was natürlich wieder auf das Membranpotential zurückwirkt. Die Ionenleitfähigkeit der Membran beruht auf bestimmten, in die Membran integrierten Proteinkomplexen (den „Kanälen"), deren Konformation von der Membranspannung abhängt. Obwohl die dynamischen Prozesse etwa bei der Informationsverarbeitung in der Retina oder bei der dendritischen Summation ein große Rolle spielen, soll hier nicht auf die mathematische Beschreibung (Hodgkin-Huxley-Theorie) eingegangen werden. Eine einfache Einführung findet sich etwa bei [Katz, 1971].

Aktionspotential. Ändert sich das Membranpotential etwa durch einen von außen injizierten Strom, so relaxiert die Membran nach kleinen Reizen wieder zum Ruhepotential. Bei großen Depolarisationen (d.h. Abschwächungen des Ruhepotentials) steigt zunächst die Natrium-Leitfähigkeit sprunghaft an, wodurch die Depolarisation weiter verstärkt

wird. Das Membranpotential steigt bis etwa +40 mV an. Durch die Schließung der Natrium-Kanäle und die verzögerte Öffnung von Kalium-Kanälen wird die Membran repolarisiert. Die (relativ geringen) Konzentrationsänderungen werden durch die K-Na-Pumpe wieder ausgeglichen. Den gesamten Zyklus bezeichnet man als Aktionspotential oder *spike*. Es handelt sich um einen binären („alles-oder-nichts") Prozess, der entweder gar nicht oder, bei überschwelliger Reizung (Depolarisation), vollständig und dann mit stets etwa gleicher Maximalamplitude abläuft. Das Aktionspotential dauert ca. 1 ms. Daraus ergibt sich eine maximale Spikefrequenz in der Größenordnung von 1000 Hz. In der Regel ist der Wert aber kleiner, da die Membran nach einem Aktionspotential für einige Millisekunden *refraktär* ist.

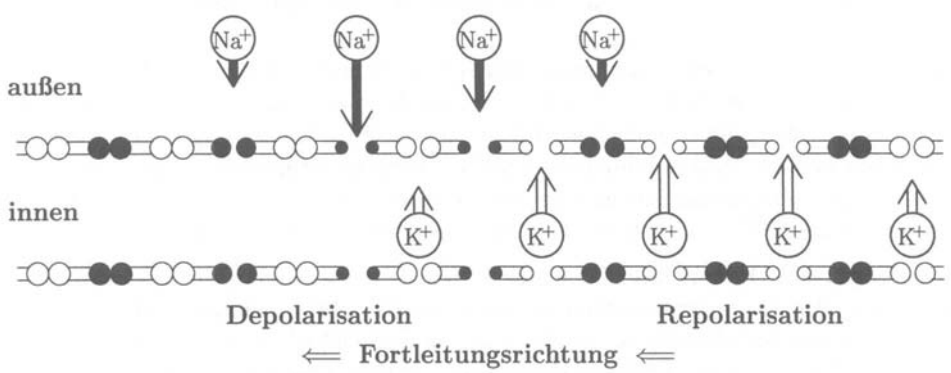


Abbildung 3.4: Schematische Darstellung der Leitfähigkeitsänderungen in der erregbaren Membran eines Axons während eines Aktionspotentials. Zu Beginn (links) werden durch die passiv vorauseilende Depolarisierung die sonst geschlossenen Na^+-Kanäle (•) geöffnet, so dass das Membranpotential schnell ansteigt (Depolarisation). Dadurch werden mit charakteristischer Verzögerung die K^+-Kanäle (○) geöffnet, wodurch (wegen der umgekehrten Verteilung von Na^+ und K^+-Ionen, vgl. Abb. 3.3b) das Membranpotential wieder absinkt (Repolarisation). Damit werden dann auch die spannungsabhängigen Kanäle wieder geschlossen. Die Länge der Pfeile deutet die Leitfähigkeiten g_{Na} bzw. g_K an.

Erregungsleitung. Die bisherige Betrachtung gilt für ein Membranstück an einem festen Ort. In der Nachbarschaft eines erregten Bereichs setzt nun durch passive Leitung (Kabelgleichung) ebenfalls eine Depolarisation ein, die zu einer Ausbreitung der Erregung führt. Im Prinzip erfolgt diese Ausbreitung in beide Richtungen der Nervenfaser; durch die Refraktärzeit wird jedoch erreicht, dass die Erregung stets gerichtet vom Ursprung wegläuft. Eine anschauliche Darstellung der Verhältnisse gibt Abb. 3.4. Durch geschickte Kombination aktiver und passiver Leitung werden Leitungsgeschwindigkeiten von bis zu 100 m/s erreicht.

3.2.3 Synaptische Übertragung

Synapsen. Die Signalübertragung von den axonalen Endverzweigungen eines Neurons auf den Dendriten eines anderen Neurons erfolgt in der Regel durch *chemische Synapsen* (vgl. Abb. 3.5). Diese bestehen aus einer *präsynaptischen Membran* am axonalen Ende der „sendenden" Zelle, dem synaptischen Spalt sowie der *postsynaptischen Membran* der

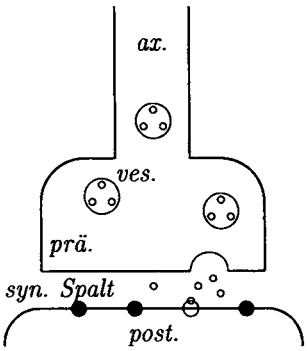


Abbildung 3.5: Schema einer chemischen Synapse. Das Axon der präsynaptischen Zelle *(ax.)* liefert die synaptischen Vesikel *(ves.)* mit dem Neurotransmitter an. Eintreffende Aktionspotentiale führen zu einer Verschmelzung dieser Vesikel mit der präsynaptischen Membran, wobei sich die Transmittermoleküle in den synaptischen Spalt *(syn. Spalt)* ergießen und zur postsynaptischen Membran diffundieren. Durch das Binden des Transmitters an spezifische Rezeptormoleküle (•, ○) werden postsynaptische Ionenkanäle geöffnet oder kompliziertere biochemische Reaktionen ausgelöst.

Empfängerzelle. Durch ein Aktionspotential kommt es in der präsynaptischen Endigung zur Ausschüttung eines *Neurotransmitters*, der auf der postsynaptischen Membran durch entsprechende *Rezeptoren* gebunden wird und in der Regel Änderungen der Membranleitfähigkeit bewirkt. Durch diesen chemischen Übertragungsmechanismus entsteht eine Vielzahl von Interaktionsmöglichkeiten zwischen den Signalen verschiedener Zellen, die für die neuronale Informationsverarbeitung von großer Bedeutung sind.

Postsynaptische Potentiale. Im einfachsten Fall werden durch die Transmittersubstanzen Ionenkanäle in der postsynaptischen (dendritischen) Membran geöffnet. Handelt es sich dabei um Na^+-Kanäle, so ist die Wirkung erregend (depolarisierend); handelt es sich um K^+-Kanäle, so entsteht eine Hemmung (Hyperpolarisation). Postsynaptische Potentiale halten für einige Millisekunden an (bis der Transmitter abgebaut oder resorbiert ist) und breiten sich auf dem Dendriten in der Regel nur passiv aus.

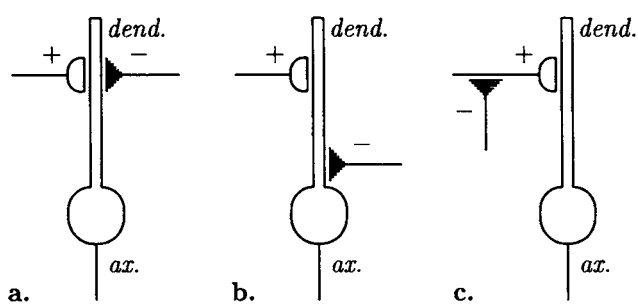


Abbildung 3.6: Dendritische Summation. **a.** Lineare Superposition bei benachbarten Synapsen. **b.** Proximale Hemmung. **c.** Präsynaptische Hemmung. *dend.:* Dendrit, *ax.:* Axon der postsynaptischen Zelle. „−" hemmende, „+" erregende Synapsen

Dendritische Summation. Typische Cortexzellen empfangen in der Größenordnung von 10^4 Synapsen, deren postsynaptische Potentiale in komplexer Weise orts-zeitlich integriert werden. Ein Aktionspotential wird nach Maßgabe dieser Summation am Anfang des Axons ausgelöst, wenn das aus allen synaptischen Einflüssen resultierende „Generatorpotential" die notwendige Schwelle übersteigt. Streng genommen müsste zur Beschreibung dieser Summation die Nervenleitungsgleichung (eine partielle DGL aus der Hodgkin-Huxley-Theorie) für die dreidimensionale Geometrie des Dendritenbaumes gelöst werden. Zusätzlich wird die Situation dadurch kompliziert, dass sich mehr als zwei Fasern zu einem synaptischen Komplex zusammenschließen können, dass es dendro-dendritische Synapsen unter Umgehung des Somas gibt, dass lokal im Dendriten erregbare Membranbereiche auftreten können etc. Qualitativ kann man u.a. folgende Interaktionen finden:

1. *Lineare Summation.* Bei benachbarten Synapsen kann man davon ausgehen, dass die beteiligten postsynaptischen Membranen *äquipotentiell* sind. Die postsynaptischen Potentiale werden dann einfach addiert.
2. *Proximale Hemmung.* Hemmende Synapsen am Ursprung eines größeren Dendriten-Astes können dessen Einfluss vollständig abschalten („Veto").
3. *Präsynaptische Hemmung.* Zuweilen wirken hemmende Synapsen bereits präsynaptisch, d.h. sie sitzen einem Axon auf. Offensichtlich beteiligen sich solche Synapsen nicht direkt an der dendritischen Summation. Sie werden nur indirekt wirksam, indem sie einkommende Aktionspotentiale „abfangen" (sog. *silent* oder *shunting inhibition*). Hierbei handelt es sich um eine multiplikative Nichtlinearität.

3.2.4 Lernen und synaptische Plastizität

Lernen ist Verhaltensänderung. Wenn Informationsverarbeitung die situationsgerechte Organisation von Verhalten ist, dann ist Lernen eine Veränderung dieser Organisation, sichtbar letztlich als eine Änderung des Verhaltens selbst. Dem Begriff des Lernens (auf der Verhaltensseite) steht auf der Seite der Struktur der der *Plastizität*, d.h. der Strukturänderung, gegenüber. Im Sinne der Rückführung der Funktion des Gehirns auf seine Struktur sollte Plastizität die Grundlage des Lernens sein. In künstlichen neuronalen Netzen wird jedoch der Begriff der Plastizität meist vorschnell mit dem Lernen (und oft auch noch mit dem Gedächtnis) gleichgesetzt. Diese reduktionistische Identifizierung des Phänomens Lernen mit dem vermuteten Mechanismus ist einem wirklichen Verständnis der Zusammenhänge eher hinderlich.

Habituation (Gewöhnung). Ein Modell für synaptische Plastizität ist die Habituation (Gewöhnung) des Kiemen-Rückziehreflexes des Seehasen *Aplysia*, einer marinen Schnecke. Es handelt sich um einen monosynaptischen Reflex, bei dem die Reizung des Siphos (Einstromöffnung) mit dem Rückzug der empfindlichen Kieme beantwortet wird. Wiederholt man diese Reizung jedoch zu oft, so ändert sich die Effizienz der beteiligten Synapse(n) und die Reaktion lässt nach. Umgekehrt kann der Reflex durch schmerzhafte Reize am Fuß des Tieres verstärkt werden. Die Neurone und Synapsen, die an diesen Vorgängen beteiligt sind, sind sehr genau bekannt. Durch häufige Reizung werden in der Präsynapse Ionenkanäle abgebaut, die normalerweise den Einstrom positiver Ionen beim Eintreffen eines Aktionspotentials ermöglichen. Dadurch wird die Wirkung dieses Aktionspotentials und letztlich die Transmitterausschüttung abgeschwächt. Durch präsynaptische Kontakte anderer Neurone kann der Abbau der Kanäle rückgängig gemacht werden. Die Postsynapse ist an diesen Vorgängen nicht beteiligt.

Assoziative Modifikation von Synapsen. Eine Grundidee der künstlichen neuronalen Netze besteht in der Modifikation der synaptischen Übertragungsstärke durch (erfolgreichen) Gebrauch (Klassische Konditionierung, Hebb-Regel). Dabei wird eine Synapse dann verstärkt, wenn sie (1.) eine Erregung übertragen hat und (2.) die postsynaptische Zelle daraufhin aktiv wird („feuert"). Es ist instruktiv, sich die Verhältnisse anhand der *klassischen Konditionierung* zu überlegen (Abb. 3.7). In diesem Beispiel wird die „Synapse" w_1 verstärkt, wenn (1.) der konditionierte Reiz CS geboten und (2.) die Zelle R (durch

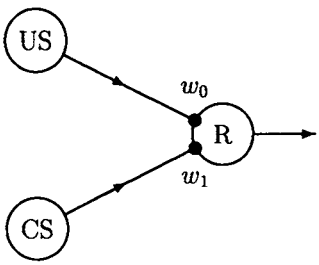


Abbildung 3.7: Modell der klassischen Konditionierung mit Hilfe von Hebbschen Synapsen. CS: *conditioned stimulus*, US: *unconditioned stimulus*, R: *Response*, w_0, w_1: synaptische Übertragungsgewichte. Vor der Konditionierung wird die Reaktion R nur durch den unkonditionierten Reiz hervorgerufen. Bietet man nun hinreichend oft CS und US gemeinsam, so wird w_1 verstärkt, da CS und R gleichzeitig aktiv sind. Schließlich kann CS die Zelle R alleine treiben. Dieses Modell setzt voraus, dass eine Verbindung zwischen CS und R bereits existiert.

gleichzeitige Darbietung von US) aktiv ist. Man bezeichnet diese Regel für die synaptische Modifikation als Hebb-Regel.[1]

Ähnliches Verhalten einzelner Synapsen ist in Form der sog. *Long-Term Potentiation, LTP* im Hippocampus und anderen Teilen der Großhirnrinde beschrieben worden (vgl. etwa [Brown *et al.*, 1990]). Hierbei steigt die Effizienz der beteiligten Synapse(n) an, wenn einlaufende Aktionspotentiale die Aktivierung der postsynaptischen Zelle nach sich ziehen. Erregt man die Zelle elektrisch, ohne präsynaptischen Reiz, oder unterbindet man die Erregung trotz präsynaptischer Aktivität, indem man die Spannung mit einer geeigneten elektronischen Schaltung *(„voltage clamp")* konstant hält, so bleibt die Bahnung (d.h. die Verstärkung der Synapse) aus. Die *long-term potentiation* hält über einige Stunden an und klingt dann wieder ab.

LTP und Hebb-Regel. LTP tritt außer im Hippocampus auch in anderen Teilen der Großhirnrinde auf und stellt den aussichtsreichsten Kandidaten für ein physiologisches Äquivalent zur Hebbschen Plastizität dar. Es gibt aber eine Reihe von Hinweisen darauf, dass die Verhältnisse in Wirklichkeit komplizierter sind (vgl. [Brown *et al.*, 1990; Shepherd, 1990]). So tritt LTP bei manchen Neuronen nur dann auf, wenn zuvor hemmende Synapsen unwirksam gemacht werden (durch Blockierung ihrer Neurotransmitter mittels bestimmter Substanzen). Dies weist darauf hin, dass die Plastizität solcher Synapsen „abschaltbar" ist. Ein weiteres Problem ist die Spezifität der LTP: wird wirklich nur die eine Synapse verstärkt, für die die Hebb-Bedingung erfüllt ist, oder können im Erfolgsfall auch dendritisch oder axonal benachbarte Synapsen mitbetroffen sein? Nicht zuletzt wegen seiner möglichen Bedeutung für die Informationsverarbeitung in neuronalen Netzen ist LTP eines der aktivsten Forschungsgebiete im Bereich der zellulären Neurobiologie.

Bisher war im Wesentlichen von den Eigenschaften einzelner Zellen und ihrer Teile die Rede. Für das Verständnis neuronaler Informationsverarbeitung sind jedoch eine Reihe abstrakterer Eigenschaften von Nervensystemen wichtig, auf die in den folgenden Abschnitten eingegangen werden soll.

[1] Donald O. Hebb (1949). Kanadischer Psychologe.

3.2.5 Rezeptive Felder beschreiben das Reiz-Reaktions-Verhältnis

Spezifität. Nervenzellen in den sensorischen Teilen des Gehirns beantworten bestimmte Sinnesreize stärker als andere. Im Allgemeinen reagieren sie nur auf Reize einer Modalität, wir beschränken uns hier auf das Sehen. Im primären visuellen Areal (V1 = Area 17) der Großhirnrinde (*Cortex*) findet man z.B. Zellen, die spezifisch auf ganz bestimmte Eigenschaften oder Teilaspekte der Sinnesreize reagieren. Typischerweise findet man folgende Spezifitäten:

1. *Position:* Die Zellen können jeweils nur von bestimmten Bereichen des Gesichtsfeldes aus erregt werden. Dieser Bereich ist das „rezeptive Feld" in der ursprünglichen Wortbedeutung.
2. *Orientierung:* Zur Reizung verwendet man häufig helle Rechtecke („Balken") auf dunklem Hintergrund. In diesem Fall hängt die Reaktion der Zelle von der Orientierung dieses Balkens ab.
3. *Ortsfrequenz:* Als Reizmuster können auch sinusförmig modulierte Grauwertverteilungen (Streifenmuster) verwendet werden. In diesem Fall hängt die Reaktion von der Ortsfrequenz und der Orientierung ab.
4. *Bewegung:* Die Reaktion praktisch aller Zellen hängt vom Zeitverlauf der Reizung ab. So werden etwa bewegte Reize bestimmter Geschwindigkeiten oder Richtung bevorzugt oder das Ein- und Ausschalten eines Reizmusters unterschiedlich beantwortet.
5. *Farbe:* Schließlich hängt die Reaktion häufig von der Farbe des Reizes ab.

Diese Liste ist natürlich nicht vollständig. Für alle diese Reizparameter kann man sogenannte *tuning*-Kurven messen, bei denen die Reaktion als Funktion des Reizparameters angegeben wird. Diese Abstimmkurven sind in der Regel nicht sehr scharf, so kann die Halbwertsbreite bei der Orientierungsspezifität durchaus 45° und mehr betragen. Es wird daher diskutiert, ob die Spezifitäten tatsächlich einzelnen Zellen oder eher größeren „Populationen" zukommen (Populationskodierung).

Modelle rezeptiver Felder. Viele Eigenschaften rezeptiver Felder lassen sich aus einer orts-zeitlichen Gewichtsfunktion w ableiten, die für jeden Reizort (x, y) und für jeden Zeitpunkt t vor der Erregung angibt, wie stark die Zelle einen kurzzeitig aufblinkenden Lichtpunkt beantwortet. Man denkt sich dann einen beliebigen Reiz aus vielen Punkten wie ein Fernsehbild zusammengesetzt, liest für jeden einzelnen Punkt die Reaktion aus der Gewichtsfunktion ab und erhält im linearen Fall die Reaktion der Zelle durch Summation (bzw. Faltung). Derartige Modelle sind als Näherungen hilfreich, müssen aber bei nichtlinearen Reiz-Reaktions-Zusammenhängen, wie sie vor allem bei den sog. „komplexen" Zellen auftreten, modifiziert werden. Typische Gewichtsfunktionen, die für die Modellierung benutzt werden, sind etwa die folgenden. Dabei sind Verschiebungen und Drehungen nicht angegeben.

1. Differenz von Gauß-Kurven (*DoG = Difference of Gaussians*)

$$w(x,y,t) = \sum_{i=1}^{2} \frac{A_i}{\tau_i} \exp\left\{-\frac{1}{2}\left(\frac{x^2}{B_{x,i}^2} + \frac{y^2}{B_{y,i}^2}\right)\right\} \exp\left\{-\frac{t}{\tau_i}\right\} \tag{3.1}$$

Dabei ist A_i die Amplitude (positiv oder negativ), $B_{x,i}$ die Breite der Gauß-Kurve in x-Richtung, $B_{y,i}$ die Breite in y-Richtung und τ_i die Zeitkonstante des für das Zeitverhalten der Einfachheit halber angenommen Tiefpasses.

2. Gabor-Funktionen

$$w(x,y) = A \exp\left\{-\frac{1}{2}\left(\frac{x^2}{B_x^2} + \frac{y^2}{B_y^2}\right)\right\} \begin{matrix}\sin\\ \cos\end{matrix} (\omega_x x + \omega_y y) \tag{3.2}$$

Zusätzlich zu den obigen Bezeichungen sind ω_x und ω_y lokale Ortsfrequenzen. Gerade Gabor-Funktionen benutzen den cos-Term, ungerade den sin-Term; in der Realität findet man meist Kombinationen dieser beiden Grundtypen.

Einfache Spezifitäten (Orientierung, Richtung, Geschwindigkeit) lassen sich bereits mit solchen linearen Modellen simulieren.

Die „Großmutterzelle". In Analogie zur Bildverarbeitung werden rezeptive Felder häufig als Merkmalsdetektoren aufgefasst, die Informationen über lokale Bildelemente (*„features"*) extrahieren (vgl. die klassischen Arbeiten von [Hubel und Wiesel, 1962]). Während dies für frühe Verarbeitungsstufen durchaus angemessen erscheint (wenn man den *feature*-Begriff auf zeitliche Phänomene erweitert), führt dieser Ansatz bei höheren Verarbeitungsstufen zu Schwierigkeiten. Werden die *features* einfach immer komplizierter, bis das Feuern einer Zelle für das Erkennen einer ganzen Szenerie steht? Für diese Idee spricht die Tatsache, dass die rezeptiven Felder tatsächlich größer werden und dass sog. Gesichterzellen, die hochspezifisch auf Gesichter reagieren, wirklich gefunden wurden. Dagegen spricht ein logisches Problem: je detaillierter die Reizbeschreibung wird, für die die Erregung einer Zelle steht, umso mehr solcher Zellen werden gebraucht. Wenn tatsächlich eine „Großmutterzelle" dafür zuständig ist, die Gegenwart der alten Dame anzuzeigen, was passiert, wenn sie den Hut abnimmt?

Neuronale Kodierung: „spezifische Sinnesenergien." Da alle Aktionspotentiale in etwa gleich sind, kann in der Form der Erregung keine Information kodiert sein. So führt z.B. eine mechanische Reizung der Netzhaut zu einer Sehwahrnehmung („Sternchen"), weil sich die mechanisch ausgelöste Erregung nicht von der „adäquat", d.h. optisch ausgelösten unterscheidet[2]. Kodierungsmöglichkeiten im Nervensystem ergeben sich aus den Spezifitäten rezeptiver Felder sowie vielfältigen orts-zeitlichen Mustern von Erregungen. So ist z.B. die Reizstärke häufig in der momentanen *spike*-Rate (Kehrwert des Zeitintervalls zwischen zwei Aktionspotentialen) kodiert.

3.2.6 Neuroanatomie des visuellen Systems

Die Sehbahn. Die ersten Stationen der visuellen Informationsverarbeitung, bei denen die Erregungen an den Somata der beteiligten Nervenzellen umgeschaltet werden, sind die Retina, das *Corpus geniculatum laterale (LGN*[3]) im Zwischenhirn und der primäre visuelle Cortex im Großhirn. Dazwischen liegen Faserverbindungen, die aus den Axonen

[2]Dies ist das von Johannes Müller bereits im 19. Jahrhundert formulierte „Gesetz der spezifischen Sinnesenergien". In der neueren Literatur findet man dafür die Bezeichung *„labeled line coding."*

[3]LGN ist die Abkürzung der englischen Bezeichnung *„lateral geniculate nucleus"*.

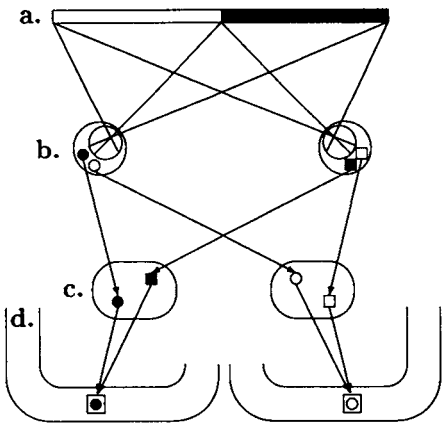


Abbildung 3.8:
Schema der Sehbahn eines Säugetieres. Das Gesichtsfeld (**a.**) wird auf die Retinae der beiden Augen (**b.**) abgebildet. In jedem Auge sind zwei Ganglienzellen angedeutet: •, ○: linkes Auge. ■, □: rechtes Auge. ○, □: linkes Gesichtsfeld. •, ■: rechtes Gesichtsfeld. Die Axone der nasal gelegenen Zellen (○,■) kreuzen im Chiasma opticum zur gegenüberliegenden Hirnhälfte, wo sie in das Corpus geniculatum laterale (LGN, **c.**) eintreten. Die Zellen des LGN werden jeweils nur von einem Auge innerviert. Erst bei der folgenden Projektion auf den visuellen Cortex (**d.**) konvergieren die Signale beider Augen auf *binokulare* Zellen.

des jeweils vorhergehenden Kerngebietes gebildet werden: Sehnerv, Chiasma und optischer Trakt zwischen Retina und LGN sowie die optische Radiation zwischen LGN und Cortex. Eine Übersicht gibt Abb. 3.8.

Zelluläre Organisation des visuellen Cortex. Der Cortex ist ein etwa 1–2 mm dickes, flächig ausgedehntes Netzwerk mit einer Zelldichte von etwa 10000 pro mm^3. Insgesamt besteht er aus etwa 10^{11} Neuronen. Während die tangentiale Organisation mehr oder weniger uniform ist, gibt es in der Vertikalen deutliche Variationen der Netzwerkparameter. Mit neuroanatomischen Methoden lassen sich typischerweise sechs Schichten unterscheiden, die sich durch unterschiedliche Verteilung von Zelltypen und Konnektivitäten auszeichen. Auch innerhalb der Schichten bestehen komplexe Verbindungsmuster; sie werden daher nur unzureichend durch die „Monolayer" etwa eines mehrschichtigen Perzeptrons modelliert. Vgl. Abb. 3.2.

Columnäre Strukturen. Auf etwas gröberem Niveau findet man eine Reihe von Ordnungsprinzipien, die unter dem Begriff der *Columne* zusammengefasst werden. Dabei variieren Verschaltungsmuster oder Eigenschaften der rezeptiven Felder in mehr oder weniger regelmäßiger Weise über horizontale Abstände in der Größenordnung von 0,1 bis etwa 5 mm. Beispiele sind die Okulardominanzstreifen, die sich durch Entmischung der Eingänge aus den beiden Augen im visuellen Cortex bilden, die Orientierungskolumnen, die einer stetigen Variation der Vorzugsorientierung der rezeptiven Felder über dem corticalen Ort entsprechen, oder die Segregation des Outputs, bei der die Zellen bereichsweise in verschiedene Folgeareale projizieren.

Areale. In der Größenordnung von einigen Millimetern bis zu Zentimetern ist der Cortex in sog. *Areale* eingeteilt, die ursprünglich durch ihre zelluläre Organisation definiert wurden. Die meisten der (bei Primaten) ca. 20 bekannten visuellen Areale enthalten jeweils eine mehr oder weniger vollständige Repräsentation des Gesichtsfelds in Form einer *retinotopen Karte*. Das bedeutet, dass sich die Positionen der rezeptiven Felder im Gesichtsfeld stetig mit der corticalen Position der zugehörigen Zelle verschieben. Umgekehrt kann man sagen, dass benachbarte Punkte des Gesichtsfelds auf benachbarte Punkte des Cortex abgebildet werden. Zwischen den Arealen gibt es viele (aber nicht ausschließlich) vollparallele Projektionen, meist in beiden Richtungen. Die topographische Organisation

Tabelle 3.1: Einige neuronal implementierbare Rechenoperationen

Rechenoperation	Physiologie
Orts-Zeitliche Summation	Superposition von postsynaptischen Potentialen
Laterale Inhibition	Summation über geeignet angeordnete Synapsen
Zeitliche Ableitungen	Verzögerte Hemmung, dynamische Lösung der Kabelgleichung
logisches „und nicht“	Präsynaptische Hemmung, Veto durch proximale Hemmung
Schwellen, logisches „und“	*spike*-Auslösung
Adaptation, Speicherung	Modifizierbare Synapsen (LTP)
Modulation des Netzwerkes	Ausschüttung weitreichender Neurotransmitter
Orts-Zeitliche Filterung	Rezeptive Felder
Ortsrepräsentation	Rezeptive Felder, neuronale Karten

entsteht im Wesentlichen durch die regelmäßige Anordnung der Inputfasern. Die intrinsische Organisation ist auf dem Zentimeter-Niveau in etwa uniform.

Einige neuronal implementierbare Rechenoperationen sind in Tabelle 3.1 zusammengestellt.

3.3 Künstliche neuronale Netze

Dieses Kapitel versucht einen systematischen Durchgang durch die im Bereich der künstlichen neuronalen Netze verwendeten Methoden und Modellbildungen. Beispiele und Anwendungen auf spezifische Informationsverarbeitungsprobleme werden in den folgenden Abschnitten besprochen.

3.3.1 Elemente neuronaler Netze

Grundgrößen für die Modellierung neuronaler Netze sind die Aktivität (Erregung) mit der zugehörigen Erregungsdynamik, die Übertragungsgewichte mit der zugehörigen Gewichtsdynamik und die Netzwerktopologie, deren Dynamik bisher nur in Ansätzen untersucht worden ist.

1. Erregung und Erregungsdynamik
   (a) *Erregungszustand (Aktivität):* Für ein Neuron i bezeichne e_i den Erregungszustand. Er kann z.B. die disketen Werte $\{-1, 1\}$, $\{0, 1\}$ oder kontinuierliche Werte aus $[0, 1]$ oder $\mathbb{R}$ annehmen. Der Vektor $\mathbf{e} := (e_1, e_2, ..., e_I)^\top$ bezeichnet den Erregungszustand des ganzen Netzes. Im zeitkontinuierlichen Fall sind e_i und $\mathbf{e}$ Funktionen von der Zeit t. Im zeit- und ortskontinuierlichen Fall geht der Vektor $\mathbf{e}$ in die globale Erregungsfunktion $e(\mathbf{x}; t)$ über.
   (b) *Aktivierungsfunktion (Übertragungsfunktion):* Die Aktivierungsfunktion $\alpha_i(\mathbf{e})$ beschreibt den zukünftigen Erregungszustand der Zelle i lokal durch die Zustände der benachbarten (= verbundenen) Zellen $\{e_j\}$. Im kontinuierlichen Fall entsteht ein „Aktivierungsfunktional“.

(c) *Globale Erregungsdynamik:* Für bestimmte Probleme (z.B. Stabilitätsanalysen) kann man sogenannte „Netzwerkenergien" definieren, die die Erregungsdynamik global beschreiben.

2. Gewichte und Gewichtsdynamik
   (a) *Übertragungsgewichte:* Die Verbindung zwischen zwei Neuronen e_i, e_j wird mit einem Gewicht w_{ij} versehen, das in die entsprechende Aktivierungsfunktion eingeht. Da die Verbindungen gerichtet sind, gilt im Allgemeinen $w_{ij} \neq w_{ji}$ (unsymmetrische Netze). Wir wollen w_{ij} als Gewicht des Übergangs $i \leftarrow j$ auffassen, d.h. in Richtung des „Einsammelns" von Erregung. Im diskreten Fall bilden die w_{ij} eine quadratische Matrix W; im kontinuierlichen Fall bilden die Gewichte den Kern $W(\mathbf{x}, \mathbf{x}'; t, t')$ eines Aktivierungsfunktionals.
   (b) *Lernregeln:* Lokale Regeln für die Verstellung von Gewichten bezeichnet man als „Lernregeln". Sie hängen nur vom gegenwärtigen Gewicht der Verbindung sowie den Erregungszuständen der beteiligten Zellen ab (unüberwachtes Lernen).
   (c) *Globale Gewichtsdynamik:* Für manche Anwendungen können globale Anforderungen an das Netzwerk formuliert werden, die dann zur Einstellung der Gewichte benutzt werden. Solche Anforderungen werden häufig als Lehrer-Signal bezeichnet (überwachtes Lernen). Daneben gibt es Zwischenformen wie Wettbewerbslernen und Verstärkungslernen (s.u.).

3. Netzwerktopologie
   Neuronale Netze sind in der Regel nicht vollständig verknüpft, sondern weisen komplexere Strukturen (z.B. Schichten u.a. Subpopulationen) auf. Die Dynamik dieser Netzwerktopologie ist bisher nur in Ansätzen untersucht.

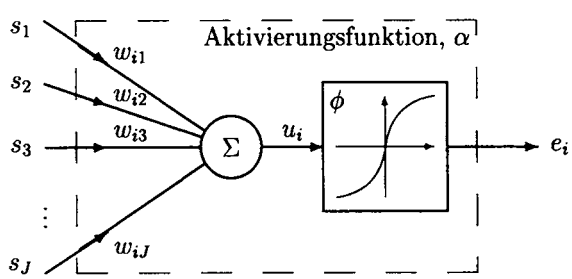


Abbildung 3.9:
Einfaches Modellneuron für neuronale Netze. s_j: Eingänge (Stimuli), w_{ij} Übertragungsgewichte, Σ Summation, u_i Potential ($u_i = \Sigma_{i=1}^{J} w_{ij} s_j$), ϕ statische Nichtlinearität („Kennlinie"), e Erregung ($e_i = \phi(u_i)$).

3.3.2 Erregungsdynamik

3.3.2.1 Aktivierungsfunktionen

Eine Aktivierungsfunktion α ordnet einem Satz von Eingangssignalen s_j eine Erregung e_i zu. Dabei gehen in der Regel Übertragungsgewichte w_{ij} sowie eine nichtlineare Kennlinie ϕ ein, seltener auch nichtlineare Interaktionen der Inputs. Im Allgemeinen sind die s_j gerade die Erregungen der übrigen Zellen des Netzwerkes, im zeitdiskreten Fall etwa zum vorherigen Zeittakt.

Logische und semilineare Neurone. [McCulloch und Pitts, 1943] führen die als „logisches Neuron" bekannte Regel

$$e_i = \phi\left(\sum_j w_{ij}s_j - \theta\right) \quad \text{mit} \quad \phi(u) := \begin{cases} 0 & \text{für } u < 0 \\ 1 & \text{für } u \geq 0 \end{cases} \tag{3.3}$$

ein, wobei e_i, die w_{ij} und θ diskret sind ($e_i \in \{0,1\}$, $w_{ij} \in \{-1,1\}$, $\theta \in \mathbb{N}$). Widrow & Hoff lassen kontinuierliche Gewichte zu (*„Adaptive Linear Neuron, ADALINE"*). Der Schwellwert wird wie ein konstanter Eingang mit Gewicht $w_{i0} = -\theta$ behandelt, so dass eine einheitliche Behandlung der Adaptivität möglich wird [Widrow und Hoff, 1960]. Kontinuierliche Werte für alle beteiligten Größen und stetige oder differenzierbare Nichtlinearitäten charakterisieren das sogenannte „semilineare Neuron". Die Aktivierungsfunktion sowie einige häufig benutzte Kennlinien lauten:

$$e_i = \phi\left(\sum_j w_{ij}s_j\right) \quad \text{mit} \tag{3.4}$$

$$\phi(u) = \frac{2}{\pi}\arctan(\lambda u); \quad \phi(u) = \tfrac{1}{1+\exp\{-\lambda u\}}; \quad \text{oder} \quad \phi(u) = \begin{cases} 0 & \text{für } u < 0 \\ x & \text{für } 0 \leq x \leq 1 \\ 1 & \text{für } u > 1 \end{cases}$$

Während diskrete Erregungsgrößen dem Alles-oder-Nichts Charakter der Aktionspotentiale entsprechen, werden kontinuierliche Größen als momentane *spike*-Raten interpretiert.

Explizite Repräsentation von Ort und Zeit. Zur Beschreibung dynamischer Vorgänge in kontinuierlicher Zeit eignen sich Aktivierungsfunktionen der Form

$$\frac{du_i}{dt}(t) = -cu_i(t) + \sum_j w_{ij}s_j(t-\tau_j) - \theta; \quad e_i(t) = \phi(u_i(t)). \tag{3.5}$$

Dabei bezeichnen c die Zeitkonstante und $\tau_j \in \mathbb{R}_0^+$ Latenzzeiten (vgl. [Hirsch, 1989]). Will man überdies auch den Ort explizit berücksichtigen (indem man die Neurone nicht mehr durchnumeriert, sondern mit einem Punkt im Raum identifiziert), so erhält man Integralgleichungen der Form

$$e(\mathbf{x}) = \int W(\mathbf{x},\mathbf{x}')s(\mathbf{x}')d\mathbf{x}'. \tag{3.6}$$

Durch Kombination mit Gl. 3.5 entstehen Integro-Differential-Gleichungen, wie sie etwa von [Wilson und Cowan, 1973] oder [Mallot und Giannakopoulos, 1996] verwendet werden.

Probabilistische Modelle. Betrachtet man die Stimuli und die Erregung eines Neurons als Zufallsvariable σ_j und η_i mit diskreten Werten in $\{0,1\}$, so erhält man stochastische Aktivierungsfunktionen. In Anlehnung an Ansätze aus der statistischen Physik wählt man etwa:

$$P\{\eta_i(t) = 1\} = \phi_T\left(\sum_j w_{ij}\sigma_j - \theta\right); \qquad \phi_T(u) := \frac{1}{1+e^{-u/T}}. \tag{3.7}$$

Der Parameter T der Verteilung wird dann häufig als „Temperatur" bezeichnet ($\rightarrow$ *simulated annealing*); ϕ_T entspricht einer sigmoiden Kennlinie im deterministischen Fall.

Nichtlineare Interaktion der Eingänge. Aktivierungsfunktionen mit nichtlinearer Interaktion der Eingänge s_j sind z.B. die sogenannte Sigma-Pi Regel (multilineare Neurone) oder präsynaptische Hemmung (Abs. 3.2.3), modelliert als Division eines erregenden und eines hemmenden Eingangsterms.

Abgestimmte Neurone. Bisher wurde die Erregung e_i stets über die Projektion des Reizes auf den Gewichtsvektor definiert. Zuweilen ist es jedoch günstiger, die Erregung als eine Art Ähnlichkeit des Reizvektors mit einem für die Zelle charakteristischen Merkmalsvektor $\mathbf{t}_i = (t_{i1}, ..., t_{iJ})^\top$ zu definieren. Der Vektor $\mathbf{t}_i$ hat die gleiche Dimensionalität wie der Gewichtsvektor, den er ersetzt. Eine häufig verwendete Aktivierungsfunktion ist z.B.:

$$e_i = G(\|\mathbf{s} - \mathbf{t}_i\|) \quad \text{mit} \quad G(x) := \exp\{-\frac{1}{2\sigma^2}x^2\}, \quad \sigma \in \mathbb{R}^+ \tag{3.8}$$

Im Prinzip können für G dabei auch andere „glockenförmige" Funktionen verwendet werden (vgl. [Poggio und Girosi, 1990]).

3.3.2.2 Globale Verfahren

Konsensfunktion und Netzwerkenergie. Bisher haben wir Erregungsdynamiken betrachtet, die lokal mit Hilfe von Aktivierungsfunktionen beschrieben werden. Die Zustandsänderung eines Neurons hängt dabei lediglich vom momentanen Zustand der Neurone ab, von denen es Eingänge erhält. Oft ist man jedoch an der Entwicklung globaler Zustandsgrößen interessiert (z.B. Energien oder Liapunov-Funktionen), z.B. bei der Untersuchung von Konvergenzverhalten oder bei der Konstruktion von Netzen für bestimmte Optimierungsprobleme. Als ein solches globales Maß der Netzwerkaktivität werden Funktionen der Form

$$k(\mathbf{e}) := \frac{1}{2}\sum_{i,j} w_{ij}e_ie_j + \sum_{j=1} w_{j0}e_j = \frac{1}{2}\mathbf{e}^\top W\mathbf{e} + \mathbf{w}_0^\top\mathbf{e} \tag{3.9}$$

untersucht. Dabei bezeichnet $-w_{j0}$ die Schwellen. Diese Funktion nimmt dann große Werte an, wenn zwei Zellen, die durch starke Gewichte w_{ij}, w_{ji} miteinander verbunden sind, auch tatsächlich gleichzeitig aktiv sind. (Man beachte, dass k jeweils nur von den Mittelwerten $(w_{ij} + w_{ji})/2$ abhängt. Daher auch der Faktor 1/2 in Gl. 3.9.) Sind die Schwellen $w_{j0} = 0$, so folgt aus der Cauchy-Schwartzschen Ungleichung sofort, dass k bei gegebener Norm von $\mathbf{e}$ durch die Eigenvektoren von W maximiert wird.

Eine andere Interpretation von k ergibt sich aus der Überlegung, dass für die Aktivierungsfunktion $\dot{\mathbf{e}} = W\mathbf{e} + \mathbf{w}_0$ (vgl. Gl. 3.5) mit symmetrischer Gewichtsmatrix $W = W^\top$ die Beziehung

$$\text{grad}\, k(\mathbf{e}) = \dot{\mathbf{e}} \tag{3.10}$$

gilt. In diesem Fall ist k also das Potential der Netzwerkdynamik. Für unsymmetrische W ist k eine Liapunov-Funktion (d.h. $k(\mathbf{e}(t))$ ist monoton fallend und nach unten beschränkt für jede Lösung $\mathbf{e}(t)$).

Anwendung finden solche globalen Funktionen bei der Boltzmann-Maschine oder in Spin-Glas Modellen (vgl. [Aarts und Korst, 1989]).

Simulated Annealing. Beschreibt man die Erregungsdynamik (oder auch die Gewichtsdynamik) eines neuronalen Netzes durch globale „Energiefunktionen", so stellt sich das Problem der Minimierung solcher Funktionen in Abhängigkeit von der Erregungs- oder Gewichtsverteilung. Handelt es sich um ein diskretes Netz, so ist das Problem endlich und die Optimierungsaufgabe heißt *kombinatorisch.* Ein wichtiger Algorithmus für kombinatorische Optimierung ist das *simulated annealing*, bei dem das langsame Erstarren einer Schmelze und die damit verbundene Minimierung der Energie (oder eines anderen thermodynamischen Potentials) simuliert wird. Wegen seiner Bedeutung sei der Algorithmus kurz skizziert:

1. *Formulierung des Problems:* Ein kombinatorisches Optimierungsproblem besteht aus einer Menge möglicher Zustände (Lösungsraum $\mathcal{E}$) und einer Kostenfunktion $f : \mathcal{E} \longrightarrow \mathbb{R}$.
2. *Erzeugung neuer Zustände:* Zu einem Zustand $\mathbf{e}^{alt}$ des Systems benötigt man eine Regel zur Erzeugung neuer „benachbarter" Zustände $\mathbf{e}^{neu}$ (z.B. Austausch zweier Städte beim *travelling salesman problem*).
3. *Fluktuation:* Ausgehend von einem Zustand $\mathbf{e}^{alt}$ mit Kosten $f(\mathbf{e}^{alt})$ erzeugt man mit der Nachbarschaftsregel 2 einen Zustand $\mathbf{e}^{neu}$. Dieser Zustand wird nach einer probabilistischen Regel als aktueller Zustand übernommen:

$$(3.11)\quad p\{\text{accept } \mathbf{e}^{neu}\} = \begin{cases} 1 & \text{falls } f(\mathbf{e}^{neu}) \leq f(\mathbf{e}^{alt}) \\ \exp\left(\frac{1}{T}(f(\mathbf{e}^{alt}) - f(\mathbf{e}^{neu}))\right) & \text{falls } f(\mathbf{e}^{neu}) > f(\mathbf{e}^{alt}) \end{cases}.$$

   $T \in \mathbb{R}^+$ heißt Kontrollparameter; er entspricht der Temperatur beim Abkühlen.

Man startet mit einer „hohen" Temperatur, entwickelt bis zum Gleichgewicht (d.h. bis $f(\mathbf{e})$ nicht mehr stark schwankt) und erniedrigt langsam die Temperatur, ohne dabei weit vom Gleichgewicht abzuweichen. (Vgl. etwa [Aarts und Korst, 1989].)

3.3.3 Grundtypen von neuronalen Netzen

Eine einheitliche Einteilung der verschiedenen Typen neuronaler Netze hat sich noch nicht durchgesetzt. Nach der Dimensionalität von Input und Output kann man z.B. folgende Unterscheidung treffen:

1. *Klassifikatoren* teilen den Raum der Inputvektoren in Klassen ein; im einfachsten Fall detektiert ein einzelnes Output-Neuron ein Muster einer bestimmten Klasse. $\mathbb{R}^J \to \{0, 1\}$.
2. *Assoziatoren* ordnen bestimmten Inputvektoren bestimmte Outputvektoren zu. Dabei ist die Anzahl der assoziierten Paare relativ klein (z.B. gleich der Dimension des Inputraums) und jedenfalls endlich. Nicht gespeicherte Inputvektoren können mit Zwischenwerten oder mit einem der gespeicherten Outputvektoren beantwortet werden. Im zweiten Fall spricht man von einem klassifizierenden Assoziator. $\{\mathbf{s}^{(1)}, ..., \mathbf{s}^{(P)}\} \subset \mathbb{R}^J \to \{\mathbf{e}^{(1)}, ..., \mathbf{e}^{(P)}\} \subset \mathbb{R}^I$.
3. *Abbildende Netzwerke* implementieren eine Zuordnung zwischen beliebigen Inputvektoren und zugehörigen Outputvektoren. Sie werden benutzt, um Funktionen bzw. Operatoren zu approximieren. $\mathbb{R}^I \to \mathbb{R}^J$.

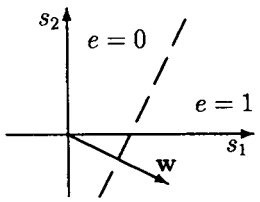


Abbildung 3.10: Interpretation eines Modellneurons mit einfacher Schwelle ($e \in \{0,1\}$) als linearer Klassifikator. Die Trennfläche (gestrichelt) ist durch den Gewichtsvektor $\mathbf{w}$ (ihren Normalenvektor) und die Schwelle $-w_0$ gegeben. Für Inputvektoren $\mathbf{s}$ links dieser Trennlinie ist die Erregung 0, für Inputs rechts der Trennlinie 1.

3.3.3.1 Klassifizierende Netzwerke: Das Perzeptron

Semilineare Neurone sind lineare Klassifikatoren. Ein einzelnes Modellneuron mit der Aktivierungsfunktion

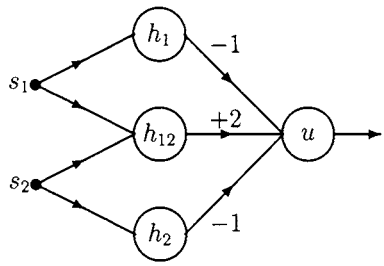
$$e := \begin{cases} 1 & \text{für } \sum_{j=0}^{J} w_j s_j \geq 0 \\ 0 & \text{für } \sum_{j=0}^{J} w_j s_j < 0 \end{cases} \tag{3.12}$$

kann als linearer Klassifikator auf dem Raum der Inputvektoren $\mathbf{s} = (s_1, ..., s_J)^\top$ aufgefasst werden (Abb. 3.10). s_0 ist konstant 1, so dass $-w_0$ die Schwelle des Neurons ist. Das Potential u spielt hier die Rolle einer „Diskriminanzfunktion"(vgl. Abs. 3.4.1); sie hängt (bis auf die Schwelle) linear vom Input $\mathbf{s}$ ab. Wegen $u = \sum_{j=0}^{J} w_j s_j = \mathbf{w}^\top \mathbf{s} + w_0$ ist die Trennfläche der Klassifikation durch die Hyperebene $w_0 + \mathbf{w}^\top \mathbf{s} = 0$ gegeben[4] (Vgl. etwa [Duda und Hart, 1973].)

Lineare Klassifikatoren können kein „exklusives Oder" implementieren. Klassen, deren Trennflächen keine Hyperebenen sind, können so nicht gebildet werden. Will man z.B. das „exklusive Oder" (XOR) zweier Eingänge implementieren, so sollten die Inputvektoren $(0,0)$ und $(1,1)$ in eine, die Vektoren $(1,0)$ und $(0,1)$ jedoch in die andere Klasse fallen. Eine solche Klassifikation ist mit einer Trenn*ebene* aus geometrischen Gründen nicht möglich. Obwohl dies ein eher akademisches Problem ist, das sich etwa in Zusammenhang mit der Frage der Universalität einer Rechenmaschine stellt, zeigt sich bereits hier eine prinzipielle Limitierung des Klassifikationsansatzes. Mögliche Auswege sind (a.) die Entwicklung mehrschichtiger Klassifikatoren („Perzeptrons"), (b.) verbesserte Kodierung der untersuchten Daten in den Inputs (*feature*-Extraktion) oder (c.) die Verwendung nichtlinearer Inputinteraktionen (z.B. Sigma-Pi-Neuronen). Die Verwendung nichtlinearer Aktivierungsfunktionen erinnert an den Polynom-Klassifikator $u = w_0 + \sum_{i=1}^{n} w_i s_i + \sum_{i,j=1}^{n} w_{ij} s_i s_j + ...$ der konventionellen Mustererkennung.

Das dreischichtige Perzeptron. Beim mehrschichtigen Perzeptron wird der Inputvektor $\mathbf{s}$ von mehreren in einer sogenannten Assoziationsschicht oder „verborgenen" Schicht angeordneten Klassifikatoren parallel zu Signalen h_j verarbeitet. Diese Merkmalsdetektoren bilden dann den Input einer zweiten Klassifikatorstufe (Abb. 3.11). Allgemein kann man zeigen, dass ein solches Perzeptron für die Frage „ist die Anzahl der gesetzten Pixel ungerade" mindestens ein Element (Assoziationszelle) enthalten muss, das alle Pixel der Eingangsschicht anschaut. Derartige Aussagen gelten auch für andere Perzeptrons. So kann z.B. die Frage „ist die Figur zusammenhängend" nicht mit lokalen Perzeptrons geklärt werden [Minsky und Papert, 1988].

[4] Da alle Vektoren als Spaltenvektoren aufgefasst werden, ist $\mathbf{w}^\top \mathbf{s}$ das Skalarprodukt von $\mathbf{w}$ und $\mathbf{s}$.



Input		Assoziation			Ausgang	
s_1	s_2		h_1	h_2	h_{12}	u	$\phi(u)$
		$(w_i$	-1	-1	$+2)$		
0	0		0	0	0	0	0
1	0		1	0	1	1	1
0	1		0	1	1	1	1
1	1		1	1	1	0	0

Abbildung 3.11: Einfaches dreischichtiges Perzeptron für die Lösung des XOR-Problems. Rechts die „Wahrheitstafel" für die beteiligten logischen Neurone.

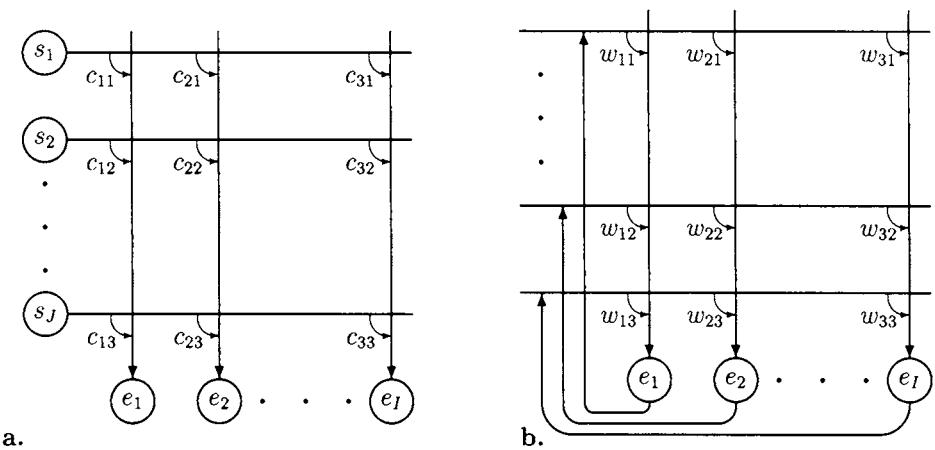


Abbildung 3.12: **a.** Heteroassoziator. **b.** Autoassoziator

3.3.3.2 Assoziierende Netzwerke: Auto- und Heteroassoziator

Der Heteroassoziator besteht aus zwei Zellpopulationen. Wir betrachten ein Netzwerk aus zwei Teilpopulationen (Schichten) $\mathcal{S}, \mathcal{E}$ von Neuronen mit den Erregungsvektoren $\mathbf{s} = (s_1, ..., s_J)^\top$ und $\mathbf{e} = (e_1, ..., e_I)^\top$. Innerhalb der Subpopulationen bestehen keine Verbindungen, während jede Zelle aus $\mathcal{S}$ mit jeder Zelle aus $\mathcal{E}$ verbunden ist; das zugehörige Gewicht sei c_{ij}. (Beachte, dass diese Definition von der bisherigen abweicht: c_{ii} meint nicht die Kopplung einer Zelle auf sich selbst, sondern die Kopplung zweier Zellen mit gleicher Nummer in $\mathcal{S}$ und $\mathcal{E}$.) Mit der linearen Aktivierungsfunktion (ohne Schwelle und Kennlinie) hat man dann:

$$e_i = \sum_{j=1}^{J} c_{ij} s_j, \text{ bzw. } \mathbf{e} = C\mathbf{s} \text{ mit } \{c_{ij}\} = C. \tag{3.13}$$

Bezeichnet man mit W die Verknüpfungsmatrix auf der Menge aller Neurone $\mathcal{F} := \mathcal{S} \cup \mathcal{E}$ (d.h. $\mathbf{f} := (s_1, s_2, ..., s_J; e_1, e_2, ...e_I)^\top$), so erhält man folgenden Zusammenhang:

$$\mathbf{f} \mapsto W\mathbf{f} = \begin{pmatrix} 0_{JJ} & 0_{JI} \\ C & 0_{II} \end{pmatrix} \mathbf{f}. \tag{3.14}$$

Dabei bezeichnet 0_{IJ} eine $I \times J$-Matrix, deren sämtliche Koeffizienten verschwinden.

Verknüpfungsmatrix und äußeres Produkt. Es sei $\mathbf{s}$ ein spezieller Stimulusvektor auf $\mathcal{S}$, der mit einem Erregungsvektor $\mathbf{e}$ auf der Ausgangspopulation $\mathcal{E}$ beantwortet oder „assoziiert" werden soll. Gesucht ist also eine Matrix C mit der Eigenschaft $\mathbf{e} = C\mathbf{s}$. Eine solche Matrix erhält man aus der Kovarianzmatrix (oder dem äußeren Produkt) von Input- und Outputvektor. Wir setzen $C := \mathbf{e}\mathbf{s}^\top$ bzw. $c_{ij} = e_i s_j$ und erhalten

$$C\mathbf{s} \;=\; \left(\mathbf{e} \cdot \mathbf{s}^\top\right)\mathbf{s} = \mathbf{e}\left(\mathbf{s}^\top\mathbf{s}\right) = \mathbf{e}\|\mathbf{s}\|^2. \tag{3.15}$$

Die Koeffizienten $c_{ij} = e_i s_j \|\mathbf{s}\|^{-2}$ haben also die gewünschte Eigenschaft.

Speicherung mehrerer Assoziationen. Wir betrachten nun einen orthonormalen Satz von Inputvektoren $\mathbf{s}^{(p)}$ (d.h. $\mathbf{s}^{(p)}\mathbf{s}^{(q)} = \delta_{pq}$),[5] die jeweils mit einem gegebenen Outputvektor assoziiert werden sollen. Wir wählen $C := \sum_q \mathbf{e}^{(q)} \cdot \mathbf{s}^{(q)\top}$ und erhalten:

$$C\mathbf{s}^{(p)} \;=\; \left(\sum_q \mathbf{e}^{(q)} \cdot \mathbf{s}^{(q)\top}\right)\mathbf{s}^{(p)} = \sum_q \mathbf{e}^{(q)}\left(\mathbf{s}^{(q)\top}\mathbf{s}^{(p)}\right) = \sum_q \mathbf{e}^{(q)}\delta_{pq} = \mathbf{e}^{(p)}. \tag{3.16}$$

Wegen der vorausgesetzten Orthonormalität kann dieses Verfahren auch bei geeigneter Vorkodierung nur funktionieren, solange die Anzahl der Musterpaare kleiner oder gleich J (Anzahl der Inputneurone = Dimension des Merkmalsraumes) ist. Im Falle P zu assoziierender Paare kann man allgemein so vorgehen: Man bezeichnet mit $S := [\mathbf{s}^{(1)}; \mathbf{s}^{(2)}; ...; \mathbf{s}^{(P)}]$ und $E := [\mathbf{e}^{(1)}; \mathbf{e}^{(2)}; ...; \mathbf{e}^{(P)}]$ die aus den P Input- bzw. Outputvektoren spaltenweise zusammengesetzten Matrizen. Dann lässt sich zeigen, dass die Verknüpfungsmatrix C die Form

$$C = ES^\top(SS^\top)^{-1} \tag{3.17}$$

annehmen muss, um die gewünschten Assoziationen mit minimalem quadratischen Fehler zu liefern. Ist $P \leq J$ und sind die $\mathbf{s}^{(q)}$ linear unabhängig, so liefert C sogar exakte Lösungen. $SS^\top$ ist die Kovarianzmatrix der Inputmenge, von der wir der Einfachheit halber annehmen wollen, dass sie regulär ist. Die Matrix $S^\top(SS^\top)^{-1}$ heißt *Moore-Penrose Pseudo-Inverse* von S.

Autoassoziator. Beim Autoassoziator (Abb. 3.12b) wird nicht zwischen einer Input- und einer Outputschicht unterschieden; vielmehr betrachtet man eine vollständig verknüpfte Population von N Zellen. Das Input-Muster stellt dann eine Anfangsbedingung dar, von der ausgehend der Assoziator durch seine intrinsische Dynamik einen Endzustand (Attraktor) anstrebt, der das gewünschte Outputmuster ist. Zu jedem Attraktor gehört ein Attraktionsgebiet, d.h. eine Teilmenge des Inputraumes, mit der Eigenschaft, dass das Netzwerk Inputs aus diesem Gebiet mit dem zugehörigen Attraktor beantwortet.

Im zeitdiskreten, linearen Fall ergibt sich durch Iteration der Gleichung $\mathbf{e}(t+1) = W\mathbf{e}(t)$, $(\mathbf{e}(0) := \mathbf{s})$ im Endergebnis eine Projektion[6], mit der Gleichung

$$\mathbf{e}^{end.} = V\mathbf{s}, \quad \text{mit} \quad V := \lim_{t\to\infty} W^t. \tag{3.18}$$

[5] Hochgestellte Indizes numerieren die Mustervektoren (Trainingssatz).

[6] Eine lineare Abbildung $P : \mathbb{R}^n \to \mathbb{R}^n$ mit der Eigenschaft $P^2 = P$ heißt (lineare) *Projektion*. Abschaulich ist das Bild von P der Unterraum, auf den projiziert wird. Der Kern von P markiert die Projektionsrichtung.

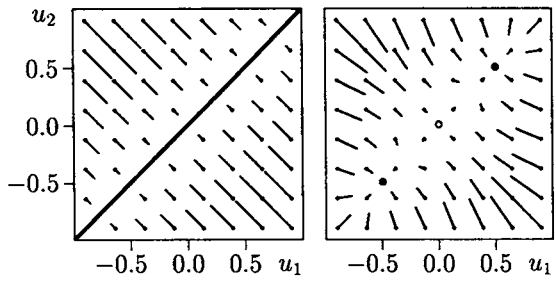


Abbildung 3.13: Dynamik des Autoassoziators aus Gl. 3.19. **a**, linearer Fall: Alle Punkte auf der Diagonalen sind Fixpunkte. Eingänge werden entlang der Linien projiziert. **b**, nichtlinearer Fall: $(0,0)$ ist instabiler Fixpunkt. Die zugehörige Projektionslinie trennt die Attraktionsgebiete der beiden stabilen Attraktoren (•). Die „Nadeln" deuten die Richtungen von $\dot{\mathbf{u}}$ an.

Interessant sind nur die Fälle, in denen der größte Eigenwert von W eins ist; andernfalls konvergiert die Folge W^t nicht oder gegen null. Ist 1 etwa k-facher Eigenwert von W $(0 < k \leq N)$, so spannen die zugehörigen Eigenvektoren einen k-dimensionalen linearen Unterraum auf, auf den die Abbildung V den Eingangsraum projiziert. Jeder Punkt im Bild von V, $\{V\mathbf{e} : \mathbf{e} \in \mathbb{R}^N\}$, ist ein Fixpunkt. Alle Punkte des Eingangsraumes, die auf ihn projiziert werden, liegen in einer $(N-k)$-dimensionalen Hyperebene durch den Fixpunkt (vgl. Abb. 3.13). Man kann diese Hyperebene als eine Art „Attraktionsgebiet" auffassen, doch handelt es sich im Gegensatz zu echten Attraktionsgebiete nicht um eine offene Menge. Benachbarte Startwerte können daher immer auf verschiedene Fixpunkte laufen (vgl. auch Abs. 3.4.3).

Im nichtlinearen Fall kann es mehrere isolierte Attraktoren mit echten Attraktionsgebieten geben. Anschaulich entsprechen die Attraktoren z.B. den Minima eines Potentialgebirges und die Grenzen der Attraktionsgebiete den Wasserscheiden in diesem Gebirge. Die Klassifikationseigenschaft nichtlinearer Autoassoziatoren ist also stärker als im linearen Fall: Ergebnisse „zwischen" den Attraktoren können nicht auftreten.

Beispiel. Abb. 3.13 zeigt die Dynamik des aus zwei Zellen gebildeten Autoassoziators

$$\begin{pmatrix} \dot{u}_1 \\ \dot{u}_2 \end{pmatrix} = -\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} \phi(u_1) \\ \phi(u_2) \end{pmatrix}, \quad \mathbf{u}(0) = \mathbf{s} \tag{3.19}$$

(vgl. Gleichung 3.5). Die Gewichtsmatrix hat die Eigenwerte 1 (Eigenvektor $(1,1)^\top$) und 0.5 (Eigenvektor $(-1,1)^\top$). Im linearen Fall (Abb. 3.13a) ist $\phi(u) \equiv u$ und alle Punkte auf der Diagonalen $u_1 = u_2 =: u$ sind Fixpunkte, auf die das System jeweils von den Startwerten $(u-\lambda, u+\lambda)^\top$ für alle $\lambda \in \mathbb{R}$ zuläuft. Im nichtlinearen Fall (Abb. 3.13b) wurde $\phi(u) = \frac{2}{\pi}\arctan(2u)$ gewählt. Man erhält dann einen instabilen Fixpunkt $\mathbf{u}_0 = (0,0)$ sowie zwei stabile Attraktoren $\mathbf{u}_{1,2} = (\pm 0.5, \pm 0.5)$ mit den echten Attraktionsgebieten $\{u_1 > u_2\}$ und $\{u_1 < u_2\}$.

3.3.3.3 Abbildende Netzwerke

Viele Anwendungen in der Signalverarbeitung erfordern die mehr oder weniger stetige Transformation eines Erregungsvektors in einen anderen. Neuronale Netzwerke können hier eingesetzt werden, um solche hochdimensionalen Abbildungen (diskret: Funktionen,

kontinuierlich: Operatoren) zu approximieren. Der Übergang zu den assoziativen Netzwerken ist fließend: beim Autoassoziator ist die Abbildung des Inputraumes auf die Outputs (Attraktoren) bereits stückweise stetig.

Filteroperationen. Ein kontinuierliches, einschichtiges Netzwerk mit Inputverteilung $s(\mathbf{x})$ und Erregungsverteilung $e(\mathbf{x})$ (vgl. Gl. 3.6) kann als lineares Filter im Sinne der Bildverarbeitung aufgefasst werden. Hängt die Koppelstärke zweier Neuronen an den Positionen $\mathbf{x}$ und $\mathbf{x}'$ nur von ihrem (gerichteten) Abstand $\mathbf{x} - \mathbf{x}'$ ab, so ergibt sich z.B. eine einfache Faltungsoperation:

$$e(\mathbf{x}) = \int w(\mathbf{x} - \mathbf{x}')s(\mathbf{x}')d\mathbf{x}'. \tag{3.20}$$

In der Neurobiologie entspricht w dem rezeptiven Feld der Zelle an der Position $\mathbf{x}$ (vgl. Gln. 3.1, 3.2). Die Idee, dass rezeptive Felder mit den Merkmalsdetektoren der Bildverarbeitung in Verbindung gebracht werden können, hat großen Einfluss auf beide Disziplinen ausgeübt.

Selbstorganisierende Merkmalskarten. Während über die Selbstorganisation noch zu reden sein wird (→ Gewichtsdynamik), soll hier die Idee der „Merkmalskarte" (*feature-map*) bereits kurz skizziert werden. Im Unterschied zur Filteroperation ist hier $W(\mathbf{x}, \mathbf{x}')$ aus Gleichung 3.6 nicht von der Form $w(\mathbf{x} - \mathbf{x}')$. Vielmehr werden von jeder „Zelle" $\mathbf{x}$ andere Reizeigenschaften betrachtet. Ein Beispiel ist die Selbstorganisation orientierungsspezifischer Kantendetektoren, die mit stetig variierender Vorzugsorientierung in der neuronalen Schicht angeordnet sind (vgl. Abs. 3.4.4). Allgemeiner kann auf diese Weise der für eine gegebene Reizklasse jeweils optimale Filtersatz gefunden werden (etwa im Sinne der Hauptachsentransformation). Dieser Typ neuronaler Netze erzeugt also Datenrepräsentationen.

Mehrschichtige diskrete Netzwerke. Als Anwendung eines Satzes von Kolmogorov kann man zeigen, dass jede stetige Abbildung $f : [0,1]^n \longrightarrow \mathbb{R}^m$ durch ein dreischichtiges vorwärts gekoppeltes Netzwerk mit n Inputzellen, m Outputzellen und $2n+1$ Zellen in der mittleren Schicht exakt implementiert werden kann. Dabei müssen für alle Zellen (fast) beliebige nichtlineare Kennlinien zugelassen werden. Leider ist der Beweis nicht konstruktiv (vgl. [Hecht-Nielsen, 1990]). Ein Verfahren zur Approximation von solchen Funktionen in Netzen mit differenzierbaren Nichtlinearitäten ist die sogenannte *back-propagation*. Mächtigere Verfahren erhält man durch Einsatz abgestimmter Neurone (Gl. 3.8) in der mittleren Schicht (vgl. Abs. 3.4.1: Radialbasisfunktionen).

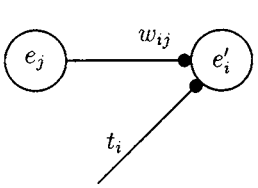


Abbildung 3.14: Komponenten von Lernregeln. e_j sei die Erregung des präsynaptischen Neurons zur Zeit t, e'_i die des postsynaptischen Neurons zur Zeit $t+1$. w_{ij} ist das aktuelle Übertragungsgewicht und t_i ein „Lehrersignal", das den gewünschten Zustand der postsynaptischen Zelle zur Zeit $t+1$ wiedergibt. Regeln mit Lehrersignal heißen „überwacht".